ISSN: 1646-9895

risti

Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Dezembro 20 • December 20



©AISTI 2020 http://www.aisti.eu

Nº E38

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Edição / Edition
N.º E38, 12/2020

ISSN: 1646-9895

Indexação / Indexing
Academic Journals Database, CiteFactor, Dialnet, DOAJ, DOI, EBSCO, GALE, Index-Copernicus, Index of Information Systems Journals, Latindex, ProQuest, QUALIS, SCImago, SCOPUS, SIS, Ulrich's.

Propriedade e Publicação / Ownership and Publication
AISTI – Associação Ibérica de Sistemas e Tecnologias de Informação
Rua Quinta do Roseiral 76, 4435-209 Rio Tinto, Portugal
E-mail: aistic@gmail.com
Web: http://www.aisti.eu

Director
Álvaro Rocha, Universidade de Lisboa, PT

Coordenadores da Edição / Issue Coordinators
Hugo Armando Ordoñez Erazo, Universidad del Cauca, CO

Conselho Editorial / Editorial Board
Carlos Ferrás Sexto, Universidad de Santiago de Compostela, ES
Gonçalo Paiva Dias, Universidade de Aveiro, PT
Jose Antonio Calvo-Manzano Villalón, Universidad Politécnica de Madrid, ES
Manuel Pérez Cota, Universidad de Vigo, ES
Ramiro Gonçalves, Universidade de Trás-os-Montes e Alto Douro, PT

Conselho Científico / Scientific Board
A. Augusto Sousa, FEUP, Universidade do Porto, PT
Abel Méndez Porras, Instituto Tecnológico de Costa Rica, CR
Abel Suing, Universidad Técnica Particular de Loja, EC
Adolfo Lozano-Tello, Universidad de Extremadura, ES
Adrián Hiebra Pardo, Universidad de Santiago de Compostela, ES
Alberto Fernández, Universidad Rey Juan Carlos, ES
Alberto Freitas, FMUP, Universidade do Porto, PT
Alcinia Zita Sampaio, IST, Universidade de Lisboa, PT
Alejandro Peña, Escuela de Ingeniería de Antioquia, CO
Alexandre L'Erario, Universidade Tecnológica Federal do Paraná, BR
Alicia García-Holgado, Universidad de Salamanca, ES
Alma Gomez-Rodríguez, Universidade de Vigo, ES
Ana Amélia Carvalho, Universidade de Coimbra, PT
Ana Beatriz Blanco-Ariza, Universidad Simón Bolívar, CO
Ana Isabel Veloso, Universidade de Aveiro, PT
Ana Maria Correia, ISEGI, Universidade Nova de Lisboa, PT
Ana Paula Afonso, Instituto Politécnico do Porto, PT
Anabela Mesquita, Instituto Politécnico do Porto, PT
Anacleto Correia, Escola Naval, PT

Angelica Caro, Universidad del Bío-Bío, CL
Ana Calvão, Universidade de Aveiro, PT
Ana Carla Amaro, Universidade de Aveiro, PT
Ana Melro, Universidade de Aveiro, PT
Ania Cravero, Universidad de La Frontera, CL
Aníbal Zaldivar-Colado, Universidad Autonoma de Sinaloa, MX
Antonia Mas, Universitat de les Illes Balears, ES
António Abreu, ISCAP, Politécnico do Porto, PT
António Coelho, FEUP, Universidade do Porto, PT
Antonio Fernández-Caballero, Unversidad de Castilla-La Mancha, ES
António Godinho, ISLA-Gaia, PT
Antonio Jesus Garcia Loureiro, Universidad de Santiago de Compostela, ES
Antonio Jiménez-Martín, Universidad Politécnica de Madrid, ES
Antoni Lluís Mesquida Calafat, Universidad de las Islas Baleares, ES
António Palma dos Reis, ISEG, Universidade de Lisboa, PT
António Pereira, Instituto Politécnico de Leiria, PT
Armando Mendes, Universidade dos Açores, PT
Arnaldo Martins, Universidade de Aveiro, PT
Arturo J. Méndez, Universidad de Vigo, ES
August Climent Ferrer, La Salle Open University, AD
Baltasar García Perez-Schofield, Universidad de Vigo, ES
Beatriz Rodríguez, Universidad de la Republica, UY
Beatriz Sainz de Abajo, Universidad de Valladolid, ES
Bernabé Escobar-Pérez, Universidad de Sevilla, ES
Borga Bordel, Universidad Politécnica de Madrid, ES
Bráulio Alturas, ISCTE - Instituto Universitário de Lisboa, PT
Brenda L. Flores-Rios, Universidad Autónoma de Baja California, MX
Carina Soledad González, Universidad de La Laguna, ES
Carlos Alexandre Silva, Instituto Federal de Minas Gerais, PT
Carlos Carreto, Instituto Politécnico da Guarda, PT
Carlos Morais, Instituto Politécnico de Bragança, PT
Carlos Regalao Noriega, Universidad Simón Bolívar, CO
Carlos Vaz de Carvalho, Instituto Politécnico do Porto, PT
Carmen Galvez, Universidad de Granada, ES

Carlos Rabadão, Politécnico de Leiria, PT
Carlos Rompante Cunha, Politécnico de Bragança, PT
Cesar Colazzos, Universidad del Cauca, CO
Ciro Martins, Universidade de Aveiro, PT
Cristina M.R. Caridade, ISEC, Politécnico de Coimbra, PT
Dalila Durães, Universidade do Minho, PT
Daniel Polónia, Universidade de Aveiro, PT
Daniel Riesco, Universidad Nacional de San Luis, AR
Dante Carrizo, Universidad de Atacama, CL
David Fonseca, Universitat Ramon Llull, ES
David Luis La Red Martínez, Universidad Nacional del Nordeste, AR
David Ramos Valcarcel, Universidad de Vigo, ES
Debora Paiva, Universidade Federal de Mato Grosso do Sul, BR
Diana Cecilia Yacchirema Vargas, Escuela Politécnica Nacional, EC
Dora Simões, Universidade de Aveiro, PT
Edna Dias Canedo, Universidade de Brasília, BR
Eduardo Amadeu Dutra Moresi, Universidade Católica de Brasília, BR
Eduardo Sánchez Vila, Universidad de Santiago de Compostela, ES
Edwin Juvenal Cedeño Herrera, Universidad de Panamá, PA
Enric Mor, Universitat Oberta de Catalunya, ES
Eusébio Ferreira da Costa, Escola Superior de Tecnologias de Fafe, PT
Eva Villegas, La Salle - Universitat Ramon Llull, ES
Fábio Marques, Universidade de Aveiro, PT
Fernando Bandeira, Universidade Fernando Pessoa, PT
Fernando Bobillo, Universidad de Zaragoza, ES
Fernando Moreira, Universidade Portucalense, PT
Fernando Paulo Belfo, ISCAC, Politécnico de Coimbra, PT
Fernando Ramos, Universidade de Aveiro, PT
Fernando Ribeiro, Politécnico de Castelo Branco, PT
Fernando Suárez, CPEIG, ES
Filipe Caldeira, Politécnico de Viseu, PT
Filipe Montargil, Politécnico de Lisboa, PT
Filipe Portela, Universiade do Minho, PT
Flor de María Sánchez Aguirre, Universidad César Vallejo, PE

Francisca Rosique Contreras, Universidad Politécnica de Cartagena, ES
Francisco Arcega, Universidad de Zaragoza, ES
Francisco Javier Lena-Acebo, Universidad de Cantabria, ES
Francisco Restivo, Universidade Católica Portuguesa, PT
Franyelit Suarez, Universidad de las Américas, EC
Gabriel Alberto García-Mireles, Universidad de Sonora, MX
Gabriel Guerrero-Contreras, Universidade de Cádiz, ES
Gerardo Gonzalez Filgueira, Universidad da Coruña, ES
Gladys Tenesaca Luna, Universidad Técnica Particular de Loja, EC
Gloria Maritza Valencia Vivas, Universidad de las Fuerzas Armadas, EC
Gloria Piedad Gasca-Hurtado, Universidad de Medellín, CO
Guilhermina Lobato Miranda, Universidade de Lisboa, PT
Hélder Gomes, Universidade de Aveiro, PT
Hélder Zagalo, Universidade de Aveiro, PT
Hélia Guerra, Universidade dos Açores, PT
Henrique S. Mamede, Universidade Aberta, PT
Higino Ramos, Universidad de Salamanca, ES
Inês Domingues, CI-IPOP, PT
Isabel de la Torre, University of Valladolid, ES
Isabel Pedrosa, Instituto Politécnico de Coimbra, PT
Isidro Calvo, Universidad del País Vasco (UPV/EHU), ES
Isidro Navarro, Universidad Politécnica de Cataluña, ES
Ismael Etxeberria-Agiriano, Universidad del País Vasco (UPV/EHU), ES
Ivaldir de Farias Junior, Universidade de Pernambuco, BR
Ivan Garcia, Universidad Tecnologica de la Mixteca, MX
João Paulo Ferreira, ISEC, Politécnico de Coimbra, PT
João Reis, Universidade de Aveiro, PT
João Roberto de Toledo Quadro, CEFET/RJ, BR
Jacinto Estima, Universidade Europeia, PT
Javier Garcia Tobio, CESGA-Centro de Supercomputacion de Galicia, ES
Javier Medina, Universidad Distrital Francisco José de Caldas, CO
Jeimy Cano, Universidad de los Andes, CO
Jezreel Mejia, Centro de Investigación en Matemática (CIMAT), MX
João Balsa, FC, Universidade de Lisboa, PT

João Paulo Costa, Universidade de Coimbra, PT
João Tavares, FEUP, Universidade do Porto, PT
João Vidal de Carvalho, ISCAP, Politécnico do Porto, PT
Joaquim Ferreira, Universidade de Aveiro, PT
Joaquim Reis, ISCTE - Instituto Universitário de Lisboa, PT
Jorge Bernardino, ISEC, Politécnico de Coimbra, PT
Jorge Buele, Universidad Tecnológica Indoamérica, EC
Jorge Eduardo Ibarra-Esquer, Universidad Autónoma de Baja California, MX
Jorge da Silva Correia-Neto, Universidade Federal Rural de Pernambuco, BR
Jose Alfonso Aguilar, Universidad Autonoma de Sinaloa, MX
José Alvarez-Garcia, Universidad de Extremadura, ES
José Borbinha, IST, Universidade de Lisboa, PT
José Carlos Ribeiro, Politécnico de Leiria, PT
José Cascalho, Universidade dos Açores, PT
José Felipe Cocón Juárez, Universidad Autónoma del Carmen, MX
Jose J. Pazos-Arias, Universidad de Vigo, ES
José Luís Pereira, Universidade do Minho, PT
José Luís Silva, Instituto Universitário de Lisboa (ISCTE-IUL), PT
José Paulo Lousado, Instituto Politécnico de Viseu, PT
José Luis Pestrana Brincones, Universidad de Málaga, ES
José Luís Reis, ISMAI - Instituto Universitário da Maia, PT
Jose M Molina, Universidad Carlos III de Madrid, ES
José Machado, Universidade do Minho, PT
Jose Maria de Fuentes, Universidad Carlos III de Madrid, ES
Jose R. R. Viqueira, Universidade de Santiago de Compostela, ES
José Silvestre Silva, Academia Militar, PT
José Torres, Universidade Fernando Pessoa, PT
Josep M. Marco-Simó, Universitat Oberta de Catalunya, ES
Juan Angel Contreras Vas, Universidad de Extremadura, ES
Juan D'Amato, PLADEMA-UNCPBA-CONICET, AR
Juan M. Santos Gago, Universidad de Vigo, ES
Jugurta Lisboa-Filho, Universidade Federal de Viçosa, BR
Klinge Orlando Villalba-Condori, Universidad Católica de Santa María, PE
Leila Weitzel, Universidade Federal Fluminense, BR

Leonardo Bermon, Universidad Nacional de Colombia, CO
Leticia Morales Trujillo, Universidad de Sevilla, ES
Lilia Muñoz, Univerrsidad Tecnológica de Panamá, PA
Lucila Ishitani, PUC Minas, BR
Lucila Romero, Universidad Nacional del Litoral, AR
Luis Alvarez Sabucedo, Universidad de Vigo, ES
Luís Bruno, Instituto Politécnico de Beja, PT
Luis Camarinha-Matos, Universidade Nova de Lisboa, PT
Luís Cavique, Universidade Aberta, PT
Luis Chamba Eras, Universidad Nacional de Loja, EC
Luís Ferreira, Instituto Politécnico do Cávado e do Ave, PT
Luis Enrique Sánchez Crespo, Universidad de Castilla-La Mancha, ES
Luis Vilán-Crespo, Universidad de Vigo, ES
Luis Maria Romero-Moreno, Universidad de Sevilla, ES
Luisa Miranda, Instituto Politécnico de Bragança, PT
Lus Sussy Bayona Ore, Universidad Nacional Mayor de San Marcos, PE
Luz María Hernández Cruz, Universidad Autónoma de Campeche, MX
Magdalena Arcilla Cobián, Universidade Nacional de Educación a Distancia, ES
Manuel Fernández-Veiga, Universidad de Vigo, ES
Manuel Jose Fernandez Iglesias, Universidad de Vigo, ES
Marcelo Marciszack, Universidad Tecnológica Nacional, AR
Marcelo de Paiva Guimarães, Universidade Federal de São Paulo, BR
Marcelo Zambrano Vizuete, Universidad Tecnica Del Norte, EC
Marco Javier Suarez Barón, Universidad Pedagógica y Tecnológica de Colombia, CO
Marco Painho, ISEGI, Universidade Nova de Lisboa, PT
Margarita Ramirez Ramirez, Universidad Autonoma de Baja California, MX
Maria Amelia Eliseu, Mackenzie Presbyterian University, BR
Maria Cristina Marcelino Bento, UNIFATEA, BR
María de la Cruz del Río-Rama, Universidad de Vigo, ES
Maria de los Milagros Gutierrez, Universidad Tecnológica Nacional, AR
María del Mar Miras Rodriguez, Universidad de Sevilla, ES
Maria do Rosário Bernardo, Universidade Aberta, BR
Maria Hallo, Escuela Politécnica Nacional, EC
Maria Helena Garcia Ruiz, Universidad de Camtabria, ES

María J. Lado, Universidad de Vigo, ES
Maria João Ferreira, Universidade Portucalense, PT
Maria João Gomes, Universidade do Minho, PT
Maria José Angélico, Instituto Politécnico do Porto, PT
Maria José Escalona, Universidad de Sevilla, ES
Maria José Sousa, Universidade Europeia, PT
Mario Chacón-Rivas, Instituto Tecnológico de Costa Rica, CR
Mario José Diván, Universidad Nacional de La Pampa, AR
Marisol B. Correia, Universidade do Algarve, PT
Maristela Holanda, Universidade de Brasília, BR
Martín Llamas Nistal, Universidad de Vigo, ES
Martín López Nores, Universidad de Vigo, ES
Matías García Rivera, Universidad de Vigo, ES
Mercedes Ruiz, Universidad de Cádiz, ES
Miguel A. Brito, Universidade do Minho, PT
Miguel Ángel Conde, Univesidad de León, ES
Miguel Angel Olivero Gonzalez, Universidad de Sevilla, ES
Miguel Bugalho, Universidade Europeia, PT
Miguel Casquilho, IST, Universidade de Lisboa, PT
Miguel Ramón González Castro, Ence, Energía y Celulosa, ES
Mirna Ariadna Muñoz Mata, Centro de Investigación en Matemáticas (CIMAT), MX
Nelson Rocha, Universidade de Aveiro, PT
Nuno Lau, Universidade de Aveiro, PT
Nuno Melão, Politécnico de Viseu, PT
Nuno Ribeiro, Universidade Fernando Pessoa, PT
Oscar Mealha, Universidade de Aveiro, PT
Patricia Dias, Universidade do estado de Minas Gerais, BR
Patrícia Oliveira, Universidade de Aveiro, PT
Paula Prata, Universidade da Beira Interior, PT
Paulo Martins, Universidade de Trás-os-Montes e Alto Douro, PT
Paulo Pinto, FCT, Universidade Nova de Lisboa, PT
Paulo Rurato, Universidade Fernando Pessoa, PT
Paulo Urbano, FC, Universidade de Lisboa, PT
Pedro Araújo, Universidade da Beira Interior, PT

Pedro Palos, Universidad de Sevilla, ES
Pedro Sanz Angulo, Universidad de Valladolid, ES
Pedro Sobral, Universidade Fernando Pessoa, PT
Pedro Sousa, Universidade do Minho, PT
Pilar Mareca Lopez, Universidad Politécnica de Madrid, ES
Ramiro Delgado, Universidad de las Fuerzas Armadas, EC
Ramon Alcarria, Universidad Politécnica de Madrid, ES
Raul Laureano, ISCTE - Instituto Universitário de Lisboa, PT
Renato Maurício Toasa Guachi, Universidad Tecnológica Israel, EC
Rene Faruk Garzozi-Pincay, Universidad Estatal Península de Santa Elena, EC
Ricardo Andrés García León, Universidad Francisco de Paula Santander Ocaña, MX
Ricardo J. Rodríguez, Universidad de Zaragoza, ES
Ricardo Linden, FSMA, BR
Rita Oliveira, Universidade de Aveiro, PT
Rita Santos, Universidade de Aveiro, PT
Robero Marichal, Universidad de La Laguna, ES
Roberto Theron, Universidad de Salamanca, ES
Rodolfo Miranda Barros, Universidade Estadual de Londrina, BR
Román Lara, Universidad de las Fuerzas Armadas, EC
Rubén González Crespo, Universidad Internacional de La Rioja, ES
Rui Cruz, IST, Universidade de Lisboa, PT
Rui José, Universidade do Minho, PT
Rui Pedro Marques, Universidade de Aveiro, PT
Rui S. Moreira, Universidade Fernando Pessoa, PT
Samuel Sepúlveda, Universidad de La Frontera, CL
Santiago Gonzales Sánchez, Universidad Inca Garcilaso de la Vega, PE
Sara Balderas-Díaz, Universidad de Cádiz, ES
Saulo Barbara de Oliveira, Universidade Federal Rural do Rio de Janeiro, BR
Sérgi F. Lopes, Universidade do Minho, PT
Sergio Araya Guzmán, Universidad del Bío-Bío, CL
Sergio Gálvez Rojas, Universidad de Málaga, ES
Sérgio Guerreiro, Universidade Lusófona de Humanidades e Tecnologias, PT
Silvia Fernandes, Universidade do Algarve, PT
Solange N Alves de Souza, Universidade de São Paulo, BR

Telmo Silva, Universidade de Aveiro, PT
Teresa Guarda, Universidad Estatal Península de Santa Elena, EC
Tomas San Feliu, Universidad Politecnica de Madrid, ES
Thiago Dias, CEFET-MG, BR
Valéria Farinazzo Martins, Universidade Presbiteriana Mackenzie, BR
Verónica Vasconcelos, ISEC, Politécnico de Coimbra, PT
Vicente Morales, Universidad Técnica de Ambato, EC
Victor Flores, Universidad Católica del Norte, CL
Víctor H. Andaluz, Universidad de las Fuerzas Armadas ESPE, EC
Victor Hugo Medina Garcia, Universidad Distrital Francisco José de Caldas, CO
Vitor Carvalho, Instituto Politécnico do Cávado e do Ave, PT
Vitor Santos, ISEGI, Universidade Nova de Lisboa, PT
Wagner Tanaka Botelho, Universidade Federal do ABC, BR

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 12/08/2020
Aceitação/Acceptance: 02/10/2020

Editorial

La Ingeniería del Software y del Conocimiento en la industria de Base Tecnológica Actual (JIISIC'2020)

Software and Knowledge Engineering in the Current Technology Base industry (JIISIC'2020)

Hugo Ordóñez[1], César Pardo[1], Carlos Cobos[1]

hugoordonez@unicauca.edu.co; cpardo@unicauca.edu.co; ccobos@unicauca.edu.co

Grupo de I+D en Tecnologías de la Información (GTI), Departamento de Sistemas, Facultad de Ingeniería Electrónica y Telecomunicaciones, Universidad del Cauca, Calle 5 N.º 4-70, 190003, Popayán, Cauca, Colombia.

Pages: xi–xii

En la actual sociedad de conocimiento, la industria del software se enfrenta a la solución de problemas cada vez más complejos y a clientes que día a día tienen un mayor nivel de exigencia, demandando mayor agilidad en los procesos, mayor calidad en los productos, mayor seguridad, integración e inteligencia en las soluciones provistas, entre otras.

Anteriormente, el enfoque de desarrollo de soluciones software era secuencial y basada en una fase inicial de definición de requerimientos, luego, el diseño de la solución software por parte de arquitectos y diseñadores, la posterior implementación o codificación de la solución conforme al diseño, después, la ejecución de una etapa en general liviana de pruebas; y finalmente, la transferencia del producto software a un equipo de operaciones para su puesta en producción. Actualmente las cosas han cambiado significativamente, debido a las exigencias previamente mencionadas y a una más dinámica transferencia de las investigaciones realizadas en diversos temas en la academia y la industria. Muchas de estas investigaciones han permitido potencializar el crecimiento de las prácticas ágiles, con lo cual; se ha transformado el trabajo diario del Ingeniero de Software para que pueda participar en los proyectos desde la concepción hasta las operaciones. Además, hoy los equipos empiezan a incluir de forma permanente miembros con otros perfiles y habilidades tanto técnicas como blandas, por ejemplo: expertos en inteligencia artificial, ciencia de datos, internet de las cosas, seguridad, y algunas habilidades blandas como autogestión, curiosidad, comunicación, colaboración, coordinación, entre otras.

Teniendo en cuenta lo anterior, el comité organizador de las XV Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento - JIISIC 2020 definió cinco grandes temas o tópicos para la presentación de los resultados de investigación realizados recientemente. El primero de ellos se relaciona con conceptos clásicos y avanzados de la Ingeniería de Software, entre ellos: la ingeniería de requisitos, el diseño y la construcción de software, la integración y operación de productos software, la medición del software, la mejora de procesos, la calidad del software, entre otros. El segundo tópico se relaciona con los Sistemas Inteligentes, que agrupa temas como: el aprendizaje de máquina, el aprendizaje profundo, la optimización basada en metaheurísticas, las aplicaciones de estos conceptos, entre otros. El tercer tópico se relaciona con los Sistemas de Soporte a la Toma de Decisiones, que en múltiples casos toma técnicas de los sistemas inteligentes, pero se centra en el uso de los datos organizacionales para obtener información o conocimiento que soporte la toma de decisiones, e incluye temas como: el Big Data, la minería de datos, las bodegas de datos, la inteligencia de negocios, entre otros. El cuarto tópico del evento se centra en lo que se denominó Sistemas Interactivos, que agrupa la Internet de las cosas, la Web semántica, las ontologías, las realidad virtual y aumentada, la comunicación máquina a máquina, los videojuegos, entre otros. Finalmente, pero con un gran interés por parte de la comunidad organizadora de este evento, el quinto tópico se relaciona con el papel de las Mujeres en la Computación, ya que se hace necesario resaltar los trabajos que profesionales están realizando en esta temática; y con ello, fomentar la participación de la mujer en la Informática y la Computación, donde sin lugar a duda son muy importantes.

Las JIISIC'2020 se realizaron del 16 al 18 de septiembre en forma virtual y contaron con la participación de conferencistas principales de diferentes países, Ecuador, España, Estados Unidos, Francia, México y Colombia. En relación con las ponencias, se recibieron en total 98 artículos, que luego de un proceso de evaluación doble ciego por pares internacionales, se aceptaron y presentaron 57 artículos (de Chile, Costa Rica, Ecuador, España, México, Perú y Colombia). De esta manera, las JIISIC se siguen consolidando como un espacio de divulgación y debate de resultados de investigación en el área de la ingeniería del software e ingeniería del conocimiento a nivel iberoamericano.

En este número especial de la revista, se presenta un selecto grupo de artículos relacionados con los tópicos del evento, treinta (30) en total. En Sistemas Inteligentes se publican 11 artículos, en Ingeniería del software se presentan 8 artículos, en Sistemas de Soporte a la Toma de Decisiones se publican 7 artículos, en Sistemas Interactivos se presentan 3 artículos, y finalmente; en el tópico de Mujeres en Ciencias de la Computación se publica 1 articulo. Estos trabajos presentan aportes científicos, así como soluciones tecnologías que impactan directamente el desarrollo de la ingeniería de software y la ingeniería del conocimiento.

Índice / Index

EDITORIAL

La Ingeniería del Software y del Conocimiento en la industria de Base Tecnológica Actual (JIISIC'2020) ix
Hugo Ordóñez, César Pardo, Carlos Cobos

ARTIGOS / ARTICLES

EiBeLec: Propuesta de sistema adaptativo para el ecosistema lácteo..... 1
Javier A. Ballesteros-Ricaurte, Angela Carrillo-Ramos, Ramon Fabregat, Carlos Parra-Acevedo, Juan-Sebastián González-Sanabria

Análisis de exactitud de reconocimiento gestual aplicando SVM y k-NN en señales EMG 15
Gonzalo Pomboza-Junez, Juan A. Holgado-Terriza

Una Mirada al Desempeño Académico en las Pruebas Saber Pro de los Estudiantes de Ingeniería desde la Minería de Datos Educativa 29
Ricardo Timarán-Pereira, Arsenio Hidalgo-Troya, Fredy Vidal-Alegría

Una revisión sistemática de los métodos de identificación y detección de corrupción en contratación pública 43
Julio-César Luna-Ortega, Carlos-Alberto Cobos-Lozada, Martha-Eliana Mendoza-Becerra

Estrategia para la homogeneización, comparación e integración de Scrum e ISO 9001:2015..... 58
César Pardo, Miguel Morcillo, Elizabeth Suescún

SCMOnto: Una ontología para soportar la gestión de la configuración de software..... 75
Carlos Orozco, César Pardo, Sebastián Vásquez

Bloques de primos para acelerar el algoritmo Quadratic Sieve..... 91
Juan Manuel Campo Revelo, Jhon Alejandro Melo Zúñiga, Carlos Alberto Trujillo Solarte, Siler Amador Donado

Mapeo Sistemático sobre la Web Semántica de las Cosas y el Cuidado de Adultos Mayores 103
Andrea Pabón-Guerrero, Miguel-Ángel Niño-Zambrano, Lider-Julian Rojas-Bolaños

Clasificación de noticias criminales basada en procesamiento del lenguaje natural y algoritmos de aprendizaje automático..... 117
Camilo Ernesto Sarmiento Torres, Néstor Diaz, y Rubiel Vargas Cañas

Inteligencia de Negocios para el Análisis de la Accidentalidad Vial en la Ciudad de Popayán 130
Carlos H. Bolaños M., Inés A. Damián F., Héctor F. Muñoz M., Andrés F. Negrete G., Miguel A. Tunubalá M., Christian G. Arias I., María A. Varona T.

Niveles de Autoeficacia Matemática en Estudiantes Mujeres y Varones Aspirantes a la Carrera de Ingeniería 142
Claudia Mego-Sanchez, Ling Huaman-Sarmiento, Ivan Iraola-Real, Alonso Iraola-Arroyo

Evaluación de la precisión de los modelos de clasificación binaria para la identificación de noticias verdaderas o falsas en Costa Rica..... 156
Eilyn Salazar Miranda, Gustavo Tenorio Arce, Luis Naranjo Zeledón

Propuesta Arquitectónica de un Sistema de Recomendación Híbrido Adaptativo..... 171
Daniela Varela, Jose Aguilar, Julián Monsalve-Pulido, Edwin Montoya

Clasificación de Imágenes de Cámaras de Vigilancia Volcánica usando Visión Artificial 185
César D. Martínez, Roberto C. Uribe, Carlos Ospina, Rubiel Vargas-Cañas

Aplicación de métodos lineales de reducción de dimensión y redes neuronales convolucionales para la predicción de estructura secundaria de proteínas..... 199
Diego Fernando Peña, Néstor Díaz, Víctor Andrés Bucheli

Computación Cuántica Adiabática aplicada a la solución del Problema de la Mochila Binaria..... 214
Danilo López-Sandoval, Carlos-Alberto Cobos-Lozada

Generación automática de resúmenes extractivos de múltiples documentos basados en SGVNS e IHEA 228
Juan-Manuel Díaz C., Anderson Buitrón P, Martha-Eliana Mendoza B

Vari::Ation - De modelos de características a modelos de configuración 245
Alonso Bobadilla Montoya, Samuel Sepúlveda Cuevas

DataOps, una alternativa para la gestión de Data Pipelines 259
Laura Vargas-Rueda, Sergio Pongutá-Díaz, Juan-Sebastián González-Sanabria

Plataforma Software Extensible para Campus Inteligente Basada en Microservicios 270
Henry Jiménez, Emilio Cárcamo, Gabriel Pedraza

Diseño analítico de reglas de decisión en procesos timetabling para reasignación de espacios académicos 283
Monica K. Duran-Vaca, J. A. Ballesteros-Ricaurte, A. M. González-Amarillo

Anna: Modelo para el análisis de requisitos de software y la identificación automática de variables, valores y generación de casos de pruebas funcionales de software 297
Dilan Mejía, Hugo Ordóñez, César Pardo

Comparación de funciones kernel para la predicción de la oferta energética fotovoltaica 310
Héctor Mora-Paz, Jaime A. Riascos, J.A Salazar-Castro, Germán Mora, Andrés Pantoja, Javier Revelo-Fuelagán, Laura Mancera-Valetts, Diego Peluffo-Ordoñez

Mapeando la evolución investigativa de la web semántica: análisis y visualización cienciométrica 325
Santiago Guerrero Narváez, Miguel Ángel Niño Zambrano, Gustavo Adolfo Ramírez González

Trading Algorítmico para la Predicción de Series de Tiempo Financieras: Una Revisión Sistemática 337
Jesús-Rodrigo Navia-Rodríguez, Carlos-Alberto Cobos-Lozada, Martha-Eliana Mendoza-Becerra

Pautas del modelo de usabilidad PACMAD para la construcción de un aplicativo móvil de gestión de actividades académicas 358
Celinda Isabel Cueto De la hoz, Carlos Javier Alcalá Orozco, Gabriel Elías Chanchí Golondrino, Martin Emilio Monroy Ríos

Mitigación del ataque replay en automóviles usando SDR y CNN 373
Juan Carlos Martínez Quintero, Edith Paola Estupiñán Cuesta, Saira Vanessa Rodriguez Diaz

Entornos Virtuales para Asistir el Aprendizaje de las Pruebas Software: Meta-Análisis para una Familia de Experimentos..... 389
Juan Ucán Pech, Omar S. Gómez, Julio Díaz Mendoza, Raúl Aguilar Vera

Algoritmo de optimización de colonia de hormigas aplicado a TSP, una revisión sistemática..... 404
Francisco Obando-Vidal, Néstor Díaz-Mariño, Ember Martínez-Flor

Evaluación del impacto en las guías de diseño aplicadas al contexto de ambientes virtuales de aprendizaje (AVA) para dispositivos móviles 418
Manuel Pastrana, Jhon Cano, Julián Mera

Priorização de Requisitos de Software Derivados de Modelos de Processos de Negócio: abordagem utilizando o Modelo de Kano 431
Rodrigo Oliveira Zacarias, Karina Terra de Souza, Thayná Albernaz Pessanha, Simone Vasconcelos Silva, Aline P. V. de Vasconcelos, Alline Sardinha C. Morais

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 20/08/2020
Aceitação/Acceptance: 31/10/2020

EiBeLec: Propuesta de sistema adaptativo para el ecosistema lácteo

Javier A. Ballesteros-Ricaurte[1,2], Angela Carrillo-Ramos[2], Ramon Fabregat[3], Carlos Parra-Acevedo[2], Juan-Sebastián González-Sanabria[1]

javier.ballesteros@uptc.edu.co; angela.carrillo@javeriana.edu.co; ramon.fabregat@udg.edu; ca.parraa@javeriana.edu.co; juansebastian.gonzalez@uptc.edu.co

[1] Escuela de Ingeniería de Sistemas y Computación, Universidad Pedagógica y Tecnológica de Colombia, 150003, Tunja, Colombia.

[2] Doctorado en Ingeniería, Pontificia Universidad Javeriana, 110231, Bogotá D.C., Colombia.

[3] Comunicaciones en Banda Ancha y Sistemas Distribuidos (BCDS), Universitat de Girona, 17003, Girona, España.

Pages: 1-14

Resumen: Este artículo presenta una propuesta de sistema adaptativo para el ecosistema lácteo del departamento de Boyacá en Colombia. El ecosistema requiere manejo de información contextual (por ejemplo, dispositivo, rol del usuario, ubicación, visualización de información, etc.) para que el sistema se pueda adaptar y brindar una experiencia personalizada a cada usuario, según sus características y las de su contexto. Para el desarrollo del sistema adaptativo, se propone una arquitectura software, basada en servicios, que admite múltiples configuraciones. Para ilustrar la propuesta se presentan dos servicios, uno para cada uno de los dos roles principales: el ganadero y el gobierno. Para el ganadero, se presenta el servicio "Generar ruta" que indica, entre varias opciones, dónde vender la leche para obtener un mejor precio, y la ruta desde su ubicación hasta el sitio de compra. Para el gobierno se presenta el servicio "Consultar contagio" que proporciona información sobre la propagación de una enfermedad.

Palabras-clave: Sistema adaptativo; Línea de productos de Software; Sistema de información web; Ecosistema lechero; Enfermedades infecciosas bovinas.

EiBeLec: Proposed Adaptive System for the Dairy Ecosystem

Abstract: This paper proposes an adaptive system for the dairy ecosystem of the Boyacá Department in Colombia. The ecosystem requires context information manipulation (e.g. device, user role, location, information visualization). Adaptation is needed in order to provide a tailored experience to each user according to her/his characteristics and those of her/his context. To face these challenges, we focus on the variability of the system and propose a service-based software architecture that supports multiple configurations. To illustrate the proposal, two services are presented, one for each of the two main actors: the rancher and the government. For the rancher, we present the Generate route service that indicates, among

several options, where to sell the milk to obtain a better price. For the government, the Consult contagion service, which presents information about the spread of a disease.

Keywords: Adaptive system; Software product line; Dairy ecosystem; Bovine infectious diseases.

1. Introducción

Los sistemas de información se utilizan para proveer a los usuarios con diferentes servicios como, por ejemplo, de salud humana, industria y agricultura, así como para el control (Bradhurst et al., 2016; Picault et al, 2017), y la erradicación de enfermedades infecciosas que ocurren en diferentes regiones ganaderas (Richards et al., 2014). Estas herramientas deben usar grandes volúmenes de datos para generar reportes que sean aportes para los procesos de toma de decisiones de los actores involucrados. Sin embargo, en la mayoría de los casos, la información se distribuye y organiza en diferentes formatos, pudiendo estar estructurada o no, pero siendo en muchos casos insuficiente la información disponible. Además, esta información depende del desarrollo de políticas, estrategias y competencias de parte del gobierno.

Este trabajo está enfocado en el ecosistema lácteo en Boyacá y los actores involucrados en este ecosistema generan datos no estructurados que se recopilan en diferentes formatos (por ejemplo, texto plano, hojas de cálculo, bases de datos) (Ballesteros-Ricaurte et al., 2018). Adicionalmente, no todo el mundo tiene acceso a esta información y de la que se dispone es insuficiente para ayudar a los ganaderos y al gobierno a que tomen la mejor decisión posible.

En la literatura analizada no se encontraron sistemas de información o aplicaciones que superen estos retos. En algunos casos, los datos sobre enfermedades no son reportados porque los ganaderos no tienen la obligación de hacerlo, o porque temen perder su ganado si lo hacen. Sin embargo, existen trabajos que muestran la forma en la que los datos epidemiológicos son utilizados, a fin de tomar decisiones respecto a enfermedades contagiosas (Barrett et al., 2008; Beckman, 2014; Carr & Roberts, 2010; Damianos, Wohlever, Kozierok, & Ponte, 2003; Picault et al., 2017; Castillejo et al., 2019). De forma similar, algunos expertos (Peeler & Otte, 2016; Singer, Salman, & Thulke, 2011), consideran políticas públicas de salud animal y modelado de enfermedades para la toma de decisiones basada en evidencias (Godsk, 2009), aunque no tienen en cuenta ni a los usuarios ni al contexto.

Por otro lado, se considera que el análisis de trabajos relacionados con la epidemiología de enfermedades infecciosas en humanos podría servir como inspiración para nuestra propuesta. Algunos artículos como los de (Beckman, 2014; Damianos et al., 2003) presentan una interfaz gráfica que permite la captura y representación de la información, y son orientados a usuarios expertos en el tema, es decir, los resultados proporcionados por las herramientas requieren que los usuarios tengan un conocimiento en análisis de datos epidemiológicos.

Para construir un sistema adaptativo para el ecosistema lácteo se identificaron tres retos principales: i) que sea configurable, para poder hacer cambios en enfermedades, actores

o contexto sin generar conflictos en otros componentes o servicios; ii) que se pueda adaptar al despliegue considerando el contexto tecnológico de cada actor y la forma en que la información puede presentarse en cada dispositivo; y iii) que la adaptación del contenido, dependiendo de las variables, pueda contemplar los diferentes niveles de granularidad de la información que puede ser presentada a los actores.

Teniendo en cuenta lo anterior, se propone EiBeLec, un sistema adaptativo sensible al contexto para apoyar la toma de decisiones del ecosistema lácteo. EiBeLec permite a los actores interactuar con servicios sobre enfermedades, contagios en la región, notificaciones sobre campañas preventivas, precios de leche y carne, rutas de distribución, y visualización de la información. Cada servicio tiene una adaptación que considera diversas características contextuales lo que permite a los usuarios tener una experiencia personalizada.

A fin de evaluar el sistema, se explican con detalle dos servicios, uno para el actor ganadero y otro para el actor gobierno. Para el primero, se describe el servicio "Generar ruta" que le permite al ganadero seleccionar el lugar donde puede vender su leche a un mejor precio, y además, le muestra la ruta desde su ubicación hasta el sitio de venta. Este servicio considera el perfil de usuario, su dispositivo hardware y el tipo de visualización de la información que el ganadero prefiere. Para el actor gobierno se detalla el servicio "Consultar contagio" que permite visualizar la presencia de enfermedades infecciosas bovinas.

El trabajo se organiza de la siguiente manera: la sección 2 presenta los conceptos fundamentales para entender el tema tratado. La sección 3 explica la motivación para esta investigación, y describe la arquitectura adaptativa propuesta. Finalmente, la sección 4 expone las conclusiones más relevantes y presenta algunas posibilidades de trabajo futuro.

2. Marco conceptual

El desarrollo de arquitecturas de sistemas adaptativos ha estado orientado frecuentemente a áreas como la educación, el turismo y los procesos de eneñanza/aprendizaje. Sin embargo, en el campo de la salud animal no se han encontrado avances significativos. Esta sección presenta una descripción general de las enfermedades infecciosas bovinas y de la generación de rutas; además, se describen brevemente los conceptos relacionados con la adaptación de la información en un contexto computacional.

2.1. Enfermedades infecciosas

Diferentes países trabajan en los problemas generados por las enfermedades infecciosas bovinas para evitar una expansión que pueda causar epidemias y pandemias (Siettos & Russo, 2013). Estas enfermedades infecciosas también generan problemas sociales, enfermedades humanas y pérdidas económicas (McConnel et al., 2015).

Estas enfermedades causan pérdida de peso, baja producción de leche, disminución en la eficiencia de la producción de los animales, tumores malignos en el tejido linfático y la muerte de los animales. Todo esto tiene un gran impacto económico en las granjas bovinas (Pulido-Medellín et al., 2017) debido al costo de los tratamientos, las muertes

prematuras y el reemplazo consecuente de animales enfermos, lo que provoca un decrecimiento en la producción de leche y restricciones en importación y exportación de algunos países.

2.2. Generar ruta

Generar ruta es un servicio que consiste en añadir funcionalidades y adaptaciones en el formato de la presentación del contenido, de forma que los actores del ecosistema de leche encuentren información acerca de los centros de acopio de leche y las plantas procesadoras que están en la región (Mojica et al., 2007), cuál es el precio promedio de compra de la leche, cuál es la ruta que debe seguirse para llegar al lugar y si el centro de acopio o la planta procesadora ofrecen el servicio de recolección de leche.

El proceso comienza cuando el ganadero hace una oferta de cierto número de litros de leche y proporciona la ubicación de su granja. El servicio debe informar sobre los precios actuales ofrecidos por las plantas procesadoras de lácteos. Después de que el ganadero seleccione una de las opciones, el servicio presenta la ubicación e información del centro de acopio o de la planta procesadora. Después, se le informa al ganadero si se le va recoger la leche o si debe llevarla a un centro de acopio o planta procesadora. Con toda esta información, el ganadero puede tomar la decisión con la que obtenga la mejor oferta por su producto.

2.3 Sistemas adaptativos

La adaptación en sistemas de información es un proceso en el cual la información que el usuario requiere le es proporcionada considerando sus necesidades, preferencias, gustos, dispositivo de acceso y, en general, sus características contextuales (Lancheros, 2014). Para alcanzar este objetivo, es necesario tomar en consideración no sólo el contenido de la información presentada, sino la forma en la que se muestra en el dispositivo de acceso del usuario (Carrillo et al., 2009). Es de resaltar que, aunque ha habido trabajos sobre epidemias de enfermedades humanas con desarrollo de software que incluye un componente adaptativo para las políticas (Bisset et al., 2014), a la fecha no se encuentran investigaciones que los apliquen para enfermedades de origen bovino.

3. EiBeLec

EiBeLec es un sistema que propone un ambiente flexible y adaptativo para el uso de los servicios, tomando en cuenta cómo un usuario interactúa con los servicios a través de diversos dispositivos (por ejemplo, teléfono inteligente, PC) y qué rol juega ese usuario en el ecosistema lácteo.

3.1. Descripción

El sistema adaptativo propuesto está basado en el ecosistema lácteo de Boyacá, Colombia. El departamento contiene 123 municipios y cubre alrededor de 23000 kilómetros cuadrados. También cuenta con 748,700 cabezas de ganado distribuidas en 52,493 granjas. En términos geográficos, la zona incluye montañas, mesetas y valles. Las montañas usualmente están a más de 1,000 metros sobre el nivel del mar.

El ecosistema lácteo es uno de los elementos fundamentales en la economía del departamento, tanto por la producción de leche como por la fabricación de productos lácteos. Algunos de los procesos del ecosistema presentan ciertos inconvenientes. En este artículo, se analizaron dos de los procesos principales: el proceso de venta de leche por el ganadero y la consulta, por parte del gobierno, de la presencia y contagio de una enfermedad infecciosa bovina.

La venta de leche por parte del ganadero es un proceso que requiere tiempo e información apropiada para conseguir el mejor trato por parte de las plantas procesadoras de lácteos que son las que compran la leche usualmente y las que dictan el precio por litro. Este valor se determina teniendo en cuenta diversos factores como la calidad de la leche, las enfermedades reportadas y la distancia recorrida para recolectar la leche. También consideran información sobre el clima, ya que condiciones de lluvia o heladas pueden afectar la producción de leche y la recolección. Una de las principales funcionalidades propuestas para EiBeLec es proporcionar información sobre las rutas para recolectar leche más beneficiosas tanto para plantas procesadoras de lácteos como para los ganaderos.

Por otro parte, el actor gobierno (gobernador del departamento, alcalde de cada municipio y/o secretarios de agricultura, finanzas y salud, entre otros) requiere reportes de presencia, posibles brotes y predicción de contagio de ciertas enfermedades infecciosas bovinas. EiBeLec debe proporcionar la funcionalidad de recolectar y presentar la información en un área dada. Además, el sistema debe ofrecer diferentes opciones de reporte, dependiendo del rol de cada interesado, de forma que éstos puedan tomar las decisiones más apropiadas para cada situación en particular.

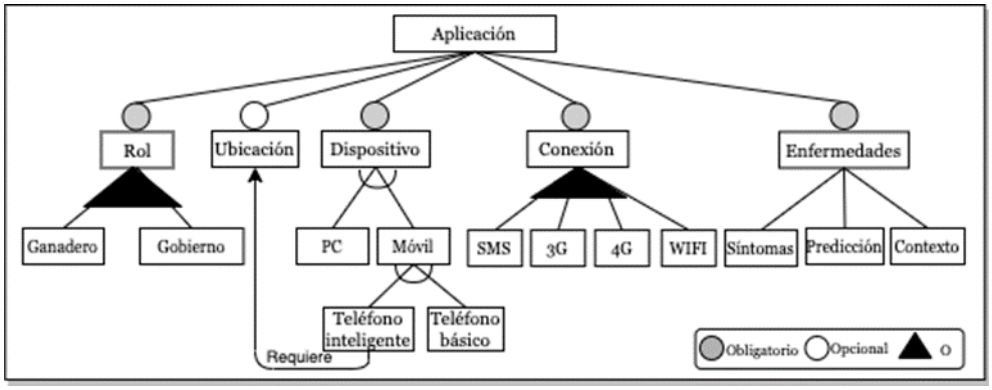


Figura 1 – Diagrama de contexto de características

Los requerimientos, descritos previamente, representaron una serie de configuraciones que pueden variar dependiendo de información del contexto: rol de usuario, ubicación, distancias, enfermedades, etc. Una forma de lidiar con la complejidad de múltiples configuraciones es usar un diagrama de características. En este tipo de diagrama, diversas características de un sistema se presentan en un árbol. Cada característica puede ser obligatoria u opcional. Además, puede haber restricciones entre dichas

características. La semántica detrás de un diagrama de características corresponde a todas las combinaciones de características opcionales posibles que respeten las restricciones. La Figura 1 ilustra las características para EiBeLec. Las características en el diagrama representan las piezas principales de información del contexto que afectan la configuración del sistema.

En primer lugar, el sistema considera el rol que está utilizando entre los dos que se han trabajado en este artículo: el rol gobierno y el rol ganadero. También considerará la ubicación del usuario ya que ésta tiene un impacto en los procesos realizados por el sistema y en la información presentada. El dispositivo usado para acceder al sistema también afecta a la forma en la que el sistema presenta la información. En este caso, se consideraron tanto las capacidades del hardware como la conectividad del dispositivo. Finalmente, está la información sobre las enfermedades, la cual afecta a la mayoría de las funcionalidades presentadas en el sistema.

3.2. Arquitectura

Para construir un sistema adaptativo que pueda representar todas las características definidas en la Figura 1, se propone una arquitectura basada en componentes. Un objetivo de esta arquitectura es la separación de asuntos mediante los componentes que son proveedores de datos y los que son consumidores. Con este tipo de arquitectura, una reconfiguración se puede hacer en dos niveles: (1) reemplazando un componente, o (2) modificando el comportamiento interno de un componente. En cualquier caso, la reconfiguración debe ser un proceso transparente para el resto de los componentes, mientras que los servicios y la información continúan siendo proporcionados en cualquier configuración.

La arquitectura propuesta por EiBeLec incluye diversos componentes para la recolección y el procesamiento de la información y la agregación de contexto, como se ilustra en la Figura 2.

3.3. Componente de contexto

La información inicial sobre enfermedades y usuarios en el ecosistema lácteo está actualmente en archivos físicos y digitales (archivos de texto y hojas de cálculo) (Ballesteros-Ricaurte et al., 2018), y en bases de datos privadas en laboratorios clínicos, grupos de investigación del área veterinaria e instituciones del gobierno. Se definió un componente de contexto que se encarga de añadir esta información y almacenarla en la base de datos de EiBeLec que es utilizada por los componentes propuestos en la arquitectura (Figura 2).

Los tres tipos principales de información de los que se dispone en EiBeLec son:

- Enfermedades: todos los actores del sistema pueden reportar información sobre enfermedades infecciosas bovinas.
- Perfiles de usuario: los roles tienen un impacto en la forma en la que los servicios se proporcionan. Por esta razón, es importante validar y almacenar un perfil para cada actor involucrado en el ecosistema lácteo. Un perfil incluye gustos, preferencias, ubicación (para los usuarios de dispositivos móviles) y relaciones con otros usuarios.

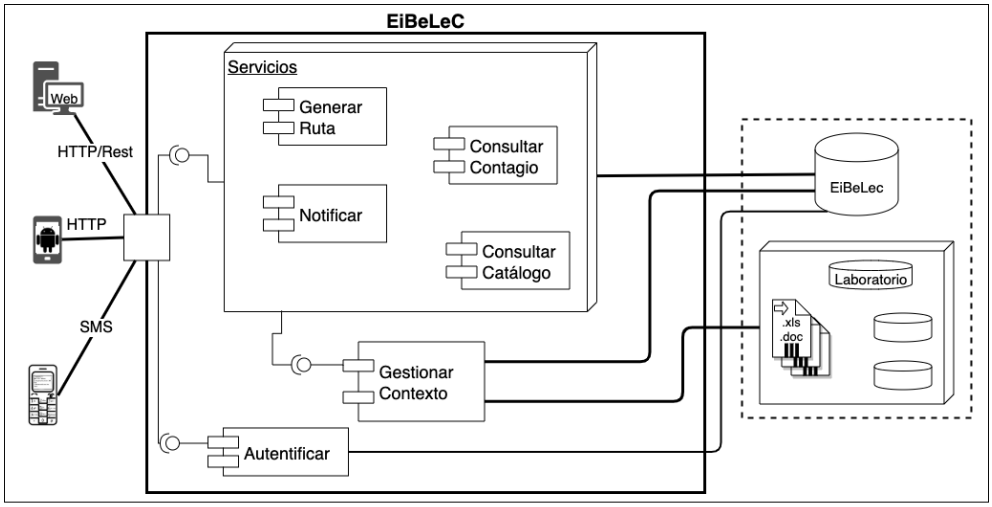


Figura 2 – Arquitectura de EiBeLec

- Dispositivos: para proporcionar una experiencia a la medida, es importante definir y procesar las características del hardware y del software del dispositivo que permite la conexión con el sistema. Actualmente, EiBeLec apoya tres tipos de dispositivos: computadores de escritorio que se conectan con el sistema vía navegador web, dispositivos móviles con sistema operativo Android, y teléfonos celulares no inteligentes que sólo se utilizan para enviar notificaciones de mensajes SMS (Short Message Service).

3.4. Componente de servicios

Este componente está conformado por cuatro (4) servicios que explicaremos a continuación: “Generar ruta”, “Consultar contagio”, “Consultar catálogo” y “Notificar”. De los dos primeros se hará una explicación detallada y de los otros dos sólo se explicará brevemente en que consisten.

3.4.1. Servicio “Generar ruta”

Este servicio organiza la información relacionada con el precio de la leche, la localización de las granjas y de las plantas procesadoras de lácteos en el área, para presentar los resultados de la consulta de las rutas para llevar la leche desde el ganadero hasta la planta procesadora.

En la Figura 3 el servicio “Generar ruta” se ilustra utilizando la notación de Business Process Modeling (BPMN) (Hitpass, Freund, & Rucker, 2011). El proceso representa la interacción entre el ganadero y el sistema. El usuario ganadero ingresa al sistema desde su dispositivo para conocer el precio de la leche que está ofreciendo cada compañía láctea. El sistema le muestra al ganadero el precio de la leche ofrecido por cada planta procesadora e información adicional sobre la ubicación de la planta procesadora, de si le va a recoger la leche en su granja y de cuál es la forma de pago. Considerando esta

información, el ganadero puede seleccionar a qué planta procesadora de lácteos va a vender la leche para obtener el mayor beneficio. Una vez seleccionada la planta procesadora a la que se la va a vender, el sistema muestra al ganadero la ruta para llegar desde su ubicación hasta la sede la planta procesadora.

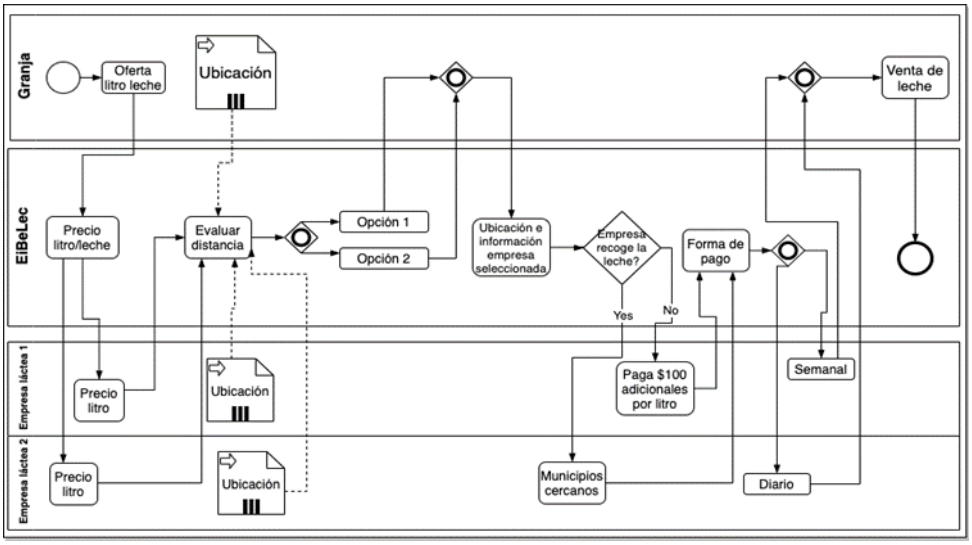


Figura 3 – Proceso de venta de leche por los ganaderos a las plantas procesadoras

La Figura 4 muestra un ejemplo de cómo le es presentada al ganadero la información sobre el componente generar ruta para tres dispositivos: navegador web, teléfono celular no inteligente y Smartphone. En el ambiente web (Figura 4a), se presenta la información sobre el precio de la leche, la ubicación de la compañía, la forma de pago y la ruta desde la granja hasta la planta procesadora de lácteos. En el Smartphone (Figura 4c), adicional a la información básica, el sistema muestra el mapa con la ruta para llegar desde la ubicación actual del usuario hasta la planta procesadora de lácteos. Finalmente, el teléfono celular no inteligente (Figura 4b), sólo permite que se presente un SMS con la información de una planta procesadora de lácteos, el precio de la leche y distancia entre la ubicación del usuario y la planta procesadora. Este es un mensaje que se envía diariamente a los ganaderos registrados y no es consecuencia de una interacción con el sistema.

Las plantas procesadoras de lácteos hacen la compra a través de cooperativas de productores de leche para reducir los costos y tener un control de calidad. Las cooperativas deben llevar a cabo procesos como ubicación de tanques de enfriamiento para la leche en diferentes áreas y crear conciencia entre los ganaderos sobre buenas prácticas de ordeñar. Esto es obligatorio para comprar la leche. Estas cooperativas deben recolectar la leche y llevarla a los tanques de enfriamiento para que un camión pueda recoger la leche y llevarla hasta la planta procesadora. La Figura 5 describe el proceso de recolección de leche realizado por los camiones de las cooperativas. Cada camión debe seguir una ruta y respetar un horario. Adicionalmente, cada camión tiene un número de litros permitidos para transportar.

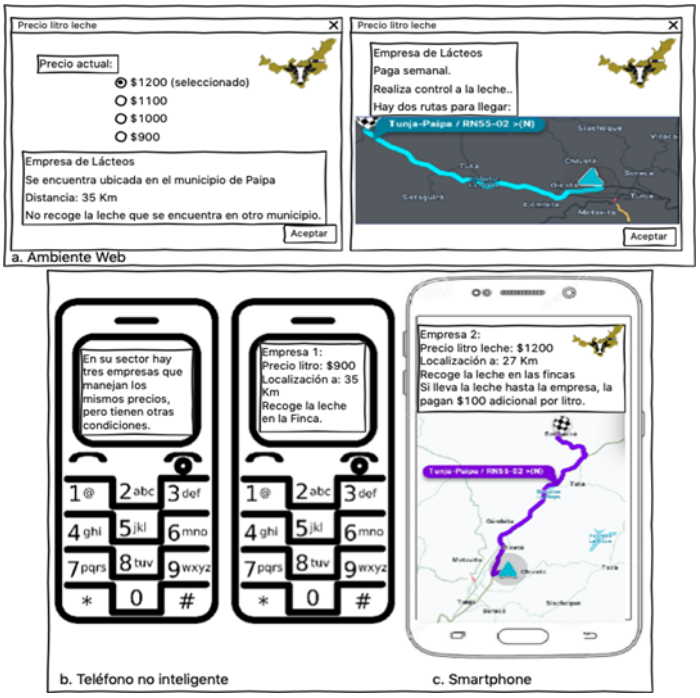


Figura 4 – Resultados de consultar el componente Ruta, usando dos ambientes

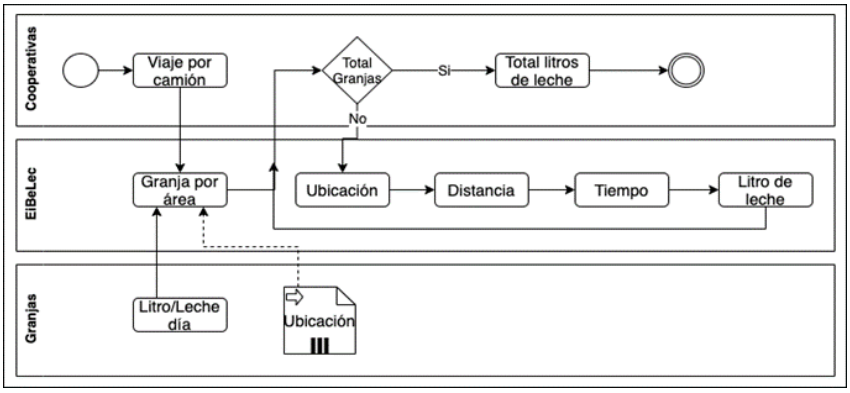


Figura 5 – Proceso de recolección y repartición de leche

El servicio “Generar ruta” entrega información a las cooperativas sobre la ubicación de las granjas, la cantidad de litros/leche, la distancia entre las granjas y la ruta que debe seguirse en el viaje. La Figura 6 muestra el resultado de la consulta hecha por la cooperativa al sistema EiBeLec para conocer la ruta que el camión del tanque debe seguir para recolectar la leche en un sector definido. La información muestra las granjas que deben ser visitadas, la distancia entre cada granja y el precio por litro de leche que se está pagando.

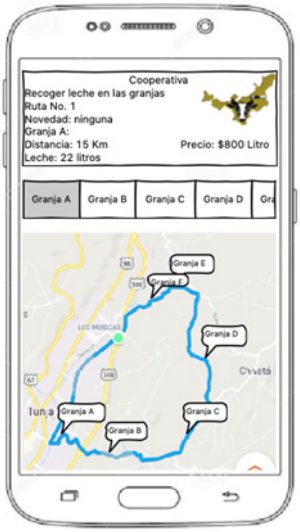


Figura 6 – Servicios de ruta para el conductor de camión en un Smartphone

3.4.2. Servicio "Consultar contagio"

El servicio "Consultar contagio" reúne el modelo predictivo y el modelo epidemiológico para pronosticar la propagación de una enfermedad infecciosa. También entrega los resultados al componente de ruta para validar si hay presencia de una enfermedad o no. Los resultados que se entregan a los usuarios tienen diferentes formatos, dependiendo de las preferencias de cada uno.

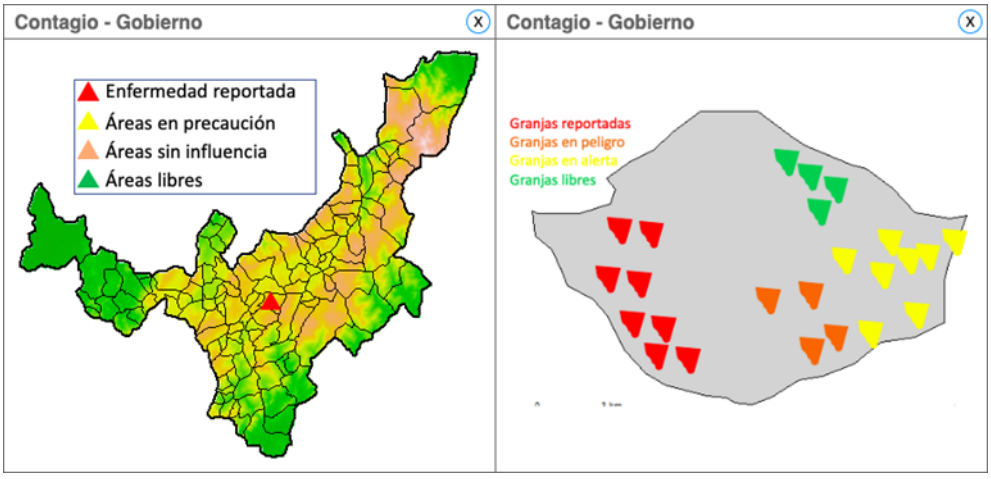


Figura 7 – Resultados esperados del servicio de consultar contagio

Este servicio está basado en la composición y la colaboración de otras unidades de servicio (Peper & Schneider, 2008). Para este escenario, existe una relación entre la

ruta y los servicios de catálogo que proporcionan información sobre: enfermedades reportadas, síntomas, y la relación con las variables contextuales. El componente de contagio consulta a los usuarios sobre la información y es responsable de la organización y coordinación de los recursos, de la información y de presentar la misma al usuario final.

Cuando el servicio "Consultar contagio" es requerido por el actor gobierno para verificar si hay posibilidades de que se difunda la enfermedad, el resultado es un mapa térmico del departamento, indicando dónde se localiza la infección, las áreas en estado de alerta, y las áreas libres de infección. Un ejemplo de este mapa térmico se ilustra en la parte izquierda de la Figura 7. El sistema simula cómo la enfermedad se puede expandir teniendo en cuenta las variables de la enfermedad. Además, si la consulta se hace por parte del gobierno de un municipio, el resultado se limita a ese municipio y no a todo el departamento, como se ilustra en el lado derecho de la Figura 7.

3.4.3. Servicio "Consultar catálogo"

Este servicio valida las consultas que los usuarios pueden hacer, verificando los datos, procesando la información y reportando el resultado a cada usuario.

3.4.4. Servicio "Notificar"

Es el servicio encargado de informar a otros componentes y usuarios sobre eventos (campañas de vacunación, precio de la leche, reporte de brote de enfermedad) o cambios que ocurren (reconfiguración de un componente).

4. Conclusiones

En la literatura existen trabajos de arquitecturas de software adaptativas para resolver problemas en áreas como transporte, turismo, comunicaciones, entre otras. Así mismo, en el campo de la agricultura y ganadería se presentan desarrollos en el campo de epidemiología y granjas (Castillejo et al., 2019; Picault et al., 2017); sin embargo, en el caso particular del ecosistema lácteo, no se encontró un sistema adaptativo que tomara en cuenta los diferentes actores, el contexto, los dispositivos de comunicación y que presentara servicios que ayuden a los usuarios en la toma de decisiones. Por esta razón, se propuso el sistema adaptativo EiBeLec, que está basado en una arquitectura de componentes, con servicios adaptables teniendo en cuenta el perfil de los actores del ecosistema lácteo.

Eibelec cumple las características propuestas para sistemas adaptativos, dado que es un sistema configurable que permite analizar sólo determinadas variables que se quieren tener en cuenta y que permite la opción de incluir nuevos actores con determinadas funcionalidades cuando el contexto lo requiera. Así mismo, posee la facilidad de visualizarse adecuadamente según el dispositivo del usuario debido a que cuenta con la escalabilidad según el dispositivo y disponibilidad de acceso según la capacidad conexión de acceso del mismo. Por último, el contenido que se muestra se adapta al perfil (ganadero o gobierno) del que haga la consulta. En cada elemento el sistema presenta funcionalidades acordes a la toma de decisiones que el usuario requiere, variando principalmente el nivel de detalle o granularidad.

Como trabajo futuro será necesario validar la arquitectura de EiBeLec y los servicios descritos. Además, para el servicio "Consultar contagio" se propone evaluar técnicas de aprendizaje automático para predecir la expansión de las enfermedades infecciosas bovinas. También es necesario enriquecer los servicios basados en el rol de usuario e implementar el sistema en un ambiente controlado para realizar pruebas de adaptabilidad y usabilidad con diferentes usuarios finales que tengan distintos perfiles para observar el comportamiento de EiBeLec para los diversos roles.

Agradecimientos

Esta investigación es financiada por el Gobierno Colombiano a través de la beca 733 de Colciencias, orientado a la Formación de Capital Humano.

Referencias

Ballesteros, J. A., Carrillo, A., Parra, C., & Gómez, J. (2018). Adaptive System to Support Decision-making of Dairy Ecosystem in Boyacá Department. In *Proceedings of the 14th International Conference on Web Information Systems and Technologies*, (pp. 253-260). Seville, Spain. doi: 10.5220/0006947502530260

Barrett, C. L., Bisset, K. R., Eubank, S. G., Feng, X., & Marathe, M. V. (2008). EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, (pp. 1-12). Austin, TX, USA. doi: 10.1145/1413370.1413408

Beckman, R. (2014). ISIS: A Networked-Epidemiology Based Pervasive Web App for Infectious Disease Pandemic Planning and Response. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14)*, (pp. 1847-1856). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/2623330.2623375

Bisset, K., Chen, J., Deodhar, S., Feng, X., Ma, Y. & Marathe, M. (2014). INDEMICS: An interactive high-performance computing framework for data intensive Epidemic Modeling. ACM Trans Model Comput Simul, 24(1). doi: 10.1145/2501602

Bradhurst, R. A., Roche, S. E., East, I. J., Kwan, P., & Garner, M. G. (2016). Improving the computational efficiency of an agent-based spatiotemporal model of livestock disease spread and control. *Environmental Modelling and Software*, 77, 1-12. doi: 10.1016/j.envsoft.2015.11.015

Carr, S., & Roberts, S. (2010). Planning for infectious disease outbreaks: a geographic disease spread, clinic location, and resource allocation simulation. In *Proceedings of the 2010 Winter Simulation Conference* (pp. 2171-2184). Baltimore, MD, USA. doi: 10.1109/WSC.2010.5678858

Carrillo, A., Aragón, F., Cárdenas, J., Cristancho, J., Higuera, M. C., & Marín, D. (2009). Aspectos a considerar para adaptar el contenido y el despliegue de la información. *Avances en Sistemas e Informática*, 6(2), 99-112.

Castillejo, P., Cürüklü, B., Fresco, R., Johansen, G., Bilbao-Arechabala, S., Martínez-Rodríguez, B., ... Santic, M. (2019). The AFarCloud ECSEL Project. In *22nd Euromicro Conference on Digital System Design (DSD)*, (pp. 414-419). Kallithea, Greece. doi: 10.1109/DSD.2019.00066

Damianos, L., Wohlever, S., Kozierok, R., & Ponte, J. (2003). MiTAP: A case study of integrated knowledge discovery tools. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, (p. 10). Big Island, HI, USA. doi: 10.1109/HICSS.2003.1173927

Godsk, T. (2009). Methods and Software Architecture for activity recognition from position data: applied to cow activity recognition. (Tesis doctoral). Ph.D. program in Computer Science, Aarhus University, Denmark.

Hitpass, B., Freund, J. & Rucker, B. (2011). BPMN Manual de referencia y guía práctica. Santiago de Chile. BPM Center.

Lancheros, D. J. (2014). *Sistema adaptativo de apoyo a los procesos de enseñanza-aprendizaje orientados a estudiantes con o sin discapacidad*. (Tesis Doctoral). Doctorado en Ingeniería, Pontificia Universidad Javeriana, Bogotá D. C.

McConnel, C., Lombard, J., Wagner, B., Kopral, C., & Garry, F. (2015). Herd factors associated with dairy cow mortality. *Animal*, 9(8), 1397-1403. doi: 10.1017/S1751731115000385

Peeler, E. J., & Otte, M. J. (2016). Epidemiology and Economics Support Decisions about Freedom from Aquatic Animal Disease. *Transboundary and Emerging Diseases*, 63(3), 266-277. doi: 10.1111/tbed.12278

Peper, C., & Schneider, D. (2008). Component engineering for adaptive ad-hoc systems. In *Proceedings of the 2008 International workshop on Software engineering for adaptive and self-managing systems*, (pp. 49-56). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/1370018.1370028

Picault, S., Huang, Y. L., Sicard, V., Beaudeau, F., & Ezanno, P. (2017). A Multi-Level Multi-Agent Simulation Framework in Animal Epidemiology. In Demazeau, Y., Davidsson, P., Bajo, J., Vale, Z. (Eds.). In *Advances in Practical Applications of Cyber-Physical Multi-Agent Systems: The PAAMS Collection*. Springer. doi: 10.1007/978-3-319-59930-4_17

Pulido, M., González, W., Bayona, H., & Chavarro, G. (2017). Determinación de Leucosis enzootica bovina mediante las claves Hematológicas de Göttigen y Elisa en Boyacá. *Revista de la Facultad de Ciencias Veterinarias*, 58(1), 10-16.

Richards, K. K., Hazelton, M. L., Stevenson, M. A., Lockhart, C. Y., Pinto, J., & Nguyen, L. (2014). Using exceedance probabilities to detect anomalies in routinely recorded animal health data, with particular reference to foot-and-mouth disease in Vietnam. *Spatial and Spatio-Temporal Epidemiology*, 11, 125-133. doi: 10.1016/j.sste.2014.08.002

Siettos, C. I., & Russo, L. (2013). Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4), 295-306. doi: 10.4161/viru.24041

Singer, A., Salman, M., & Thulke, H. H. (2011). Reviewing model application to support animal health decision making. *Preventive Veterinary Medicine*, 99(1), 60-67. doi: 10.1016/j.prevetmed.2011.01.004

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 25/08/2020
Aceitação/Acceptance: 22/10/2020

Análisis de exactitud de reconocimiento gestual aplicando SVM y k-NN en señales EMG

Gonzalo Pomboza-Junez[1], Juan A. Holgado-Terriza[2]

wpomboza@unach.edu.ec; jholgado@ugr.es

[1] Carrera de Informática Aplicada a la Educación, Universidad Nacional de Chimborazo, Riobamba, Ecuador.

[2] Departamento de Lenguajes y Sistemas Informáticos, Universidad de Granada, Granada, España.

Pages: 15-28

Resumen: El uso del gesto, sea manual, facial y postural, establece una forma de comunicación hombre-máquina que aún debe ser estudiada profunda y ampliamente. Esta investigación establece la factibilidad del reconocimiento de los gestos o posturas de la mano usando un wearable (brazalete MYO) para la captura las señales electromiográficas (EMG) producidas por los músculos del antebrazo al formar un gesto. La información es capturada usando electrodos de superficie para luego, mediante el uso de clasificadores, lograr su reconocimiento. Participaron veintiún voluntarios y se analizaron doscientos setenta y tres gestos. Se evaluaron dos clasificadores, a saber, k-Nearest Neighbor (k-NN) y Support Vector Machines (SVM). El clasificador basado en SVM con núcleos Radial (93,03%) y Polinomial (97,81%) logró los mejores resultados. Además, se implementó un diccionario gestual que puede ser usado en investigaciones similares especialmente para tareas de control e interacción hombre-máquina.

Palabras-clave: Señales EMG; Interfaz de control gestual; reconocimiento gestual; Clasificador SVM; diccionario gestual.

Accuracy analysis of gesture recognition applying SVM and k-NN on EMG signals

Abstract: The use of the gesture using hands, face, and body positions, establish a form of man-machine communication that has yet to be studied deeply and widely. The purpose of this paper is to illustrate the feasibility of gestural recognition performed with the hand by using a wearable (MYO), which captures the electromyographic (EMG) signals produced by the forearm´s muscles, precisely by forming and maintaining the gesture. The EMG signals are captured using surface electrodes and applied in classifier algorithms to achieve gesture recognition. Twenty-one volunteers participated in this research and two hundred and seventy-three gestures were analyzed. Two classifiers were evaluated, namely k-Nearest Neighbor (k-NN) and Support Vector Machines (SVM). SVM-based classifier with Polynomial kernel (97.81%) and Radial kernel (93.03) achieved the best results. A gestural dictionary of hand poses was implemented that can be used for similar research, especially in human-machine control tasks interaction.

Keywords: EMG signals; gesture control interface; gestural recognition; SVM classifier; gestural dictionary.

1. Introducción

La forma más sencilla de lograr la interacción con la máquina es, sin duda, aquella que refleja el deseo o la intención del usuario en forma rápida, sencilla y confiable. A este tipo de interacción se la conoce Interfaz Natural de Usuario (INU) en la cual la intuición juega un papel muy importante para ejecutar una tarea o una acción. Esto fue demostrado por Ibañez et al. (Ibañez, Soria, Teyseyre, & Campo, 2014), que utilizaron los movimientos intuitivos del usuario (manos, brazos, piernas, posiciones) como comandos para controlar un juego. De la misma manera, según Mitra et al. (Mitra & Acharya, 2007), las manos son el mejor recurso de comunicación que posee el hombre. La riqueza gestual en el ser humano es muy amplia, al punto que existen lenguajes gestuales basados en signos, como el ASL (American Sign Language), que utilizan principalmente las manos. Además, desde este punto de vista, el gesto permite a los usuarios expresar una idea o un concepto (Lian, Hu, & Wang, 2014), que puede ser convertido en una acción o un comando.

Se han propuesto muchos métodos para lograr el reconocimiento del gesto, incluidas soluciones ópticas (usando cámaras) (Hasan & Abdul-Kareem, 2014), mediante el contacto (por ejemplo, pantalla táctil) (Poppinga, Sahami Shirazi, Henze, Heuten, & Boll, 2014), y dispositivos fijados o implantados en el cuerpo (por ejemplo, electrodos EMG) (C. Chen, Jafari, & Kehtarnavaz, 2015). Por su complejidad, estas soluciones han tenido que superar varios problemas (óptica con problemas de oclusión, contacto con problemas de captura-sensibilidad, dispositivos fijados al cuerpo con problemas de identificación y reconocimiento) (Hsieh & Liou, 2015) para convertirse en métodos operativos, fiables y portables (Ionescu et al., 2014).

El uso de señales electromiográficas (EMG) aplicadas al reconocimiento de movimientos o gestos (faciales, corporales o manuales) no es nuevo, y las investigaciones han hecho notables avances tanto en su captura como en su aplicación un ejemplo de ello es la prótesis de miembros superiores e inferiores, los mecanismos de ayuda a personas con discapacidades motoras, etc. La señal EMG provoca la contracción o relajación de un músculo produciendo, en consecuencia, un movimiento o gesto. El movimiento muscular genera patrones los cuales definen su comportamiento o permiten su identificación. Para encontrar estos patrones, los algoritmos de clasificación son necesarios e indispensables para lograr un reconocimiento gestual preciso. La señal EMG también contiene características aleatorias que deben ser procesadas antes de aplicar un clasificador. Existe un consenso en la comunidad científica para aplicar métodos de extracción de características en el dominio del tiempo (TD) y en el dominio de la frecuencia (FD). Kakoty et al. (Kakoty, Hazarika, & Gan, 2016) evalúan las características del tiempo, la frecuencia, la frecuencia del tiempo y el análisis de los componentes principales (PCA) del dominio tiempo-frecuencia. Mientras que Menon et al. (Menon et al., 2017) estudiaron los efectos de la información temporal y espacial proporcionada al clasificador, fuera de línea, y analizaron sus interdependencias. En cualquier caso, la aplicación de las características, cualquiera sea el método de extracción, es fundamental cuando se intenta lograr la máxima precisión de la clasificación. A continuación, se indican algunos

de los métodos más utilizados para clasificar y analizar las señales EMG: Red neuronal convolucional (CNN) (Zhai, Jelfs, Chan, & Tin, 2017), Redes neuronales artificiales (RNA), Análisis discriminante lineal (LDA) (Menon et al., 2017; Zhai et al., 2017), Máquinas vectoriales de apoyo (SVM)-RBF (Kakoty et al., 2016), Vecino más cercano (k-NN) (J. Kim, Mastnik, & André, 2008), entre otros. Estos dos últimos algoritmos, k-NN y SVM, han sido utilizados con resultados muy alentadores a la hora de clasificar las señales EMG (Singh, Arora, Shukla, & Mittal, 2015).

Por otro lado, varias investigaciones discuten el uso de dos tipos de electrodos (sEMG, superficial e iEMG, intramuscular) para la captura efectiva de señales EMG (Yugan, Gobee, & Vickneswari, 2014). Las soluciones con iEMG son muy escasas ya que se consideran demasiado invasivas y sólo se recomiendan en casos obligatorios, al contrario de lo que sucede con sEMG. Los electrodos sEMG deben estar correctamente ubicados para garantizar la lectura de la señal que, por lo general, tiene una resolución ultra baja (μV-mV) (Al-Timemy, Bugmann, Escudero, & Outram, 2013). Otros estudios muestran que es posible controlar dispositivos y máquinas, utilizando un pequeño número de electrodos sEMG integrados al usuario. En este ámbito, el brazalete MYO®, incorpora electrodos de tipo sEMG y permite ser usado en el usuario sin mayor complicación. Este dispositivo no requiere ningún accesorio especial y es adaptable a diferentes perfiles de usuarios. El brazalete lee las señales EMG de los usuarios y envía los datos al equipo receptor (empotrado o computador) usando protocolo Bluetooth con características de bajo consumo (BLE). El brazalete es invariable a las condiciones de iluminación, a oclusiones y es independiente de la relación usuario-posición (W. G. Pomboza-Junez & Holgado-Terriza, 2016). Los gestos utilizados en este estudio se basan en posturas estáticas y se describen en la Sección 2.

Este trabajo se compone de las siguientes secciones: La sección 2 determina los antecedentes conceptuales mientras la sección 3 describe los materiales y métodos. La sección 4 muestra nuestros resultados y consideraciones. Por último, la sección 5 presenta las conclusiones y la sección 6 los trabajos futuros.

2. Antecedentes conceptuales

2.1. Diccionario Gestual

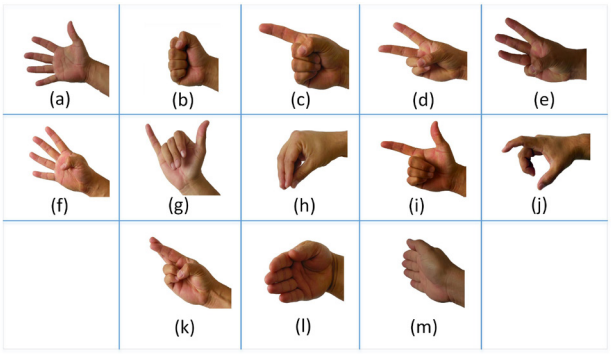


Figura 1 – Gestos de la mano analizados en esta investigación.

El diccionario gestual que en esta investigación se analiza, usa una sola mano y ofrece varias alternativas si pensamos en una interfaz basada en gestos de una mano. Utilizamos gestos que son fáciles de hacer y aprender para los usuarios (Nielsen, Störring, Moeslund, & Granum, 2004), como se muestra en la Figura 1.

2.2. Algoritmos de clasificación.

SVM

Las Máquinas de Vectores de Soporte (SVM) ofrecen un nuevo y prometedor método para clasificar tanto datos lineales como no lineales (Figura 2). SVM puede utilizarse tanto para la predicción como para la clasificación. Para más detalles, véase (Wallraven et al., 2003). Los clasificadores SVM encuentran hiperplanos que dividen las clases; el mejor hiperplano es el que tiene el mayor margen entre las dos o más clases. El hiperplano correspondiente puede definirse como:

$$w^T \cdot x + b = 0 \tag{1}$$

dónde w es un vector de peso, x es un vector de datos y b es un término separado que ofrece una mayor libertad para encontrar el hiperplano óptimo para clasificar los datos. Al aplicar núcleos el clasificador lineal puede servir como un clasificador no lineal. Los métodos de los núcleos de los SVMs muestran que, los SVMs funcionan bien para espacios de altas dimensiones, ya que encuentran el hiperplano con el mayor margen para clasificar diferentes conjuntos de datos. Hay tres tipos de funciones núcleo involucradas: funciones polinómicas, de base radial y sigmoidales.

$$k(\overline{X}, \overline{Y}) = (\overline{X} \cdot \overline{Y} + 1)^p \tag{2}$$

$$k(\overline{X}, \overline{Y}) = \exp\left(-\frac{\left\| \overline{X} - \overline{Y} \right\|^2}{2\sigma^2} \right) \tag{3}$$

$$k(\overline{X}, \overline{Y}) = \tanh\left(\gamma(\overline{X} \cdot \overline{Y}) + \delta \right) \tag{4}$$

La ecuación (3) ofrece un clasificador que es un polinomio de grado p, la ecuación (4) ofrece un clasificador en base radial gaussiana, y la ecuación (5) es una función de la activación neuronal sigmoidal, específicamente del tipo perceptrón (para ciertos valores de δ y γ).

k-NN

El clasificador k-NN encuentra un grupo de objetos k en el conjunto de entrenamiento los cuales están más cerca del objeto de prueba (Figura 3) y se basa en almacenar los

vectores característicos y las etiquetas de las clases de los objetos de entrenamiento. El vecino más cercano da la mejor clasificación (Wu et al., 2008). k-NN localiza las instancias k más cercanas a la instancia de consulta y determina su clase identificando la etiqueta de clase más frecuente (Kotsiantis, 2007). Para determinar la distancia relativa entre las instancias más significativas k-NN utiliza una métrica de distancia. Las más significativas son Mikowsky, Manhattan, Chebychev, Euclidiano, Canberra, y la Correlación de Rango de Kendall. En este trabajo, usamos la distancia euclidiana para facilitar la implementación y la velocidad de cálculo que se expresa en la ecuación (6). Para más detalles, puede revisar la encuesta realizada por (Wettschereck, Aha, & Mohri, 1997).

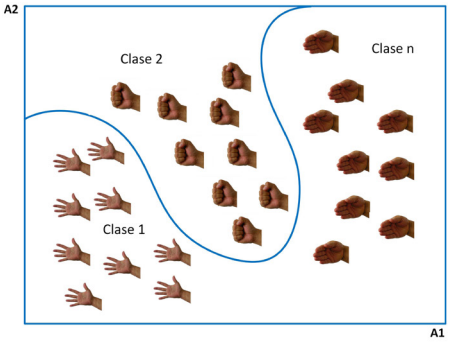


Figura 2 – Clasificador SVM con base Sigmoidal.

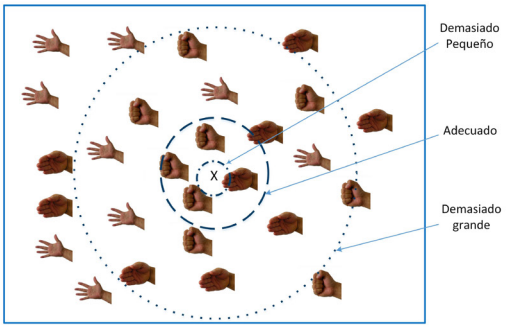


Figura 3 – Clasificador k-NN con valores pequeños, medianos y grandes de k.

$$Euclideo: D(x,y) = \left(\sum_{i=1}^{m} | x_i - y_i |^2 \right)^{1/2} \quad (5)$$

k es un factor clave en el algoritmo k-NN, ya que no debe ser tan pequeño que se vea afectado por el ruido, ni tan grande que el vecino más cercano pueda influir en él (Cover & Hart, 1967; Kim et al., 2008; Parada & Melia-Segui, 2017). Varias cuestiones clave afectan al rendimiento de k-NN según la elección de k (Figura 3).

Selección de Características

En este trabajo se ha usado la extracción de características en el dominio del tiempo, implícitas en la señal. Estos incluyen los resultados de la Rectificación de Onda Completa de una señal EMG (FW-REMG), y, el EMG Integrado (IEMG), las cuales se muestran en la Figura 4. Además, estos valores proporcionan la información necesaria sobre el cambio que experimenta la señal del EMG a lo largo del tiempo.

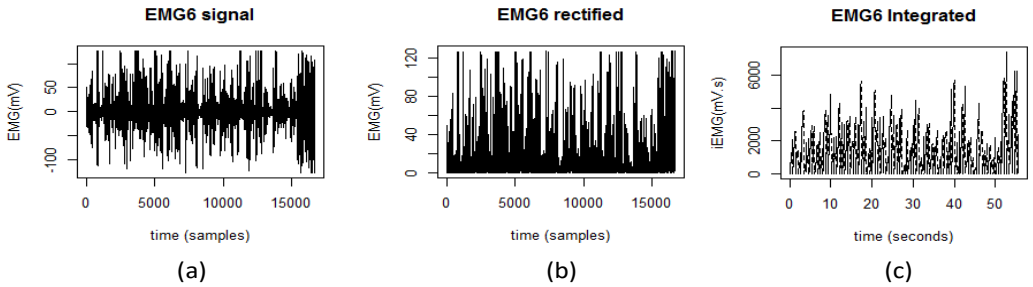


Figura 4 – Ejemplo del canal seis de lectura del gesto (i). (a) RAW. (b) Rectificación de onda completa a la señal. (c) Señal EMG con característica de integración.

Rectificación: Se prefiere el FW-REMG porque se conserva toda la energía de la señal para el análisis (Kaur, Mathur, Bhatia, & Verma, 2015). A veces se considera un paso intermedio antes de la fase de integración en la señal. La rectificación de onda completa se lleva a cabo convirtiendo todas las amplitudes negativas en amplitudes positivas, los picos negativos se "suben" al lado positivo del eje del tiempo, en otras palabras, es lo mismo que tomar el valor absoluto de la señal EMG. Se define como:

$$x_3(t) = | x_2(t) | \tag{6}$$

dónde, $x_2(t) =$ Señal filtrada del sensor, y $x_3(t) =$ Señal rectificada.

Integración de EMG: IEMG es una característica de la señal EMG que puede ser calculada tomando la suma del valor absoluto de la señal EMG digitalizada. Está definido por:

$$x_i(t) = \int_{t_1}^{t_2} x_3(t)dt \tag{8}$$

dónde, $x_i(t) =$ Señal integrada, y $x_3(t) =$ Señal rectificada.

La extracción de características nos permitió extraer la información útil que está oculta en la señal de EMG de la superficie y eliminar la parte no deseada y las interferencias. Se pueden incluir otros análisis en el dominio del tiempo y la frecuencia (por ejemplo, el cuadrado medio de la raíz, la frecuencia media, la amplitud máxima, la frecuencia media, etc.). Sin embargo, hemos intentado tomar las características más significativas de la señal EMG que no reducen la potencia de procesamiento del sistema incorporado, asegurando al mismo tiempo la correcta identificación del gesto basada en las características de las señales EMG.

3. Materiales y métodos

3.1. Captura de los datos

Participantes

Veintiún voluntarios participaron en los experimentos de captura de datos. Las edades oscilaban entre 20 y 50 años (Media=33,44; SD=8,7) (6 mujeres y 15 hombres). Todos ellos eran propietarios de un teléfono inteligente y manifestaron tener experiencia con las utilidades de "manos libres". Además, todos los voluntarios sabían cómo usar la computadora y todos los hombres indicaron que tenían experiencia previa en el uso de videojuegos.

Sensor

El brazalete MYO® (Figura 5.a) fue seleccionado para capturar las señales EMG que se producen en el antebrazo del usuario al realizar un gesto con la mano. El brazalete viene equipado con un microprocesador de tipo Cortex ARM-M4 de bajo consumo y compuesto por un conjunto de 8 sensores de electromiografía de superficie (sEMG) y un sensor inercial (IMU, unidad de medición inercial). Cada sensor sEMG se agrupa en barras paralelas que han sido colocadas radialmente en el brazalete, describiendo una circunferencia, y unidos por una banda flexible, como se muestra en la Figura 5.b. La frecuencia de muestreo, según el fabricante, es de 200 Hz para los sensores sEMG y de 50 Hz para la IMU de 9 ejes.

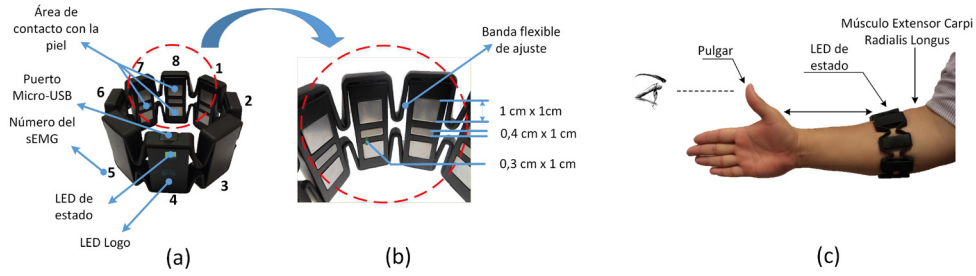


Figura 5 – (a) MYO y componentes. (b) Dimensiones de los sEMG. (c) Alineación del brazalete en el antebrazo del usuario.

Protocolo experimental.

El área de pruebas consistía en un gran escritorio y una cómoda silla con respaldo. Los voluntarios colocaban el brazalete en el antebrazo de la mano derecha (no existieron voluntarios zurdos). El procedimiento para la toma de muestras fue el siguiente: el brazalete se alineaba en el antebrazo del voluntario tomando como referencia el pulgar, con la palma vertical al piso, y trazando visualmente una línea recta entre aquel y el led de estado del brazalete (ver figura 5.c). Esta forma de alineación del brazalete evitó la pérdida de calibración al pasar de un usuario a otro. Además, permitió mantener la configuración por defecto en el brazalete.

El investigador explicaba brevemente (de forma verbal y visual) al voluntario el gesto que debía realizar y mantener, antes de cada prueba, con el objetivo de minimizar datos inconsistentes o no pertenecientes al gesto. Cada prueba tuvo una duración de 10 s y era controlada por software. El inicio lo daba el investigador, para cada gesto, coordinando la indicación verbal con el inicio de la prueba. Durante este tiempo, el sistema registraba los datos generados. Se controlaba siempre la correcta ejecución del gesto. Una señal auditiva indicaba el final de cada prueba. Los datos generados por el voluntario, correspondientes a las señales EMG del gesto, eran capturados por los electrodos sEMG y transferidos, por Bluetooth, a un empotrado. El software de captura fue escrito en lenguaje C y los datos almacenados en un archivo. El brazalete no presentó problemas durante las pruebas.

3.2. Procesamiento de datos y experimentos

El experimento, por cada voluntario, alcanzó un total de 150 s. Se analizaron 273 gestos y un volumen aproximado de datos que alcanzó las 581.000 muestras (entre 28.000 y 30.000 muestras por voluntario). El software, creado para las pruebas, cerraba el archivo de datos y mostraba automáticamente, en pantalla, el siguiente gesto a ser realizado por el voluntario, en un ciclo repetitivo hasta que fuesen todos ejecutados. La figura 6, muestra gráficamente una parte de los datos recopilados del gesto (i) por cada uno de los 8 canales.

Se utilizó el procesamiento fuera de línea para evaluar el rendimiento de la clasificación. Para el procesamiento numérico se utilizó R (Project for Statistical Computing). Los clasificadores SVM y k-NN se construyeron utilizando el paquete e1071, entre otros disponibles en R, para el procesamiento de datos y señales.

Procesamiento y experimentos numéricos

Se utilizaron 8 canales y se analizaron todas las lecturas de los electrodos sEMG; por lo tanto, cada electrodo entregaba una lectura asociada a un músculo, en ese lugar específico pues, todos los músculos, en mayor o menor grado, se encuentran relacionados con el gesto (figura 6).

Los resultados se obtuvieron después de la separación del conjunto de datos para el entrenamiento y la verificación (70-30 para SVM y 80-20 para k-NN). Se investigaron combinaciones de técnicas de reducción de características y clasificación para lograr un máximo rendimiento. Los clasificadores se entrenaron utilizando un conjunto de entrenamiento, mientras que se utilizó un conjunto de validación para optimizar los parámetros de cada método.

En SVM se utilizó el procedimiento “uno contra uno” para realizar la clasificación. Los datos fueron procesados usando SVM con funciones de núcleo tipo: Base Radial Gaussiana (RBF), Polinomial y Sigmoidal. El núcleo requiere dos parámetros: C y γ (gamma). Para afinar estos parámetros y encontrar aquellos que permitan una mejor generalización, hemos utilizado una búsqueda por intervalos en C y γ a través de la validación cruzada (n=10). Los parámetros se seleccionaron sobre la base de varios valores de C [0.0001,0.001,0.01,0.1,1,10,100,1000,10000] y γ [0.00125,0.0625,0

.0125,0.125,0.25,0.5,0.625,0.75]. El mejor parámetro se usó para optimizar a C en el conjunto de entrenamiento solamente. La tabla 1 muestra la precisión alcanzada por cada núcleo en SVM.

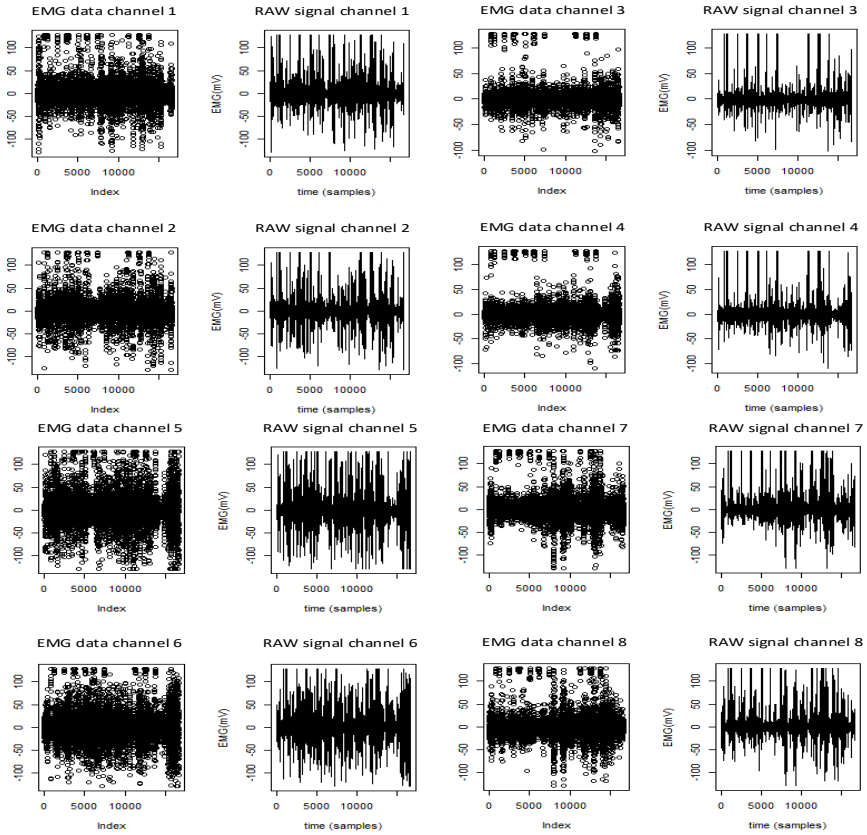


Figura 6 – Muestras de los 8 canales de las señales generadas al realizar el gesto (i). El sensor fue ubicado en el brazo del usuario de acuerdo con lo indicado.

En k-NN el conjunto de datos se clasificó utilizando k {1,3,5,...,27}. Los mejores resultados se obtuvieron para k =7. Este algoritmo mira al vecino más cercano, usando la distancia euclidiana de la ecuación 6. La Tabla 1 muestra la precisión para cada valor de k.

4. Resultados y discusión

En esta investigación, se estudió un amplio conjunto de datos, a los cuales se agregaron características de rectificación e integración, como resultado del análisis de la señal EMG. Se obtuvo una precisión de 93,03% para SVM con núcleo RBF. Fue notorio que, al experimentar con un núcleo polinomial, SVM alcanzó una precisión de 97,81%.

Es necesario indicar que, los resultados para SVM con núcleo sigmoidal, se mantuvieron bajos e incluso por debajo de los inicialmente esperados (apenas alcanzaron el 12,62%). La precisión de SVM con núcleos RBF y polinomial fue alta, con una diferencia entre ellos de 4,78. Por otro lado, la precisión alcanzada por el clasificador k-NN fue de 70,24%. Esta precisión, aunque baja, permite una rápida implementación por lo simple de su algoritmo. Además, no requiere un tiempo de aprendizaje significativo. Sin embargo, si el objetivo es la alta precisión de reconocimiento gestual, los núcleos RBF y polinómico son la mejor alternativa en SVM para reconocimiento gestual. Las matrices de confusión de SVM polinomial (figura 7.a) y k-NN (figura 7.b) se muestran en la figura 7.

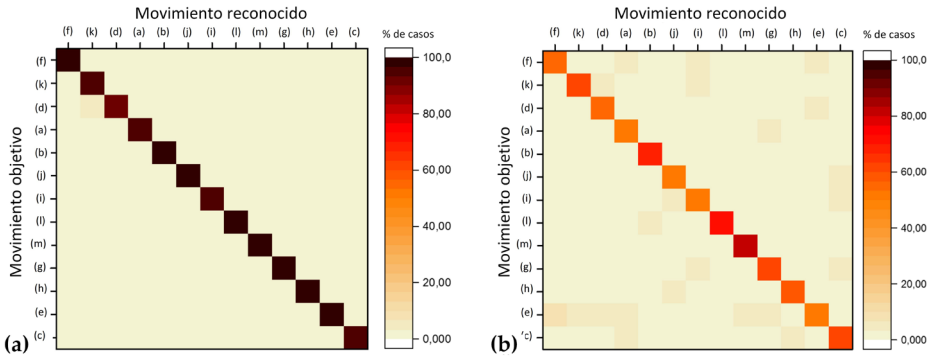


Figura 7 – Matrices de confusión. (a) Matriz algoritmo SVM con núcleo polinómico. (b) Matriz algoritmo k-NN. Ambas matrices muestran la distribución de errores para las lecturas con todos los canales EMG. Los resultados de confusión en la diagonal son las tasas de clasificación correctas mientras que los resultados fuera de la línea diagonal son los errores.

Para SVM, el mejor par (C, γ) estaba en el rango establecido en la Sección 2.2, con una tasa de error de 0,0697 para el núcleo RBF, con porcentajes de exactitud calculados para cada gesto, como se muestra en la Tabla 1. El núcleo polinomial alcanzó una tasa de error de 0,0219 con los mismos parámetros. La Tabla 1, muestra también los niveles de precisión para cada uno de los núcleos probados y el grado de reconocimiento gestual. Los niveles de precisión son aceptables, pero podrían mejorarse eliminando los gestos con lecturas muy similares, por ejemplo, (f), (k), (d), (a), (i), (j), (g) y (h).

Para k-NN, se probaron varios valores de k (k {1,3,...,27}), logrando la mejor precisión para k=7 utilizando el mismo conjunto de datos. El algoritmo k-NN alcanzó 70,24% de precisión en la clasificación. La tasa de error alcanzó 0,2976, con porcentajes de exactitud calculados para cada gesto, como se muestra en la Tabla 1. En este sentido, los niveles de precisión son aceptables, ya que todos superan el 80% excepto el gesto (e). Esto podría mejorarse eliminando los gestos con lecturas muy similares, por ejemplo, el gesto (a) y (f), (e) y (f), (d) y (k), (d) y (e), (h) y (j). La similitud se produce en la señal generada por los músculos, y no necesariamente en la representación espacial del gesto (Figura 1).

En resumen, los dos métodos de aprendizaje automático utilizados, SVM y k-NN, aplicados al reconocimiento gestual han demostrado ser adecuados. Sin embargo, la precisión de clasificación demostrada por SVM con respecto a k-NN fue mucho mayor.

En el primer caso, SVM con núcleo RBF logró una diferencia de 22,79% con respecto k-NN, y en el segundo, con núcleo polinomial logró diferencia aún mayor con respecto a k-NN (27,57%). SVM con núcleos RBF y polinomial logra la mayor precisión, mejor rendimiento y mejores resultados individuales y globales.

Biblioteca gestual		Nivel de Exactitud de algoritmos			
Gesto		k-NN	SVM		
			RBF	Polinomial	Sigmoidal
(a)	Mano abierta	81.37	95.94	98.32	50.88
(b)	Mano cerrada	91.77	97.77	99.17	55.80
(c)	Uno	82.82	95.54	98.41	49.87
(d)	Dos	80.77	94.53	97.09	50.77
(e)	Tres	76.25	95.24	99.10	50.91
(f)	Cuatro	80.5	95.62	99.20	53.86
(g)	Seis	84.76	96.71	99.13	54.23
(h)	Siete	83.66	96.93	99.48	54.48
(i)	Ocho	83.24	94.94	98.02	47.64
(j)	Nueve	83.5	97.00	99.60	50.78
(k)	Diez	83.01	95.23	98.44	52.48
(l)	Palma dentro	89.11	97.19	99.04	53.04
(m)	Palma fuera	89.92	98.25	99.56	59.87

Tabla 1 – Gestos y niveles de exactitud por cada gesto

5. Conclusiones

Esta investigación muestra el alto nivel de precisión que el algoritmo de clasificación SVM con núcleo RBF y polinomial es capaz de lograr en comparación con el algoritmo k-NN, ambos aplicados al reconocimiento gestual. SVM logra clasificar correctamente todos los gestos propuestos, en la biblioteca gestual de experimentación, a partir de las señales EMG obtenidas de los voluntarios al formar y mantener cada gesto. También hemos encontrado que, al agregar características de rectificación de onda completa (FW-REMG) e integración (IEMG) de la señal EMG los resultados de clasificación y reconocimiento son muy aceptables. Finalmente, se determinó que, la precisión del clasificador SVM (97,81%) con núcleo polinomial es mayor que la alcanzada por k-NN (70,24%) usando, en ambos casos, señales EMG capturadas con electrodos EMG de superficie incorporados en el brazalete MYO.

6. Trabajos futuros

Hemos pensado en la implementación de una interfaz de reconocimiento gestual en viviendas, como parte de un sistema integral de control domótico. Implementar algoritmos de reconocimiento con Deep Learning. Implementar algoritmos que logren

introducir el gesto en el control de sistemas de seguridad y monitoreo. En el campo educativo, las posibilidades son mucho más amplias, se podría lograr un aprendizaje significativo mucho más amplio trabajando con niños de educación básica e inicial a través del gesto.

Referencias

Al-Timemy, A. H., Bugmann, G., Escudero, J., & Outram, N. (2013). Classification of Finger Movements for the Dexterous Hand Prosthesis Control With Surface Electromyography. IEEE Journal of Biomedical and Health Informatics, 17(3), 608–618. https://doi.org/10.1109/JBHI.2013.2249590

Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). Improving Human Action Recognition Using Fusion of Depth Camera and Inertial Sensors. IEEE Transactions on Human-Machine Systems, 45(1), 51–61. https://doi.org/10.1109/THMS.2014.2362520

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Hasan, H., & Abdul-Kareem, S. (2014). Static hand gesture recognition using neural networks. Artificial Intelligence Review, 41(2), 147–181. https://doi.org/10.1007/s10462-011-9303-1

Hsieh, C.-C., & Liou, D.-H. (2015). Novel Haar features for real-time hand gesture recognition using SVM. Journal of Real-Time Image Processing, 10(2), 357–370. https://doi.org/10.1007/s11554-012-0295-0

Ibañez, R., Soria, Á., Teyseyre, A., & Campo, M. (2014). Easy gesture recognition for Kinect. Advances in Engineering Software, 76, 171–180. https://doi.org/10.1016/j.advengsoft.2014.07.005

Ionescu, B., Suse, V., Gadea, C., Solomon, B., Ionescu, D., Islam, S., & Cordea, M. (2014). Using a NIR camera for car gesture control. IEEE Latin America Transactions, 12(3), 520–523. https://doi.org/10.1109/TLA.2014.6827882

Kakoty, N. M., Hazarika, S. M., & Gan, J. Q. (2016). EMG Feature Set Selection Through Linear Relationship for Grasp Recognition. Journal of Medical and Biological Engineering, 36(6), 883–890. https://doi.org/10.1007/s40846-016-0188-y

Kaur, M., Mathur, S., Bhatia, D., & Verma, S. (2015). siGnum: Graphical user interface for EMG signal analysis. Journal of Medical Engineering & Technology, 39(1), 19–25. https://doi.org/10.3109/03091902.2014.973615

Kim, J., Mastnik, S., & André, E. (2008). EMG-based hand gesture recognition for realtime biosignal interfacing. In Proceedings of the 13th international conference on Intelligent user interfaces - IUI '08 (Vol. 39, p. 30). New York, New York, USA: ACM Press. https://doi.org/10.1145/1378773.1378778

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. Informatica, 31, 249–268. https://doi.org/10.1115/1.1559160

Lian, S., Hu, W., & Wang, K. (2014). Automatic user state recognition for hand gesture based low-cost television control system. IEEE Transactions on Consumer Electronics, 60(1), 107–115. https://doi.org/10.1109/TCE.2014.6780932

Menon, R., Di Caterina, G., Lakany, H., Petropoulakis, L., Conway, B., & Soraghan, J. (2017). Study on interaction between temporal and spatial information in classification of EMG signals in myoelectric prostheses. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 4320(c), 1–1. https://doi.org/10.1109/TNSRE.2017.2687761

Mitra, S., & Acharya, T. (2007). Gesture Recognition: A Survey. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 37(3), 311–324. https://doi.org/10.1109/TSMCC.2007.893280

Naik, G. R., & Nguyen, H. T. (2015). Nonnegative matrix factorization for the identification of EMG finger movements: Evaluation using matrix analysis. IEEE Journal of Biomedical and Health Informatics, 19(2), 478–485. https://doi.org/10.1109/JBHI.2014.2326660

Nielsen, M., Störring, M., Moeslund, T. B., & Granum, E. (2004). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In Gesture-Based Communication in Human-Computer Interaction (pp. 409–420). https://doi.org/10.1007/978-3-540-24598-8_38

Parada, R., & Melia-Segui, J. (2017). Gesture Detection Using Passive RFID Tags to Enable People-Centric IoT Applications. IEEE Communications Magazine, 55(2), 56–61. https://doi.org/10.1109/MCOM.2017.1600701CM

Pomboza-Junez, G., & Holgado-Terriza Juan, A. (2015). Control of home devices based on hand gestures. In 2015 IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin) (pp. 510–514). IEEE. https://doi.org/10.1109/ICCE-Berlin.2015.7391325

Pomboza-Junez, W. G., & Holgado-Terriza, J. A. (2016). Gestural Interaction with Mobile Devices. In Proceedings of the XVII International Conference on Human Computer Interaction - Interacción '16 (pp. 1–2). New York, New York, USA: ACM Press. https://doi.org/10.1145/2998626.2998672

Poppinga, B., Sahami Shirazi, A., Henze, N., Heuten, W., & Boll, S. (2014). Understanding shortcut gestures on mobile touch devices. In Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14 (pp. 173–182). New York, New York, USA: ACM Press. https://doi.org/10.1145/2628363.2628378

Singh, A., Arora, S., Shukla, P., & Mittal, A. (2015). Indian Sign Language gesture classification as single or double handed gestures. In 2015 Third International Conference on Image Information Processing (ICIIP) (pp. 378–381). IEEE. https://doi.org/10.1109/ICIIP.2015.7414800

Wallraven, C., Caputo, B., & Graf, A. (2003). Recognition with local features: the kernel recipe. Proceedings Ninth IEEE International Conference on Computer Vision, (Iccv), 1–8. https://doi.org/10.1109/ICCV.2003.1238351

Wettschereck, D., Aha, D. W., & Mohri, T. (1997). A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms. Artificial Intelligence Review, 11(1/5), 273–314. https://doi.org/10.1023/A:1006593614256

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., … Steinberg, D. (2008). Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1–37. https://doi.org/10.1007/s10115-007-0114-2

Yugan, V., Gobee, S., & Vickneswari, D. (2014). Design and Development of Lower Limb Exoskeleton for Rehabilitation. In The 15th International Conference on Biomedical Engineering - IFMBE Proceedings (Vol. 43, pp. 516–519). https://doi.org/10.1007/978-3-319-02913-9

Zhai, X. L., Jelfs, B., Chan, R. H. M., & Tin, C. (2017). Self-Recalibrating Surface EMG Pattern Recognition for Neuroprosthesis Control based on Convolutional Neural Network. Under Review, 11(July), 1–11. https://doi.org/10.3389/fnins.2017.00379

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 11/08/2020
Aceitação/Acceptance: 15/10/2020

Una Mirada al Desempeño Académico en las Pruebas Saber Pro de los Estudiantes de Ingeniería desde la Minería de Datos Educativa

Ricardo Timarán-Pereira[1], Arsenio Hidalgo-Troya[1], Fredy Vidal-Alegría[2]

ritimar@udenar.edu.co; arsenio.hidalgo@udenar.edu.co; fvidal@unimayor.edu.co

[1] Universidad de Nariño, Ciudad Universitaria Torobajo,520001, Pasto, Colombia.

[2] IU Colegio Mayor del Cauca, Edificio Bicentenario, 190001, Popayán, Colombia.

Pages: 29-42

Resumen: En este artículo se presentan los resultados obtenidos al aplicar la minería de datos educativa para descubrir factores asociados al desempeño académico de los estudiantes colombianos de las facultades de ingeniería que presentaron las pruebas Saber Pro entre los años 2012 y 2014. La metodología utilizada fue CRISP-DM, la guía de referencia más ampliamente aplicada en el desarrollo de proyectos de minería de datos. De las bases de datos del ICFES se seleccionó la información socioeconómica, académica e institucional de los estudiantes de ingeniería y se construyó, limpió y transformó un repositorio de datos con los resultados obtenidos en estas pruebas. Se aplicó la técnica de clasificación basada en árboles de decisión para descubrir factores asociados al buen o mal desempeño académico de los estudiantes en las pruebas Saber Pro. Se documentaron los patrones descubiertos y estos podrán incorporarse al conocimiento existente sobre desempeño académico en las pruebas Saber Pro.

Palabras-clave: Pruebas Saber Pro; Ingeniería; Desempeño Académico; Árboles de Decisión; Minería de Datos Educativa.

A Look at the Academic Performance in the Saber Pro Tests of Engineering Students from Educational Data Mining

Abstract: This paper presents the results obtained by applying educational data mining to discover factors associated with the academic performance of Colombian engineering students who presented the Saber Pro tests between 2012 and 2014. The methodology used was CRISP- DM, the most widely applied reference guide in the development of data mining projects. From the ICFES databases, the socioeconomic, academic and institutional information of the engineering students was selected, and a data repository was built, cleaned and transformed with the results obtained in these tests. The decision tree-based classification technique was applied to discover factors associated with the good or bad academic performance of students on the Saber Pro tests. The patterns discovered were documented and these could be incorporated into existing knowledge about academic performance on the Saber Pro tests.

Keywords: Saber Pro Tests; Engineering; Academic Performance; Decision Trees; Educational Data Mining.

1. Introducción

Según el Instituto Colombiano para la Evaluación de la Educación-ICFES (icfes, 2011; icfes, 2012a; icfes, 2012b), uno de los objetivos del examen de estado de calidad de la educación superior Saber Pro, es comprobar el grado de desarrollo de competencias de los estudiantes próximos a culminar los programas académicos de pregrado que ofrecen las instituciones de educación superior. El examen está compuesto por pruebas que evalúan competencias genéricas y específicas. Todos los estudiantes deben presentar los módulos de competencias genéricas sin importar el programa de formación que cursen; incluyen competencias de razonamiento cuantitativo, comunicación escrita, lectura crítica, competencias ciudadanas e inglés.

En las pruebas Saber Pro no se pretende que los estudiantes de todas las formaciones desarrollen las competencias genéricas a un mismo nivel, pero si es importante determinar cómo se asocian los factores socioeconómicos, académicos e institucionales del estudiante para obtener un determinado nivel de desempeño en estas pruebas (Timarán et.al, 2016).

Según Timarán et al., (2016) los estudios que se han realizado hasta el momento con respecto al análisis de los resultados de las pruebas Saber Pro, tales como (Ramírez & Teichler,2014; UNAL, 2012; Zapata, 2011), se basan en información procesada mediante un análisis estadístico, donde se han preocupado más por la generalización de los resultados, sin tener en cuenta otras interrelaciones, que por lo general están ocultas y que únicamente se pueden descubrir utilizando un tratamiento más complejo de los datos, que es posible con la minería de datos.

La aplicación de las técnicas y herramientas de la minería de datos en los diferentes contextos educativos se conoce como minería de datos en la educación EDM (del inglés *Educational Data Mining*). Según Huapaya et al. (2012), EDM busca desarrollar nuevos métodos para explorar la información que se genera dentro de los ambientes educativos con el fin de entender la forma en que los estudiantes aprenden. Se puede utilizar sus técnicas para explicar y/o predecir cualquier fenómeno dentro del campo educativo (Timarán et al., 2013). Utilizando las técnicas de minería de datos, se puede predecir, con un porcentaje muy alto de confiabilidad, la probabilidad de deserción de cualquier estudiante (Valero, 2009; Valero et al., 2010; Timarán et al., 2017). En el ámbito educativo la aplicación de la minería de datos como técnica de análisis se ubica en el entorno del sistema educativo y en concreto en educación superior. Las instituciones de educación superior pueden usar la minería de datos en la educación para hacer análisis comprensivos de las características de sus estudiantes, métodos evaluativos, develando procesos exitosos o, por el contrario, detectando fraudes o inconsistencias (Valero et al, 2010).

En este artículo se presentan los resultados de aplicar la minería de datos educativa para descubrir factores asociados al desempeño académico de los estudiantes colombianos de las facultades de ingeniería que presentaron las pruebas Saber Pro entre los años 2012 y 2014.

2. Metodología

La investigación fue de tipo descriptivo bajo el enfoque cuantitativo, aplicando un diseño no experimental. Para el descubrimiento de patrones de desempeño académico, se aplicó la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), por ser según (Azevedo & Santos, 2008; Chapman et al., 2000; Hernández et al, 2005; Villena, 2016), una de las guías de referencia más ampliamente utilizada en el desarrollo de proyectos de minería de datos. Esta metodología contempla seis fases: análisis del problema, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación.

2.1. Análisis del problema.

En esta fase, se realizaron las actividades que permitieron profundizar y apropiar de una manera completa el problema objeto de estudio, los objetivos y los requisitos de esta investigación, que posibilitaron la recolección de los datos correctos para interpretar adecuadamente los resultados

2.2. Comprensión de los datos.

En esta fase, se realizó la recolección inicial de datos, para establecer un primer contacto con el problema, familiarizarse con ellos, identificando su calidad y establecer las relaciones más evidentes que permitieron definir las primeras hipótesis. Como fuentes se utilizaron los datos de los resultados de los estudiantes colombianos de ingeniería que presentaron las pruebas Saber Pro, que se encontraban disponibles, al momento de realizar esta investigación, en las bases de datos del ICFES. Los datos más actualizados eran los resultados de las pruebas Saber Pro de los años 2012, 2013 y 2014.

Se seleccionaron únicamente los registros de estudiantes de programas profesionales de ingeniería y sus resultados en las pruebas Saber Pro, quedando un repositorio con 130.936 registros y 96 atributos. Este repositorio se identificó como SBP130936A96, el cual sirvió de base para las subsiguientes fases.

Con el fin de tener una comprensión preliminar de los datos se realizó una caracterización de las variables socioeconómicas de los estudiantes de ingeniería que presentaron las pruebas Saber Pro. Los resultados se muestran en la tabla 1.

Variable socioeconómica		N	%
Género	Femenino	45019	34,4%
	Masculino	85917	65,6%
Grupos de edad	Menor de 23 años	48159	36,8%
	23-31	68948	52,7%
	32-40	11093	8,5%
	41-49	2325	1,8%
	50 o más años	411	0,3%

Estado Civil	Soltero(a)	112531	85,9%
	Casado(a)	9650	7,4%
	Viudo(a)	75	0,1%
	Separado(a) / Divorciado(a)	796	0,6%
	Unión libre	7884	6,0%
Estrato social	Uno	11280	8,6%
	Dos	40864	31,2%
	Tres	52256	39,9%
	Cuatro	18082	13,8%
	Cinco	5714	4,4%
	Seis	2607	2,0%
	Rural	133	0,1%
Tiene personas a cargo	No	98412	75,2%
	Si	32524	24,8%
Trabaja	No	59130	45,2%
	Si	71806	54,8%
Ingreso familiar	Menos de 1 SM	8357	6,4%
	Entre 1 y menos de 2 SM	38766	29,6%
	Entre 2 y menos de 3 SM	33524	25,6%
	Entre 3 y menos de 5 SM	28068	21,4%
	Entre 5 y menos de 7 SM	10929	8,3%
	Entre 7 y menos de 10 SM	5899	4,5%
	10 o más SM	5393	4,1%
Total		**130936**	**100,0%**

Tabla 1 – Caracterización socioeconómica de los estudiantes de ingeniería

Según la tabla 1, por género, la mayoría son hombres, con un 65.6%; por edad el mayor porcentaje se encuentra en el rango de 23 a 31 años con un 52.7%. Igualmente, el estado civil prevalente es de solteros con un 85.9%, observándose porcentajes un tanto similares entre estudiantes casados y en unión libre que suman el 13.4%. Una altísima mayoría de los estudiantes, aproximadamente el 80%, pertenecen a los estratos sociales más bajos (uno, dos y tres), solo el 13.8% se ubican en el estrato medio y 6.6% en estratos más altos (cinco y seis). Además, un número importante, que representa el 24.8%, tiene personas a cargo. El 54.8% trabaja y más de la tercera parte (36%) de los hogares de los estudiantes tienen ingresos familiares inferiores a 2 salarios mínimos, aproximadamente una cuarta parte ingresos entre 2 y 3 salarios mínimos y tan solo 38.3% más de 3 salarios mínimos.

Por otra parte, se realizó una correlación entre los resultados de los estudiantes en las competencias genéricas y el ponderado general. A través del coeficiente de correlación de Pearson se establece como se asocian linealmente, entre sí, los resultados obtenidos en las cinco competencias genéricas por los estudiantes de los diferentes programas de Ingeniería en Colombia que presentaron las pruebas en el período 2012-2014. Los resultados se presentan en la tabla 2.

Competencia	Lectura Crítica	Comunicación Escrita	Razonamiento Cuantitativo	Inglés	Competencias Ciudadanas
Lectura Crítica	1	0,288**	0,576**	0,475**	0,636**
Comunicación Escrita		1	0,224**	0,244**	0,275**
Razonamiento Cuantitativo			1	0,513**	0,525**
Inglés				1	0,442**
Competencias Ciudadanas					1

Tabla 2 – Matriz de correlaciones entre competencias genéricas

De acuerdo con la tabla anterior, todas las correlaciones resultan altamente significativas (p valor <0,01) por el tamaño de los datos. Siguiendo la clasificación de Cohen para la interpretación del coeficiente de Pearson, se observó que lectura crítica presenta una correlación alta (r>0,5) con razonamiento cuantitativo y competencias ciudadanas, una correlación moderada (0,3<r<=0,5) con inglés y una correlación baja con comunicación escrita (r<=0,3); esta última competencia también presenta una correlación baja con el resto de las competencias. Razonamiento cuantitativo tiene una correlación alta con inglés y competencias ciudadanas, mientras que estas últimas competencias presentan una correlación moderada.

2.3. Preparación de los datos.

En esta fase, se seleccionaron los datos a los cuales se les aplicaría una determinada técnica de limpieza, generación de variables adicionales, integración de diferentes orígenes de datos, cambios de formato y eliminación de atributos. Como resultado de esta fase se obtuvo un repositorio de datos limpio y transformado, con 130.936 registros y 17 atributos, listo para aplicarle las técnicas de minería de datos y al cual se le denominó SBP130936A17.

Con el fin de facilitar la detección de patrones de rendimiento académico se discretizaron los valores de los puntajes numéricos obtenidos por los estudiantes de ingeniería en las cinco competencias genéricas, en los valores “sobre la media nacional” o “bajo la media nacional”. El resumen de estos resultados se muestra en la tabla 3.

Competencia		N	%
Lectura Crítica	Bajo la media	54831	41,9%
	Sobre la media	75885	58,1%
Comunicación Escrita	Bajo la media	66933	51,7%
	Sobre la media	62594	48,3%
Razonamiento Cuantitativo	Bajo la media	41253	31,6%
	Sobre la media	89462	68,4%
Inglés	Bajo la media	63982	48,9%
	Sobre la media	66740	51,1%
Competencias Ciudadanas	Bajo la media	59149	45,3%
	Sobre la media	71565	54,7%
Puntaje general	Bajo la media	50406	38,5%
	Sobre la media	80530	61,5%

Tabla 3 – Resultados en las competencias genéricas de los estudiantes de ingeniería

Como se deduce de la tabla 3, los estudiantes de programas de Ingeniería presentan mejores resultados académicos en las pruebas de razonamiento cuantitativo y lectura crítica, teniendo en cuenta que el 68.1% y el 58.1% obtuvo, respectivamente, puntajes por encima de la media. En las demás competencias se presentan unos resultados normales, los puntajes por encima y debajo de la media, se reparten de manera equivalente

2.4. Modelado.

Para el descubrimiento de patrones asociados al desempeño académico en las pruebas Saber Pro, se construyó un modelo de clasificación basado en árboles de decisión, utilizando el algoritmo J48 de la herramienta WEKA (Witten, Frank & Hall, 2011), el cual implementa al algoritmo C.45 (Quinlan, 1993). Se escogió este modelo porque según la experiencia de algunos autores (Han & Kamber, 2001; Sattler & Dunemann, 2001; Timarán et. al, 2016), para este tipo de proyectos, es probablemente el más utilizado y popular por su simplicidad y facilidad para entender los resultados obtenidos. Además, la importancia de los árboles de decisión se debe a su capacidad de construir modelos interpretables, siendo este un factor decisivo para su aplicación. Por otra parte, se escogió WEKA por ser una herramienta de minería de datos de software libre, ampliamente utilizada, distribuida bajo licencia GPL, que contiene una colección de algoritmos para realizar análisis de datos y modelado predictivo, tiene herramientas para la visualización de estos datos y provee una interfaz gráfica que unifica las herramientas para acceder fácilmente a sus funcionalidades. (Calleja, 2010; García, 2016).

Se aplicó el algoritmo J48, porque este se basa en la utilización del criterio de ganancia de información (*information gain*). De esta manera se consigue evitar que las variables con mayor número de posibles valores salgan beneficiadas en la selección. Además, el algoritmo incorpora una poda del árbol de clasificación una vez que éste ha sido

inducido (Hernández & Lorente, 2009). El parámetro que se tuvo en cuenta para la poda fue el factor de confianza C (*confidence level*), que influye en el tamaño y capacidad de predicción del árbol construido. Cuanto más baja se haga esa probabilidad, se exigirá que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar. El valor por defecto de este factor es del 25% y conforme va bajando este valor, se permiten más operaciones de poda y por lo tanto llegar a árboles cada vez más pequeños (García & Álvarez, 2010). Otro parámetro utilizado para variar el tamaño del árbol fue a través del factor M que especifica el mínimo número de instancias o registros por nodo del árbol (Witten et al, 2011).

Antes de construir el modelo, se definió el procedimiento para probar la calidad del modelo y su validez. Teniendo en cuenta que, para entrenar y probar un modelo de clasificación, se divide los datos en dos conjuntos: entrenamiento y prueba (Witten et al, 2011), se utilizó el método de validación cruzada (*Cross validation*) porque permite reducir la dependencia del resultado del experimento en el modo en el cual se realiza la partición (Hernández et al, 2005). Otra ventaja de la validación cruzada es que la varianza de los n errores de muestra parciales, permite estimar la variabilidad del método de aprendizaje con respecto al conjunto de datos. Para esta investigación, se utilizó 10 particiones (*10-fold cross validation*) teniendo en cuenta lo recomendado por Hernández et al. (2005).

Por otra parte, se evaluó o estimó el coste del clasificador para el repositorio SBP130936A17. a través de la matriz de confusión. La matriz de confusión representa de forma detallada el número de instancias que son predichas por clase. Los valores en la diagonal son los aciertos y el resto son los errores de clasificación (Fernández, 2009).

Teniendo en cuenta los parámetros de evaluación anteriores, se procedió a construir los diferentes árboles de decisión con el algoritmo J48. Se escogió como clase el puntaje general de cada estudiante obtenido en las pruebas Saber Pro. Con el fin de obtener diferentes modelos de árboles y escoger el de mejores resultados, se establecieron cuatro porcentajes de prepoda del árbol para el factor M igual a 0.05%, 1% ,1.5% y 2% del total de registros del repositorio de datos, y tres porcentajes para el factor confianza C igual a 25%, 15% y 5% y se construyeron los diferentes modelos combinando estos factores. Se escogió el árbol construido con los parámetros M=1300 (1%) y C=15% por los mejores resultados obtenidos y por la facilidad de análisis de los patrones. Una vez construido los árboles se aplicó un proceso de pospoda para dejar las ramas y por ende las reglas más representativas, que son aquellas que sobrepasan un mínimo soporte del 1% y una confianza del 60%. En la figura 1 se muestra la precisión del árbol y su matriz de confusión. El árbol construido con los parámetros M=1300 y C=15% se muestra en la figura 2.

2.5. Evaluación.

En la fase de evaluación se verificó si el modelo se ajusta a las necesidades establecidas en el proyecto. Se evaluaron los patrones encontrados con el fin de determinar su validez, remover los redundantes o irrelevantes y traducir los patrones útiles en términos que sean entendibles para el usuario. La evaluación e interpretación de los patrones descubiertos se describe en la sección de Interpretación y Discusión de Resultados.

```
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances       88797               67.8171 %
Incorrectly Classified Instances     42139               32.1829 %
Kappa statistic                          0.297
Mean absolute error                      0.4165
Root mean squared error                  0.4565
Relative absolute error                 87.955 %
Root relative squared error             93.8084 %
Total Number of Instances           130936

   a     b   <-- classified as
 24849 25557 |     a = BAJO LA MEDIA
 16582 63948 |     b = SOBRE LA MEDIA
```

Figura 1 – Exactitud del modelo y Matriz de Confusión obtenido con Weka

```
inst_acreditada = NO ACREDITADA
|  estu_estrato = BAJO
|  |  inst_caracter_academico = UNIVERSIDAD
|  |  |  inst_prgm_metodologia = A DISTANCIA: BAJO LA MEDIA (4409.0/1633.0)
|  |  |  inst_prgm_metodologia = PRESENCIAL
|  |  |  |  estu_rango_edad = [<23): SOBRE LA MEDIA (8723.0/3169.0)
|  |  |  |  estu_rango_edad = [23-32)
|  |  |  |  |  inst_prgm_zonageo = BOGOTA: SOBRE LA MEDIA (3937.0/1290.0)
|  |  |  |  |  inst_prgm_zonageo = ATLANTICA: BAJO LA MEDIA (2844.0/1098.0)
|  |  inst_caracter_academico = INSTITUCION UNIVERSITARIA: BAJO LA MEDIA (11791.0/4223.0)
|  estu_estrato = MEDIO
|  |  inst_caracter_academico = UNIVERSIDAD
|  |  |  estu_rango_edad = [<23): SOBRE LA MEDIA (13868.0/3568.0)
|  |  |  estu_rango_edad = [23-32)
|  |  |  |  inst_prgm_zonageo = BOGOTA: SOBRE LA MEDIA (8966.0/3010.0)
|  |  inst_caracter_academico = INSTITUCION UNIVERSITARIA
|  |  |  fami_ing_fmliar_mensual = Entre 1 y menos de 2 SM: BAJO LA MEDIA (3163.0/1172.0)
|  estu_estrato = ALTO: SOBRE LA MEDIA (4610.0/926.0)
inst_acreditada = ACREDITADA
|  inst_caracter_academico = UNIVERSIDAD: SOBRE LA MEDIA (30343.0/5191.0)
|  inst_caracter_academico = INSTITUCION UNIVERSITARIA
|  |  inst_origen_general = PRIVADO: SOBRE LA MEDIA (1407.0/172.0)

Number of Leaves  :          49
Size of the tree :           63
```

Figura 2 – Árbol podado de Clasificación obtenido con Weka

2.6. Implementación.

En la fase de implementación, se documentaron los patrones descubiertos y estos podrán incorporarse al conocimiento existente sobre desempeño académico en las pruebas Saber Pro de los estudiantes de carreas de ingeniería en Colombia. Es potestad del MEN, ICFES y las instituciones de educación superior, integrar este conocimiento en sus procesos de toma de decisiones para mejorar la calidad de la educación en Colombia.

3. Interpretación y Discusión de Resultados

De acuerdo con la figura 1, se puede observar que el modelo clasifica correctamente a 88.797 instancias, que corresponde a un porcentaje de precisión del 68% e incorrectamente a 32.189 instancias, correspondiente a un porcentaje del 32%. Además. teniendo en cuenta la matriz de confusión, del total de 130936 estudiantes de ingeniería evaluados, según el modelo, el desempeño académico de 89.505 estudiantes está sobre la media, que corresponde a un 68% del total de casos y 41.431 está bajo la media, que corresponde a un 32% del total. Por otra parte, de estos 89.505 estudiantes que el modelo clasifica sobre la media, el 71,5% (63.948) son casos correctos y el 28,5% (25.557) son falsos casos. De igual manera de los 41.431 que el modelo clasifica bajo la media, el 60% (24.849) son casos correctos y el 40%(16.582) son falsos casos. Con respecto al total de casos correctos que están sobre la media según la tabla 3, en el puntaje general (80,530), la exactitud del modelo es del 79,4%. De igual manera, con respecto al total de casos correctos que están bajo la media (50.406), la exactitud del modelo es del 49.3%.

Para escoger los patrones más representativos, se tuvo en cuenta un mínimo soporte del 1% y una confianza mínima de 60%, tanto los que se ubican por encima de la media, como aquellos que se sitúan por debajo de ella. Entre los patrones más importantes, descritos en forma de reglas y acordes con los resultados de la figura 2 están:

Regla 1: Si el estudiante es de una universidad no acreditada, de estrato bajo y está matriculado en un programa de ingeniería bajo la modalidad a distancia, entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar bajo la media nacional. El 3,4% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 63% de los 4.409 estudiantes que se clasifican así, están correctamente clasificados y el 5,5% del total de estudiantes que están bajo la media, cumplen este patrón.

Regla 2: Si el estudiante es de una universidad no acreditada, de estrato bajo, está matriculado en un programa de ingeniería bajo la modalidad presencial y es menor que 23 años, entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar sobre la media nacional. El 6,7% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 63,7 % de los 8.723 estudiantes que se clasifican así, están correctamente clasificados y el 6,9% del total de estudiantes que están sobre la media, cumplen este patrón.

Regla 3: Si el estudiante es de una Institución Universitaria no acreditada y de estrato bajo, entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar bajo la media nacional. El 9% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 64,2 % de los 11.791 estudiantes que se

clasifican así, están correctamente clasificados y el 15% del total de estudiantes que están bajo la media, cumplen este patrón.

Regla 4: Si el estudiante es de una universidad no acreditada, es de estrato medio y es menor que 23 años, entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar sobre la media nacional. El 10,6% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 74,3 % de los 13.868 estudiantes que se clasifican así, están correctamente clasificados y el 12,8% del total de estudiantes que están sobre la media, cumplen este patrón.

Regla 5: Si el estudiante es de una universidad no acreditada, de estrato medio, está entre los 23 y 32 años y es de Bogotá D.C., entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar sobre la media nacional. El 6,8% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 66,4 % de los 8.966 estudiantes que se clasifican así, están correctamente clasificados y el 7,4% del total de estudiantes que están sobre la media, cumplen este patrón.

Regla 6: Si el estudiante es de una Institución Universitaria no acreditada, de estrato medio y los ingresos familiares están entre 1 y 2 salarios mínimos legales, entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar bajo la media nacional. El 2,4% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 62,9 % de los 3.163 estudiantes que se clasifican así, están correctamente clasificados y el 3.9 % del total de estudiantes que están bajo la media, cumplen este patrón.

Regla 7: Si el estudiante es de una universidad no acreditada y es de estrato alto, entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar sobre la media nacional. El 3,5% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 79,9 % de los 4.610 estudiantes que se clasifican así, están correctamente clasificados y el 4,6% del total de estudiantes que están sobre la media, cumplen este patrón.

Regla 8: Si el estudiante es de una universidad acreditada, entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar sobre la media nacional. El 23,2% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 82,9 % de los 30.343 estudiantes que se clasifican así, están correctamente clasificados y el 31,2% del total de estudiantes que están sobre la media, cumplen este patrón.

Regla 9: Si el estudiante es de una institución universitaria acreditada de carácter privado, entonces el desempeño en las pruebas Saber Pro tiene mayor probabilidad de estar sobre la media nacional. El 1,1% del total de estudiantes de ingeniería que presentaron las pruebas Saber Pro se clasifican de esa manera. El 87,8 % de los 1.407 estudiantes que se clasifican así, están correctamente clasificados y el 1,5% del total de estudiantes que están sobre la media, cumplen este patrón.

Analizando los patrones encontrados sobre el rendimiento académico en las pruebas Saber Pro de los estudiantes colombianos de los programas de ingeniería, cabe destacar que la acreditación institucional es un factor muy importante asociado al buen desempeño

en las pruebas Saber Pro. Tendencia que se mantiene en un estudio que Timarán et al. (2016) realizaron con las pruebas Saber Pro, pero con todos los estudiantes colombianos de programas profesionales que presentaron las pruebas en el año 2011. Esto hecho indica, que el cumplimiento de los doce factores de calidad señalados por el CNA (2013), dan cuenta de las condiciones de favorabilidad de todos los procesos que permiten un mejor desempeño académico de los estudiantes, en las competencias genéricas previstas por el ICFES. Por otra parte, que la IES no sea acreditada se constituye en un factor que se asocia con un bajo desempeño académico en las Pruebas Saber Pro de los estudiantes de carreras de ingeniería.

La modalidad de estudio es otro factor asociado al desempeño académico en estas pruebas. La modalidad presencial se asocia al buen desempeño (por encima de la media nacional), en cambio la modalidad a distancia se asocia al mal desempeño (por debajo de la media nacional). Este patrón coincide con el estudio realizado por por Parra et al. (2012), realizada con estudiantes de primer semestre de pregrado de la Facultad de Ingeniería de la Universidad de Antioquia, cohorte 2012-2. En este estudio se encontró que la modalidad del programa se asocia de manera importante al desempeño académico. Por ejemplo, en la condición de insuficiente se ubica el 17% de los matriculados en la modalidad presencial y el 50% de los de la modalidad virtual; cifras que indican que los estudiantes de programas presenciales tienen mejor desempeño que aquellos que están en la modalidad a distancia.

El estrato socioeconómico al cual pertenece el estudiante de ingeniería, es otro factor asociado al desempeño académico en las pruebas Saber Pro. Los estudiantes de estratos altos y medios tienen mejor desempeño académico que los de estratos bajos que están bajo la media. Estos patrones coinciden con los resultados del estudio de Garbanzo (2007), sobre desempeño académico, en los cuales concluyen que un resultado generalmente aceptable en el desempeño académico es la existencia de una asociación significativa entre el nivel socioeconómico del estudiante y su desempeño académico.

4. Conclusiones y Trabajos Futuros

El estudio realizado mostró que existen factores asociados al buen o mal desempeño académico de los estudiantes colombianos de programas de ingeniería en las pruebas Saber Pro. La acreditación institucional es uno de ellos, asociado al buen desempeño. Es importante que el MEN redoble esfuerzos en la política pública de acreditación tanto a nivel institucional, como a nivel de carreras o programas, en este caso de ingeniería, para que las IES mejoren la calidad de sus procesos de docencia, investigación e interacción social, cuyo logro permita un alto desempeño académico de sus estudiantes en los diferentes programas, en pro del mejoramiento de la calidad de la educación en Colombia.

Como trabajos futuros se plantea utilizar otras técnicas predictivas de clasificación para comparar los resultados obtenidos en esta investigación y complementar este estudio utilizando técnicas descriptivas de minería de datos como reglas de Asociación y agrupamiento que permitan relacionar cuales factores se presentan juntos y las similitudes que se pueden encontrar entre los estudiantes de acuerdo con su desempeño académico en las pruebas Saber Pro.

Referencias

Azevedo, A.,Santos, M.(2008). KDD, SEMMA and CRISP-DM: a parallel overview. In Proceedings of IADIS European Conference on Data Mining, (pp. 182-185). Amsterdam, Netherlands . ISBN: 978-972-8924-63-8.

Calleja, A. (2010). Minería de Datos con Weka para la Predicción del Precio de Automóviles de Segunda Mano. Recuperado de: https://riunet.upv.es/bitstream/handle/10251/10097/PFC_DSIC-80_Agust%C3%ADnCalleja.pdf

Chapman, P., Clinton, J., Kerber, R, Khabaza, T., Reinartz, T.,Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. CRISP-DM consortium: NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA), and OHRA Verzekeringen en Bank Groep B.V. (The Netherlands).

Fernández, G. (2009). Extracción de Información de la Web usando Técnicas de Minería de Datos. Recuperado de: http://www.tdg-seville.info/Download.ashx?id=48

Garbanzo, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde calidad de la educación superior pública. Revista Educación, 31(1):43-63.

García, J. (2016). Comenzando con Weka: Filtrado y selección de subconjuntos de atributos basada en su relevancia descriptiva para la clase. Technical report. Universidad de Malaga. Recuperado de: https://www.researchgate.net/publication/308141950.

García, M., Álvarez, A. (2010). Análisis de datos en WEKA–pruebas de selectividad. Disponible en http://www. it. uc3m. es/jvillena/irc/practicas/06-07/28. pdf. Recuperado de: http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf

Han, J., Kamber, M. (2001). Data Mining: Concepts and Techniques. San Francisco, USA: Morgan Kaufmann Publishers.

Hernández, J., Ramírez, M., & Ferri, C. (2005). Introducción a la Minería de Datos. Madrid (España): Pearson Prentice Hall. ISBN: 84-205-4091-9.

Huapaya, C., Lizarralde, F., Arona, G., & Massa, S. (2012). Minería de datos educacional en ambientes virtuales de aprendizaje. En memorias de XIV Workshop de Investigadores en Ciencias de la Computación, (pp. 996-1000). Entre Ríos, Argentina: RedUNCI. Recuperado de: http://sedici.unlp.edu.ar/handle/10915/18347.

Icfes (2011). Lineamientos Saber Pro. Recuperado de http://aprendeenlinea.udea.edu.co/lms/moodle/file.php/532/Lineamientos_SABER_PRO_2011_2_30_08_1_.pdf.

Icfes (2012a). Examen Saber Pro, Módulos de competencias genéricas y especificas. Evaluación de la calidad de la educación superior. Bogotá: 2012. Recuperado de: http://www.icfes.gov.co/examenes/.../151-saber-pro-modulos-de-competencias.

Icfes (2012b). Saber Pro: Principales resultados en Competencias Genéricas. Santa Marta, Colombia: 2012. Recuperado de: http:// www.icfes.gov.co/examenes/.../151-saber-pro-modulos-de-competencias.

Parra, C., Mejía, L., Valencia, A., Castañeda, E., Restrepo, G., Usuga, O.,& Mendoza, R. (2013). Rendimiento académico de los estudiantes de primer semestre de pregrado de la Facultad de Ingeniería de la Universidad de Antioquia: cohorte 2012-2. Medellín: Ingeniería y Sociedad. Recuperado de: http://www.udea.edu.co/portal/page/portal/bibliotecaSedesDependencias/unidadesAcademicas/FacultadIngenieria/Diseno/Archivos/Tab/Rendimiento%20acad%C3%A9mico%20de%20los%20estudiantes%5B1%5D.pdf.

Quinlan, J. (1993). *C4. 5: programs for machine learning* (Vol. 1). San Mateo, CA, USA: Morgan Kaufmann Publishers, Inc. Recuperado de http://books.google.com.co/books? hl=es&lr=&id=HExncpjbYroC&oi=fnd&pg=PR7&dq = Programs+for+Machine+Leraning&ots = nLkbbRq2Yj&sig = Y5h5CQUdtbZjs1Fjd8ilbJfyRLE.

Ramírez, C., & Teichler, U. (2014). Factores socioeconómicos y educativos asociados con el desempeño académico, según nivel de formación y género de los estudiantes que presentaron la prueba SABER PRO 2009. Bogotá: ICFES. Recuperado de: https://www.icfes.gov.co/documents/20143/233983/Factores+ socioeconomicos+y+educativos+asociados+a+d esempeno+academico+segun+formacion+y+genero+Saber+Pro+2009.pdf.

Sattler, K, Dunemann, O. (2001). SQL Database Primitives for Decision Tree Classifiers. In Paques H, Liu L, Grossman D, editors. The 10th ACM International Conference on Information and Knowledge Management, (379-386). Atlanta, USA: ACM New York.

Timarán, R., Calderón, A., & Jiménez, J. (2013). Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil. Revista Ventana Informática, No 28(ene.-jun.). Manizales, Colombia: Facultad de Ciencias e Ingeniería, Universidad de Manizales. p. 31-47. ISSN:0123-9678.

Timarán, S.R., Hernández, I., Caicedo, S.J., Hidalgo, A., & Alvarado, J.C. (2016). Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional. Bogotá: Ediciones Universidad Cooperativa de Colombia. http://dx.doi.org/10.16925/9789587600490.

Timarán, R., Jiménez, J. y Calderón, A. (2017). Detección de patrones de deserción estudiantil con minería de datos. Pasto, Colombia: Editorial Universitaria. Universidad de Nariño. ISBN:978-958-8958-38-5.

UNAL (2012). Análisis de los resultados obtenidos por la Universidad Nacional de Colombia sede Bogotá en las pruebas Saber Pro 2011–2. Bogotá: Universidad Nacional de Colombia. Recuperado de: www.unal.edu.co/diracad/evaluacion/SaberPro_2012/analisis_de_resultados.pdf.

Valero, S (2009). Aplicación de técnicas de minería de datos para predecir deserción. Puebla, México: Universidad Tecnológica de Izúcar de Matamoros. Recuperado de: http://www.utim.edu.mx/~svalero/docs/MineriaDesercion.pdf.

Valero, S., Salvador, A. y García, M. (2010). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. Recuperado de: http:// www.utim.edu.mx/~svalero/docs/e1.pdf.

Villena, J. (2016). CRISP-DM: La metodología para poner orden en los proyectos de Data Science. Recuperado de: https://data.sngular.team/es/art/25/crisp-dm-la-metodologia-para-poner-orden-en-los-proyectos-de-data-science.

Witten, I., Frank, E., & Hall, M. (2011). Data Mining: Practical Machine Learning Tools and Techniques (Third Edition). Morgan Kaufmann. ISBN:978-0-12-374856-0. https://doi.org/10.1016/C2009-0-19715-5.

Zapata, L. (2011). Factores académicos asociados al bajo rendimiento en inglés en las pruebas ECAES presentadas por los estudiantes de la Facultad de Educación en el año 2009. Recuperado de: https://docplayer.es/1310875-Factores-academicos-asociados-al-bajo-rendimiento-en-ingles-en-las-pruebas-ecaes-presentadas-por-los-estudiantes-de-la-facultad-de-educacion-en-el-ano.html.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 17/08/2020
Aceitação/Acceptance: 07/10/2020

Una revisión sistemática de los métodos de identificación y detección de corrupción en contratación pública

Julio-César Luna-Ortega[1, 2], Carlos-Alberto Cobos-Lozada[1],
Martha-Eliana Mendoza-Becerra[1]

jcluna@unicauca.edu.co; ccobos@unicauca.edu.co; mmendoza@unicauca.edu.co

[1] Universidad del Cauca, Departamento de Sistemas, Calle 5 Nº 4-70, 190003, Popayán, Colombia

[2] Clúster CreaTIC, Grupo de Investigación CreaTIC, Cl. 20N Nº 7a-34, 190003, Popayán, Colombia

Pages: 43-57

Resumen: Recientemente, se ha observado un notable crecimiento de los casos de corrupción en procesos de contratación. Esto ha motivado a que las entidades promuevan estrategias anticorrupción, la mayoría de estas estrategias son manuales y de índole social. Para poder predecir o detectar efectivamente (eficiente y eficazmente) la corrupción se hace necesario realizar procesos automáticos. La presente revisión busca establecer los conceptos fundamentales del estado de arte relacionados con soluciones que apliquen estrategias anticorrupción y con base en esto, analizar la viabilidad de crear un modelo de detección e identificación de corrupción en el proceso de contratación pública. Los resultados muestran que combinar técnicas estadísticas con tecnologías de la información (inteligencia artificial, big data, minería de datos, entre otros) es una línea de investigación prometedora en esta área de aplicación. Además, que se hace necesario desarrollar modelos automáticos específicos para la detección de los diferentes delitos de corrupción en contratación.

Palabras clave: Corrupción, Contratación, Modelo, Revisión sistemática.

A systematic review of the methods for identification and detection of corruption in public procurement

Abstract: Recently, notable growth has been observed in corruption cases in public procurement processes. This has motivated entities to promote anti-corruption strategies, most of these strategies are manual and social in nature. In order to effectively (efficiently and efficiently) predict and detect corruption, it is necessary to carry out automatic processes. This review seeks to establish the fundamental concepts of the state of the art related to solutions that apply anti-corruption strategies and based on this, analyze the feasibility of creating a model for detecting and identifying corruption in the public procurement process. The results show that combining statistical techniques with information technologies (artificial intelligence, big data, data mining, among others) is a promising line of research in

this area of application. Furthermore, it is necessary to develop specific automatic models for the detection of the different crimes of corruption in contracting.

Keywords: Corruption, Hiring, Model, Systematic review.

1. Introducción

El creciente descubrimiento de casos de corrupción en procesos de contratación pública que se ha dado en los últimos años se ha convertido en una oportunidad para la implementación de sistemas que analicen los datos y permitan detectar y prevenir nuevos casos de corrupción. A nivel mundial, el promedio de las pérdidas sufridas por las entidades víctimas de corrupción es aproximadamente de USD$250,000 por caso.

Según diversas investigaciones el proceso más permeado por actos corruptos es el de contratación pública, lo cual, en una grave contradicción, puesto que la Contratación Estatal se presenta como una herramienta utilizada por un estado social de derecho para garantizar la satisfacción del interés general. Cuando el proceso de contratación se ve afectado por un caso de corrupción, este deja de estar enfocado en la satisfacción del interés común para enmarcarse en el de terceros, afectando el bienestar de la comunidad y en el caso del Estado, alterando significativamente las finanzas éste.

El aumento de los casos de corrupción ha hecho que diversas entidades estén aplicando variadas estrategias anticorrupción, desafortunadamente la mayoría de estas son técnicas manuales y de enfoque social, que buscan prevenir, más que detectar o identificar posibles índices de corrupción. Si bien las técnicas manuales pueden ser útiles como estrategias anticorrupción con enfoques sociales, en estrategias con enfoque predictivo se hace necesario realizar un proceso automático (como modelos estadísticos y/o tecnológicos), dado que, como se menciona en (Ehlermann-Cache, 2007) la detección manual de la corrupción es intrínsecamente difícil puesto que esta se encubre por su propia naturaleza.

Esta revisión sistemática busca establecer los conceptos fundamentales del estado de arte relacionados con modelos y/o estrategias utilizadas en contra de la corrupción y con ello analizar la viabilidad de crear un modelo sistemático de detección e identificación de corrupción en el proceso de contratación pública.

La guía definida en (Kitchenham & Charters, 2007) fue utilizada como base para el proceso de la revisión sistemática utilizado en este estudio. El resto de este documento está organizado de la siguiente manera: en la sección 2 se presenta la planeación de la revisión con su protocolo de investigación, más adelante en la sección 3 se muestran los resultados obtenidos de dicha revisión, luego en la sección final se presentan las conclusiones y trabajos futuros que el grupo de investigación espera realizar.

2. Planeación de la Revisión Sistemática de la Literatura (RSL)

Los documentos relacionados con la temática de investigación se buscaron en Scopus, ACM Digital Library, IEEEXplore y ScienceDirect. Posterior a esto, se aplicaron los criterios de inclusión/exclusión para la selección de los estudios primarios y finalmente se realizó una evaluación de la calidad de los estudios incluidos.

2.1. Preguntas de investigación

Las preguntas de investigación de esta revisión buscan recopilar y categorizar la información existente acerca de la detección de corrupción en contratación pública, con el fin de realizar una revisión general del tema, identificando las carencias existentes y con esto, en un próximo paso plantear una alternativa de solución a la problemática expuesta (ver Tabla 1).

Preguntas de Investigación	Motivación
Q1. ¿Qué estudios existen sobre la detección o identificación de corrupción en contratación?	Determinar el número de publicaciones actuales y la tendencia a lo largo de los últimos años en relación con el campo de la identificación de corrupción en contratación.
Q2. ¿Qué tipos de soluciones se han implementado para la identificación de corrupción?	Determinar cuáles son las técnicas, métodos, modelos o similares que se han utilizado para la detección de corrupción en contratación y como son los informes que estas realizan.
Q3. ¿Qué tipos de corrupción en contratación se han trabajado?	Determinar qué tipos de corrupción o fraude son normalmente a los que se presta más atención en el proceso de contratación.
Q4. ¿En qué dominios de aplicación se ha tratado el problema de la identificación de la corrupción?	Determinar cuáles son los objetos (contratos, negocios, etc.) en donde se ha trabajado sobre el problema de la identificación de corrupción.
Q5. ¿Qué indicadores se utilizan en las mediciones de corrupción o fraude?	Determinar cuáles son los indicadores que se suelen utilizar a la hora de evaluar un contrato y cuál es el dominio de valores que estos pueden tomar para alertar un posible caso de corrupción.
Q6. ¿Qué problemas y desafíos identifican los autores en relación con la corrupción?	Determinar cuáles son los problemas y retos identificados por los autores que trabajan en el proceso de identificación de la corrupción en la contratación.

Tabla 1 – Preguntas de investigación.

2.2. Desarrollo del protocolo de revisión

A continuación, se resume el protocolo seguido para llevar a cabo la RSL.

Cadena de búsqueda de estudios primarios: Los términos de la cadena de búsqueda fueron construidos a través de un proceso iterativo de mejora con el análisis de los resultados reportados por cada versión previa de esta. Se hizo uso de la estrategia planteada en (Bailey et al., 2007), la cual consiste en derivar los términos de las preguntas de investigación, identificando y agrupándolos en población, intervención, comparación y resultado (ver Tabla 2). Los operadores booleanos OR incorporan palabras alternativas y sinónimos en cada una de las partes, y los operadores AND para vincular los términos principales entre sí.

Proceso de búsqueda: Las búsquedas se realizaron sobre Scopus, ACM Digital Library, IEEEXplore y ScienceDirect. Siguiendo el proceso propuesto en (Bailey et al., 2007), la búsqueda se organizó en fases; en la primera se identificaron los estudios candidatos derivados de las palabras claves de la cadena de búsqueda, una vez esto que se procedió a aplicar los criterios de selección sobre el abstract de cada uno de ellos, posteriormente los mismos criterios fueron aplicados a todo el documento de los estudios resultantes de la fase previa, finalmente se aplicó la técnica “bola de nieve”, la cual consiste en revisar

las referencias de cada estudio primario buscando identificar más estudios candidatos, así mismo se hizo una revisión de la literatura gris. La Tabla 3 resume el proceso de búsqueda realizado.

Concepto	Estructura	Términos Alternativos y Sinónimos	Descripción
Técnicas	Resultado	(“model” OR “algorithm” OR “framework” OR “technology” OR “method” OR “technique”) AND	Se delimita lo que se está buscando: algoritmos, técnicas, modelos y similares.
Corrupción	Población	(“*corruption” OR “collusion” OR “fraud”) AND	Se define el problema a nivel general: corrupción y similares.=
Tipos de delitos	Comparación	(“bribery” OR “buys influence” OR “conflicts of interest” OR “cronyism” OR “embezzlement” OR “extort*” OR “favoritism” OR “gift-giving” OR “graft” OR “influence peddling” OR “kickback” OR “nepotism” OR “parochialism” OR “patronage” OR “political interference” OR “rigging” OR “splitting” OR “theft”) AND	Se delimitan los tipos específicos de delitos de corrupción.
Objeto	Población	(“governm*” OR “agreement” OR “contract” OR “commitment” OR “arrangement” OR “settlement” OR “understanding” OR “bond” OR “deal” OR “bargain”) AND	Se menciona el objeto que tiene el problema: contratación gubernamental y palabras relacionadas con contrato.
Acciones	Intervención	(“detect” OR “avoid” OR “prevent” OR “warn” OR “caution” OR “identify” OR “understanding” OR “avert” OR “stop” OR “alert” OR “perceive” OR “distinguish” OR “notify” OR “inform” OR “recognize” OR “establish”)	Se relaciona con el tipo de soluciones que buscan identificar, evitar, prevenir (o similares) la corrupción.

*Los asteriscos significan que cualquier carácter o caracteres pueden ser incluidos en la palabra, de esta manera se consiguen permutaciones en los términos de búsqueda en los que se utiliza.

Tabla 2 – Cadena de búsqueda usada para iniciar la revisión.

Actividad	Resultados
Aplicar cadena de búsqueda	245
Eliminar duplicados	229
Aplicar criterios de selección (inclusión y exclusión)	40
Leer documento completo	8
Revisar artículos excluidos y Snowballing (bola de nieve)	16
Revisar literatura gris	19
Aplicar criterios de evaluación de calidad	19

Tabla 3 – Proceso de búsqueda.

Criterios de inclusión y exclusión: Los criterios para incluir un estudio primario fueron los siguientes: 1) Estudios escritos en el idioma inglés; 2) Estudios completos publicados en revistas, conferencias, congresos o talleres de prestigio con revisión por pares; y 3) El estudio propone un modelo, técnica, algoritmo, framework o similar. Los criterios para excluir un estudio primario fueron los siguientes: 1) El artículo no presenta un diseño de investigación bien definido; 2) El artículo no presenta una pregunta de investigación bien definida; 3) El artículo corresponde a un estudio de debate, o solo está disponible en forma de resumen o presentación. 4) El artículo es un estudio secundario, es decir, corresponde a un reporte de resultados de otras investigaciones; 5) La contribución principal del artículo no se relaciona con alguna modelo/técnica/método/algoritmo de identificación de corrupción; 6) El artículo no propone una modelo/técnica/método/algoritmo, o que lo hagan, pero la propuesta no pueda ser implementada con ayuda del computador; y 7) El artículo presenta un modelo de índole psicosocial. Por último, vale la pena mencionar que, debido a la poca existencia de investigaciones que abordan el problema presentado en esta investigación, no se tuvo en cuenta el año de publicación del artículo como un criterio para incluir o no el trabajo en la presente revisión sistemática.

Evaluación de calidad de los estudios primarios: Adicional a la selección (criterios de inclusión y exclusión) se realizó una evaluación de la calidad de los estudios primarios (Bailey et al., 2007) para ponderar la importancia de cada estudio seleccionado. Esta evaluación permitió comprender mejor las posibles comparaciones entre los estudios y guiar la interpretación de los hallazgos. Con este objetivo se desarrolló una prueba con un sistema de puntuación de tres valores (-1, 0, 1) en 6 preguntas (ver Tabla 4). La puntuación total máxima corresponde a la suma de las calificaciones parciales y el puntaje total está en el dominio [-6, 6]. El puntaje de calidad de cada estudio se presenta en la Tabla 5, donde 6 es el máximo valor posible.

No	Criterios	Puntuación		
		-1	0	1
C1	El estudio presenta una descripción detallada sobre los tipos de corrupción o fraude que se presentan en un objeto de población (ejemplo contratos)	No	Parcialmente	Si
C2	El estudio contiene un modelo/técnica/método/algoritmo/similar aplicada.	No	Parcialmente	Si
C3	El estudio tiene como finalidad aplicar acciones (identificar, detectar, prevenir, etc.) sobre un problema de corrupción	No	Parcialmente	Si
C4	El estudio expone de manera clara y detallada los resultados obtenidos tras aplicar un modelo/técnica/método/algoritmo/similar sobre un problema de corrupción en un objeto de población	No	Parcialmente	Si
C5	A qué cuartil pertenece la revista, conferencia o congreso en donde fue publicado el estudio	Q4 O NO	Q3	Q2 O Q1
C6	El estudio ha obtenido dos o más citas por año desde su publicación (según índice de citas de Google, Promedio anual)	menos de 1 cita por año	1 cita por año	2 o más citas por año

Tabla 4 – Prueba de evaluación de la calidad.

Referencia	(Porter & Zona, 1993)	(Bajari, 2001)	(Bajari & Ye, 2003)	(Parente, Schulte, Jost, Sullivan, & Klindworth, 2012)	(Chotibhongs & Arditi, 2012)	(Ralha & Silva, 2012)	(Morozov & Podkolzina, 2013)	(Shan, Chan, Le, Xia, & Hu, 2015)	(Fazekas, Tóth, & King, 2016)	(Peltier-Rivest, 2017)	(Belokrylov & Runova, 2017)	(Islam & Yousuf, 2018)	(Peltier-Rivest, 2018)	(López-Iturriaga & Sanz, 2018)	(Imhof, Karagök, & Rutz, 2018)	(Wachs & Kertész, 2019)	(Huber & Imhof, 2019)	(Owusu, Chan, & Ameyaw, 2019)	(Dorofeeva & Tokmakov, 2020)
Puntaje	6	6	6	3	6	6	3	3	5	1	3	1	3	5	6	5	6	6	1

Tabla 5 – Resultados de la evaluación de los estudios primarios ordenado por año y puntaje.

3. Resultados de la Revisión Sistemática de la Literatura (RSL)

Haciendo uso del protocolo mencionado se encontraron 19 estudios que cumplieron con los diferentes criterios de calidad establecidos. Estos estudios se analizaron y organizaron en 5 grupos (A, B, C, D y E) conforme se presentan en la Tabla 6. En esta tabla, además, se muestra para cada estudio las respuestas a cada pregunta de investigación previamente formulada. Para determinar el grupo al que pertenecía cada estudio se hizo un análisis de las respuestas obtenidas por cada uno de ellos en las diferentes preguntas de investigación, siendo la pregunta **Q2** la de mayor peso en esta decisión.

Id	Ref.	Preguntas
	(Porter & Zona, 1993)	**Q1:** Propuesta.
		Q2: Caso de estudio de un cartel con información de cuentas debidamente detalladas respecto a sus actividades y prácticas de licitación. Técnicas estadísticas. Procedimientos de prueba econométricos.
		Q3: Colusión. Aparejo de ofertas (práctica ilegal en la que los competidores se confabulan para elegir al ganador de una licitación).
		Q4: Contrato.
		Q5: No utiliza.
		Q6: Falta de información de contratación pública.

Id	Ref.	Preguntas
	(Bajari & Ye, 2003)	**Q1:** Propuesta.
		Q2: Técnicas estadísticas. Enfoque para identificar y probar la manipulación de ofertas en subastas de adquisición. Los autores destacan la importancia de la implementación de herramientas computacionales que puedan ser utilizadas para detectar la manipulación de ofertas.
		Q3: Aparejo de ofertas.
		Q4: Contrato.
		Q5: Índice de probabilidad.
		Q6: Falta de información de contratación pública. Empresas corruptas pueden aplicar acciones que ayuden a evitar la identificación de la corrupción.
A	(Chotibhongs & Arditi, 2012)	**Q1:** Propuesta.
		Q2: Técnicas estadísticas. Método paso a paso para detectar las ofertas colusorias mediante el análisis de los datos históricos de ofertas proporcionados por el propietario de una construcción. El método se basa en un modelo de regresión e implica identificar a los posibles postores del cártel utilizando la prueba residual y la prueba de estabilidad de la estructura de costos.
		Q3: Colusión. Aparejo de ofertas.
		Q4: Contrato.
		Q5: Niveles de correlación.
		Q6: La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. La corrupción puede tomar muchas formas. El modelo no puede detectar las ofertas colusorias en el momento en que sucede; se usa para detectar posible colusión en el pasado y solo si hay acceso a información histórica detallada sobre un conjunto de variables independientes.
	(Morozov & Podkolzina, 2013)	**Q1:** Propuesta.
		Q2: Técnicas estadísticas. Método econométrico de detección de manipulación de ofertas.
		Q3: Colusión. Aparejo de ofertas.
		Q4: Contrato.
		Q5: Niveles de correlación.
		Q6: La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución.

Id	Ref.	Preguntas
	(Shan et al., 2015)	**Q1:** Encuesta. Identificación de elementos importantes de medición de corrupción.
		Q2: Técnicas estadísticas. Modelo sistemático para medir la corrupción potencial en un proyecto de construcción pública.
		Q3: No enfatiza un delito.
		Q4: Contrato.
		Q5: No utiliza.
		Q6: Falta de información de contratación pública.
	(Imhof et al., 2018)	**Q1:** Propuesta.
		Q2: Técnicas estadísticas. Método para detectar la manipulación de ofertas mediante la aplicación de pantallas (índices estadísticos específicos). El método es particularmente adecuado para detección de la colusión parcial (colusión que no involucra a todas las empresas y/o todos los contratos en un conjunto de datos específico). Se logra aislar a un grupo de empresas "sospechosas" de ser un cartel local de manipulación de ofertas.
		Q3: Colusión. Aparejo de ofertas.
		Q4: Contrato.
		Q5: Índice de probabilidad.
		Q6: Falta de información de contratación pública. La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. La corrupción puede tomar muchas formas.
	(Parente et al., 2012)	**Q1:** Propuesta.
		Q2: Técnicas estadísticas. Inteligencia artificial.
		Q3: Fraude.
		Q4: Contrato.
		Q5: Índice de probabilidad.
		Q6: La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. Falta de información de contratación pública.
	(Fazekas et al., 2016)	**Q1:** Propuesta.
		Q2: Técnicas estadísticas. Inteligencia artificial. El método hace uso de Big Data para crear un indicador de corrupción a nivel de contrato.
		Q3: No enfatiza un delito.
		Q4: Contrato.
		Q5: Banderas rojas. Propone indicador de corrupción a través de Big Data.
		Q6: La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución.

Id	Ref.	Preguntas
B	(Belokrylov & Runova, 2017)	**Q1:** Propuesta.
		Q2: Técnicas estadísticas. Algoritmos computacionales. Propone un modelo basado en lógica difusa.
		Q3: Aparejo de ofertas. Sobornos.
		Q4: Contrato.
		Q5: No utiliza.
		Q6: La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. Falta de compromiso de las entidades.
	(Huber & Imhof, 2019)	**Q1:** Propuesta.
		Q2: Técnicas estadísticas. Inteligencia artificial Combinan técnicas de aprendizaje automático con pantallas estadísticas.
		Q3: Aparejo de ofertas.
		Q4: Contrato.
		Q5: Binarios. Clasificación binaría que informa si una oferta es marcada como posible colusoria, clasificando como 1 si la probabilidad de colusión de la oferta es superior a 0,5 y 0 de lo contrario.
		Q6: Debe realizarse de forma discreta, si algún cartel se entera del proceso, podría destruir evidencia. Falta de compromiso de las entidades. Información de costos es difícil de obtener.
	(Owusu et al., 2019)	**Q1:** Propuesta. Muestra una revisión de la literatura con 21 actividades que se deben realizar en las cuatro etapas del proceso de adquisición.
		Q2: Técnicas estadísticas. Algoritmos computacionales. Modelo que usa un enfoque de computación flexible (técnica de evaluación sintética difusa).
		Q3: No enfatiza un delito.
		Q4: Contrato.
		Q5: Índice de probabilidad.
		Q6: La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución.
	(Ralha & Silva, 2012)	**Q1:** Propuesta. Se enfoca en extraer información de las bases de datos del proceso de adquisiciones federales brasileñas para identificar carteles.
		Q2: Inteligencia artificial. Minería de datos: Agrupamiento y reglas de asociación.
		Q3: Carteles.
		Q4: Contrato.
		Q5: Binario.
		Q6: Falta de compromiso de las entidades.

Id	Ref.	Preguntas
C	(Islam & Yousuf, 2018)	**Q1:** Propuesta.
		Q2: Inteligencia artificial. Se ha propuesto un modelo que usa el algoritmo K-means.
		Q3: No enfatiza en un delito.
		Q4: Corrupción pública.
		Q5: Índice de probabilidad.
		Q6: Falta de compromiso de las entidades.
	(López-Iturriaga & Sanz, 2018)	**Q1:** Propuesta.
		Q2: Inteligencia artificial Propone sistema de alertas tempranas basado en red neuronal (mapas autoorganizados, SOM).
		Q3: No enfatiza un delito.
		Q4: Corrupción pública.
		Q5: Binarios.
		Q6: Falta de compromiso de las entidades.
	(Bajari, 2001)	**Q1:** Propuesta.
		Q2: Algoritmos computacionales. Los algoritmos computacionales desarrollados en este documento pueden usarse para la evaluación de los daños causados por la manipulación de ofertas.
		Q3: Colusión. Aparejo de ofertas.
		Q4: Contrato.
		Q5: Diferencia entre ofertas corruptas y competitivas.
		Q6: La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. Falta de información de contratación.
D	(Wachs & Kertész, 2019)	Q1: Propuesta.
		Q2: Ciencia de redes. Propone un marco basado en análisis de redes (grafos) para identificar grupos de empresas que interactúan intensamente en un mercado.
		Q3: Colusión. Aparejo de ofertas.
		Q4: Contrato.
		Q5: Índices de probabilidad.
		Q6: Falta de información de contratación pública. La corrupción puede tomar muchas formas.
	(Dorofeeva & Tokmakov, 2020)	**Q1:** Propuesta.
		Q2: Inteligencia artificial. Análisis de las direcciones IP de los miembros de una licitación pública.
		Q3: Aparejo de ofertas.
		Q4: Contrato.
		Q5: No utiliza.
		Q6: Falta de compromiso de las entidades. Falta de claridad en la definición del objeto a abordar.

Id	Ref.	Preguntas
	(Peltier-Rivest, 2017)	**Q1:** Estudio de caso.
		Q2: No relacionado con tecnología o estadística. Propone un marco de gobernanza, haciendo uso de la teoría del diamante del fraude.
		Q3: Aparejo de ofertas. Sobornos. Conflictos de interés. Extorsión.
		Q4: Contrato.
		Q5: Banderas rojas.
		Q6: Falta de información de contratación pública. Falta de compromiso de las entidades.
E	(Peltier-Rivest, 2018)	**Q1:** Propuesta. Muestra aspectos interesantes para tener en cuenta en el proceso de detección de la corrupción.
		Q2: No relacionado con tecnología o estadística. Introduce un modelo derivado de las teorías criminológicas (como el triángulo de fraude).
		Q3: Aparejo de ofertas. Sobornos.
		Q4: Contrato.
		Q5: Banderas rojas.
		Q6: Debe realizarse de forma discreta, si algún cartel se entera del proceso, podría destruir evidencia. Falta de compromiso de las entidades.

Tabla 6 – Agrupamiento de los estudios

3.1. Análisis de los grupos resultantes

A continuación, se muestran un análisis detallado de la conformación de cada uno de los grupos con relación a las preguntas de investigación planteadas.

Grupo A: Los estudios pertenecientes a este grupo tienen un enfoque de solución relacionada con técnicas estadísticas como Test econométricos ((Morozov & Podkolzina, 2013; Porter & Zona, 1993)), pantallas ((Imhof et al., 2018)), regresión ((Chotibhongs & Arditi, 2012)) y teoría de conjuntos difusos ((Bajari & Ye, 2003; Shan et al., 2015)). Los tipos de delitos de corrupción abordados por este grupo de estudios son en su mayoría la colusión y el aparejo de ofertas, así mismo las métricas de corrupción utilizadas tienden a ser análisis estadísticos como es el caso de un índice de probabilidad. Es de destacar que los estudios pertenecientes a este grupo se presentan desde 1993 hasta 2018.

Grupo B: La principal característica de los estudios pertenecientes a este grupo es que el modelo desarrollado involucra técnicas estadísticas con métodos de Inteligencia Artificial, destacándose la utilización de la teoría de conjuntos difusos con algoritmos de agrupamiento (Belokrylov & Runova, 2017) buscando medir el riesgo de corrupción en el proceso de contratación o técnicas de evaluación difusas con el objetivo de evaluar los niveles de vulnerabilidad con respecto a las etapas de corrupción (Owusu et al., 2019), así mismo la combinación del análisis de regresión con algoritmos predictivos para la detección analítica de fraude (Parente et al., 2012) o con algoritmos de big data

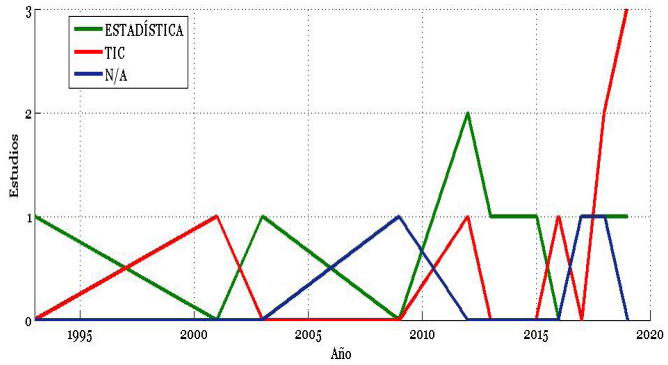
(Fazekas et al., 2016) para el desarrollo de un indicador compuesto de corrupción y finalmente la combinación de pantallas con aprendizaje de máquina (Huber & Imhof, 2019) con el objetivo de predecir la colusión a través de carteles de manipulación de ofertas. La mayoría de estos estudios no enfatizan su solución a un delito en particular, y los que los hacen se enfocan en el aparejo de ofertas. Por otro lado, similar al grupo anterior, la tendencia de las métricas de corrupción utilizadas se encuentra en un marco estadístico. Estos trabajos son relativamente nuevos, teniendo su primera publicación en 2012, así mismo, se puede apreciar que la tendencia de este tipo de artículos ha venido incrementando en los últimos años.

Grupo C: Los integrantes de este grupo son estudios en donde se aplicó una solución tecnológica basada en Minería de Datos, haciendo uso de algoritmos de agrupamiento y reglas asociación en (Ralha & Silva, 2012), donde se obtiene un modelo que permitió detectar las ofertas colusorias mediante el análisis de los datos históricos de ofertas. En (Islam & Yousuf, 2018) hacen uso del algoritmo K-means (agrupamiento) el cual permitió crear diferentes grupos de acuerdo con el nivel de corrupción de un grupo de entidades. Finalmente en (López-Iturriaga & Sanz, 2018) se usó un enfoque de red neuronal, mapas autoorganizados (agrupamiento), para predecir casos de corrupción en diferentes espacios temporales, en una región específica de un país. Los estudios pertenecientes a este grupo por lo general no se enfocan en un delito en particular, solo uno de ellos lo hace, enfocándose en la organización de carteles. Por otro lado, en este grupo prima un indicador binario como métrica de corrupción. Finalmente, y, como era de esperarse, similar al anterior grupo este tipo de soluciones es relativamente nueva teniendo su primera publicación en 2012.

Grupo D: Los integrantes de este grupo, si bien hacen uso de técnicas computacionales, difieren del anterior en que dichas técnicas no están enmarcadas dentro de la Inteligencia Artificial. Destacándose soluciones que hacen uso de algoritmos computacionales (Bajari, 2001) para encontrar diferencias entre ofertas competitivas y colusorias, o la utilización de análisis de redes (grafos) (Wachs & Kertész, 2019) para la identificación de grupos de empresas que interactúan intensamente en un mercado y detectar su potencial colusorio. Y finalmente el análisis de las direcciones IP de los miembros de una licitación pública para la identificación de posible presencia de un cartel en un proceso de licitación. (Dorofeeva & Tokmakov, 2020). Este grupo se centra en la detección de colusión y aparejo de ofertas, teniendo como principal métrica los índices de probabilidades. Finalmente, y similar a los anteriores grupos que hacen uso de técnicas de tecnologías de la información, la tendencia de este tipo de investigaciones es relativamente nueva.

Grupo E: El último grupo de investigaciones está conformado por los estudios cuyo modelo de solución no está relacionado con tecnología o estadística, en estos estudios se despliegan una serie de estrategias para prevenir o detectar la corrupción, haciendo uso de la teoría del diamante del fraude (Peltier-Rivest, 2017, 2018). Los delitos que se abordan en este grupo son la colusión y el aparejo de ofertas, teniendo como métrica el uso de banderas rojas.

Por último, en la Gráfica 1 se muestran los estudios primarios realizados por año, discriminados por la técnica utilizada, se logra apreciar que en los últimos años hay un crecimiento de los modelos que utilizan técnicas computacionales.



Gráfica 1 – Estudios primarios por año.

4. Conclusiones y Trabajo Futuro

Los resultados obtenidos en esta revisión sistemática demuestran que el uso de técnicas estadísticas para solventar el problema de la identificación temprana de la corrupción es una idea que en los últimos años ha tenido una buena acogida, así como la combinación de estas técnicas con tecnologías de la información. Lo anterior sumado al incremento en los casos de corrupción global reportados, llevan a la necesidad de seguir en el desarrollo de modelos que permitan solventar de manera más eficaz y eficiente el problema de la detección de la corrupción en procesos de contratación. Los estudios muestran que, dada la complejidad del problema, se hace necesario contar con un conjunto (batería) de herramientas especializadas para atacar los diferentes delitos que ocurren en contratación.

Respecto al trabajo futuro, el grupo de investigación pretende trabajar en lo siguiente: 1) El desarrollo de un modelo que integre diversas técnicas estadísticas y de computación, que permita la identificación temprana de la corrupción en un dominio especifico de aplicación y 2) La elaboración de uno o más índices que permitan evaluar el modelo desarrollado.

Agradecimientos

A la Universidad del Cauca y al Clúster CreaTIC por financiar parcialmente el desarrollo de la presente investigación.

Referencias

Bailey, J., Budgen, D., Turner, M., Kitchenham, B., Brereton, P., & Linkman, S. (2007). Evidence relating to Object-Oriented software design: A survey. Paper presented at the First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007).

Bajari, P. (2001). Comparing competition and collusion: a numerical approach. Economic Theory, 18(1), 187-205. doi: https://doi.org/10.1007/PL00004128

Bajari, P., & Ye, L. (2003). Deciding between competition and collusion. Review of Economics and statistics, 85(4), 971-989. doi: https://doi.org/10.1162/003465303772815871

Belokrylov, K. A., & Runova, L. P. (2017). Localization of Subjective Assessment Risks in the Public Procurement System Based on Fuzzy Logic. Journal of Advanced Research in Law and Economics, 8(2 (24)), 411-417. doi: https://dx.doi.org/10.14505//jarle.v8.2(24).09

Chotibhongs, R., & Arditi, D. (2012). Detection of collusive behavior. Journal of Construction Engineering and Management, 138(11), 1251-1258. doi: https://doi.org/10.1061/(ASCE)CO.1943-7862.0000542

Dorofeeva, J. A., & Tokmakov, M. (2020). The Use of Identical IP Address by Tender Participants as an Indication of Cartel Scheme Digital Transformation of the Economy: Challenges, Trends and New Opportunities (pp. 244-252): Springer.

Ehlermann-Cache, N. (2007). Bribery in public procurement: methods, actors and counter-measures: OECD.

Fazekas, M., Tóth, I. J., & King, L. P. (2016). An objective corruption risk index using public procurement data. European Journal on Criminal Policy and Research, 22(3), 369-397. doi: https://doi.org/10.1007/s10610-016-9308-z

Huber, M., & Imhof, D. (2019). Machine learning with screens for detecting bid-rigging cartels. International Journal of Industrial Organization, 65, 277-301. doi: https://doi.org/10.1016/j.ijindorg.2019.04.002

Imhof, D., Karagök, Y., & Rutz, S. (2018). Screening for Bid Rigging—Does It Work? Journal of Competition Law & Economics, 14(2), 235-261. doi: https://doi.org/10.1093/joclec/nhy006

Islam, M. T., & Yousuf, M. A. (2018). Development of a Corruption Detection Algorithm using K-means Clustering. Paper presented at the 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE).

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

López-Iturriaga, F. J., & Sanz, I. P. (2018). Predicting public corruption with neural networks: An analysis of spanish provinces. Social Indicators Research, 140(3), 975-998. doi: https://doi.org/10.1007/s11205-017-1802-2

Morozov, I., & Podkolzina, E. A. (2013). Collusion detection in procurement auctions. Higher School of Economics Research Paper No. WP BPR, 25. doi: http://dx.doi.org/10.2139/ssrn.2221809

Owusu, E. K., Chan, A. P., & Ameyaw, E. (2019). Toward a cleaner project procurement: Evaluation of construction projects' vulnerability to corruption in developing countries. Journal of cleaner production, 216, 394-407. doi: https://doi.org/10.1016/j.jclepro.2019.01.124

Parente, S. T., Schulte, B., Jost, A., Sullivan, T., & Klindworth, A. (2012). Assessment of predictive modeling for identifying fraud within the Medicare program. Health Manag Policy Innov, 1(2), 8-37.

Peltier-Rivest, D. (2017). The prevention and detection of corruption in pharmaceutical companies. Pharmaceuticals Policy and Law, 19(1-2), 17-31. doi: https://dx.doi.org/10.3233/PPL-170451

Peltier-Rivest, D. (2018). A model for preventing corruption. Journal of Financial Crime, 25(2), 545-561.

Porter, R. H., & Zona, J. D. (1993). Detection of bid rigging in procurement auctions. Journal of political economy, 101(3), 518-538. doi: https://doi.org/10.1086/261885

Ralha, C. G., & Silva, C. V. S. (2012). A multi-agent data mining system for cartel detection in Brazilian government procurement. Expert Systems with Applications, 39(14), 11642-11656. doi: https://doi.org/10.1016/j.eswa.2012.04.037

Shan, M., Chan, A. P., Le, Y., Xia, B., & Hu, Y. (2015). Measuring corruption in public construction projects in China. Journal of Professional Issues in Engineering Education and Practice, 141(4), 05015001. doi: https://doi.org/10.1061/(ASCE)EI.1943-5541.0000241

Wachs, J., & Kertész, J. (2019). A network approach to cartel detection in public auction markets. arXiv preprint arXiv:1906.08667. doi: https://doi.org/10.1038/s41598-019-47198-1

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 12/09/2020
Aceitação/Acceptance: 18/11/2020

Estrategia para la homogeneización, comparación e integración de Scrum e ISO 9001:2015

César Pardo[1], Miguel Morcillo[1], Elizabeth Suescún[2]

cpardo@unicauca.edu.co; miguelmorcillo@unicauca.edu.co; esuescu1@eafit.edu.co

[1] Grupo de investigación GTI, Universidad del Cauca, Carrera 2º Calle 15n Esquina Campus Tulcán 19001, Popayán, Colombia.

[2] Grupo GIDITIC, Universidad EAFIT, Carrera 49 Nº 7 Sur-50, Medellín, Colombia

Pages: 58-74

Resumen: Las organizaciones que utilizan Scrum y requieren un sistema de gestión de la calidad como ISO 9001, se enfrentan a desafíos en su esfuerzo por demostrar que sus actividades de proceso cumplen con los requisitos del estándar, y por desarrollar sus propios procedimientos, metodologías y herramientas para hacerlo. Este artículo presenta una estrategia utilizada para homogeneizar, comparar y conocer el grado de conformidad de Scrum respecto a los requisitos de los capítulos 8, 9 y 10 del estándar ISO 9001:2015, asimismo, se propone un proceso híbrido resultado de la integración de Scrum e ISO 9001. Con los resultados obtenidos, se puede observar que Scrum tiene un grado de conformidad igual o mayor que el 50% de los elementos comparados con el estándar. La estrategia de armonización y el proceso propuesto podrían ser de gran utilidad para profesionales y organizaciones que intentan aplicar y beneficiarse de las ventajas de ambas soluciones.

Palabras-clave: Scrum; ISO 9001; Armonización de múltiples modelos; Modelos híbridos; Estrategia de armonización.

Homogenize, compare, and integrate Scrum and ISO 9001: 2015

Abstract: Companies using Scrum which also require a quality management system such as ISO 9001, are facing challenges in their effort to: demonstrate that their process activities meet the requirements of ISO 9001 and to develop their own procedures, methodologies and tools to do it. This paper presents a strategy used to homogenize, compare and know the degree of compliance of Scrum regarding to the requirements of chapter 8, 9 and 10 of the standard ISO 9001: 2015, also, a hybrid process resulting from the integration of Scrum and ISO 9001 is proposed. With the results obtained, it can be seen that Scrum has a degree of conformity equal to or greater than 50% of the elements compared to the standard. The harmonization strategy and the proposed process could be useful for organizations and professionals trying to apply and benefit from the advantages of both solutions.

Keywords: Scrum; ISO 9001; Harmonization of multiple models; Hybrid models; Harmonization strategy.

1. Introducción

Desde el surgimiento de los enfoques ágiles, cada vez más organizaciones se están moviendo de los enfoques tradicionales hacia procesos ágiles iterativos e incrementales (en adelante, soluciones) (Fernandez & Fernandez, 2008), (Royce, 1987), (Larman, 2003), incluso, ha llevado a que desarrolladores sientan que a diferencia de los enfoques ágiles como Scrum, modelos como ISO 9001 y otros estándares que permiten gestionar y asegurar la calidad, se hayan vuelto conflictivos, irrelevantes o que ya no sean necesarios (Stålhane & Hanssen, 2008).

La idea de armonizar Scrum y otros enfoques ágiles con el estándar ISO 9001 es probablemente casi tan antigua como el propio desarrollo ágil, incluso los creadores del desarrollo ágil no lo consideran un problema, sin embargo, gerentes y profesionales de control de calidad sí (Stålhane & Hanssen, 2008), esto, debido a que al parecer, existe una idea generalizada que concibe que un proceso diseñado a partir de los requisitos de la norma ISO 9001 es incompatible con un proceso diseñado a partir de los elementos y prácticas definidas por un enfoque ágil como Scrum. Sin embargo, según (Qasaimeh & Abran, 2010), ISO 9001 a diferencia de otras soluciones, es genérico y puede ser aplicado al ciclo de vida del desarrollo de organizaciones de software, ayudando a alcanzar sus objetivos de calidad mediante la implementación de ciertos requisitos en todas las fases de desarrollo, por ejemplo: diseño, implementación, producción, instalación y mantenimiento (Grimes, 2002).

Además, la implementación de un sistema de gestión de la calidad (SGC), brinda a las organizaciones la capacidad de administrar y mejorar efectivamente la calidad de sus productos y servicios, y de abordar riesgos y oportunidades asociadas al contexto de mercado y a los objetivos de la organización, consiguiendo así una ventaja competitiva de cara al cliente, accediendo a nuevos nichos de mercado, entre otros (Petersen & Wohlin, 2010). El estándar ISO 9001 no sugiere un proceso estático, incluso, recomienda diseñar, establecer, implementar y mantener un proceso de diseño y desarrollo de productos y servicios que sea adecuado a sus características y necesidades para garantizar la provisión de productos y servicios (ISO, 2015).

Por otra parte, los clientes potenciales de una organización sienten confianza por un por un proceso de desarrollo certificado por ISO 9001. En este sentido, se puede establecer que si una organización tiene el certificado ISO es porque realmente sigue el estándar, y que ha sido evaluada tanto por auditores internos como externos que; a través de ciclos continuos, han identificado oportunidades de mejora a realizar en sus SGC. Además, las organizaciones que ya cuentan con la certificación ISO 9001, pueden conservar su certificado a la vez que pueden adoptar, adaptar, diseñar, introducir e instanciar un enfoque ágil como Scrum, que permite la gestión de un equipo y un producto en desarrollo partiendo de un insumo definido para la creación de productos y servicios de acuerdo con necesidades explícitas (Stålhane & Hanssen, 2008).

Teniendo en cuenta lo anterior, el objetivo de este artículo es presentar una estrategia utilizada para armonizar Scrum e ISO 9001:2015 por medio de la homogeneización,

comparación e integración de sus estructuras y elementos en común. Con esta estrategia fue posible observar que en realidad hay un gran número de elementos en común que pueden unir a estas soluciones, que características que los separen, o elementos opuestos, pero lo más importante, es posible encontrar cómo unificar sus diferencias e integrarlas para obtener beneficios en los proyectos en las organizaciones de software. Teniendo en cuenta lo anterior, el documento está organizado de la siguiente manera. La Sección 2 presenta los trabajos relacionados. La Sección 3 presenta la estrategia de armonización definida para homogeneizar y comparar Scrum con los requisitos de los capítulos 8, 9 y 10 del estándar ISO 9001:2015, también se presenta el nivel de conformidad alcanzado por Scrum respecto al estándar, se sugiere cómo tratar las no relaciones, se presenta un proceso híbrido como resultado de la integración de los elementos comparados en ambas soluciones, así como las limitaciones que se tuvieron durante la armonización de las soluciones y la solución propuesta. Finalmente, la Sección 4 presenta las conclusiones y trabajos futuros.

2. Antecedentes y análisis de trabajos relacionados

A partir del análisis de los trabajos relacionados, es posible observar que son varios y variadas las propuestas encontradas. Es importante resaltar que para el presente trabajo, se tuvieron en cuenta aquellos trabajos donde se incluían esfuerzos de armonización del estándar ISO 9001 y enfoques ágiles, algunos trabajos son: en (Qasaimeh & Abran, 2010) y (Qasaimeh & Abran, 2013) se analiza la capacidad del enfoque XP para cumplir con los requisitos del proceso de realización de producto de la ISO 9001, en (Stålhane & Hanssen, 2008) se analiza la manera de cómo llevar a cabo la armonización de enfoques ágiles, desarrollo Lean Software y el estándar ISO 9001:2000 en lo que respecta a la documentación, trazabilidad y control.

En cuanto a estudios de caso realizados, la mayoría involucra el estudio de la implementación del estándar ISO 9001 y el enfoque XP en grandes organizaciones, por ejemplo: (Silva & Cunha, 2006), (Wright, 2003), (Vriens, 2003) y (Nawrocki et al., 2002). Además, se encontró un estudio de caso interesante en (Popovic, 2015), el cual presenta la aplicación del estándar ISO 9001:2008 y Scrum, donde se utilizaron para el desarrollo del producto usaba tecnologías Java y herramientas de código abierto. Asimismo, ha sido posible encontrar un reporte de experiencia de un estudio de caso realizado en una organización de tamaño mediano en China, donde se implementa CMMI después de la aplicación del estándar ISO 9001 para un proceso iterativo basado en RUP (Shen & Ruan, 2008). Por otra parte, en (Jakobsen & Sutherland, 2009) se discute la introducción de Scrum a cinco organizaciones nivel 5 de CMMI.

En (Kuhrmann et al., 2017), se presentan los resultados de una encuesta sobre enfoques de desarrollo de software híbrido, en el estudio se presentan qué enfoques se usan en la práctica, cómo se combinan con otros enfoques y qué factores influyen en el uso y la combinación de enfoques de desarrollo de software híbrido, siendo ISO 9001 y Scrum de los más utilizados. En (Calabrese et al., 2019), se presenta una propuesta para adaptar la documentación de Scrum y los estándares IEEE recomendados para las etapas de

documentación del proceso de desarrollo, a la documentación requerida por IRAM-ISO 9001:2015 en Argentina.

En general, aunque es posible encontrar algunos trabajos relacionados con la propuesta aquí presentada, se observa que los trabajos existentes carecen de un análisis detallado que permita entender la estrategia de armonización utilizada para llevar a cabo la integración de los requisitos del estándar ISO 9001 y Scrum, por lo tanto, no es posible replicar los resultados alcanzados en esos trabajos. En este sentido, se hace necesario definir una estrategia clara que permita: primero, resolver las diferencias estructurales del estándar ISO 9001 y Scrum, segundo, que sugiera los pasos de comparación llevados a cabo para identificar las relaciones entre las dos soluciones, y, tercero, que presente el resultado de la integración de las prácticas de ambas soluciones sin afectar los beneficios que éstos ofrecen. Teniendo en cuenta lo anterior, a continuación, se presenta la estrategia diseñada para armonizar Scrum e ISO 9001:2015.

3. Armonización de Scrum e ISO 9001:2015

3.1. Estrategia de armonización

La Figura 1 presenta la estrategia de armonización definida a partir de la implementación de un conjunto de métodos propuestos en (César Pardo et al., 2012), los métodos son: (i) un método de homogeneización (HoMethod) (C Pardo et al., 2016), el cual permite identificar y resolver las diferencias estructurales entre modelos por medio de una estructura común de elementos de proceso, (ii) un método de comparación (CoMethod) (Pino et al., 2010), el cual describe un proceso para apoyar la identificación de las relaciones entre modelos, y (iii) un método integración (IMethod) (César Pardo et al., 2016), el cual establece los pasos a seguir para integrar las descripciones de los elementos deseados. La incorporación de estos métodos nos permitió llevar a cabo la armonización paso a paso de las soluciones. Con el fin de organizar y gestionar las personas y actividades en toda la estrategia, esta tuvo en cuenta dos roles: el *intérprete* y el *revisor*, junto con las actividades de los tres métodos: **Método 1**: Homogeneización (HoMethod), este método implicó: (i) adquirir conocimiento acerca de las soluciones involucradas; (ii) analizar la estructura y terminología; (iii) identificar los requisitos, y (iv) corresponder los elementos de cada solución en una plantilla común de elementos de proceso; **Método 2**: Comparación (CoMethod), implicó: (i) diseñar la comparación; (ii) realizar la comparación; (iii) presentar los resultados y (iv) analizar los resultados, y **Método 3**: Integración (IMethod), implicó: (i) diseñar la integración; (ii) establecer los criterios de integración; (ii) llevar a cabo la integración; (iv) analizar los resultados y (v) presentar el modelo integrado. En la siguiente sección se presenta de manera detallada las tareas de la estrategia de armonización que se siguieron para armonizar Scrum e ISO 9001:2015.

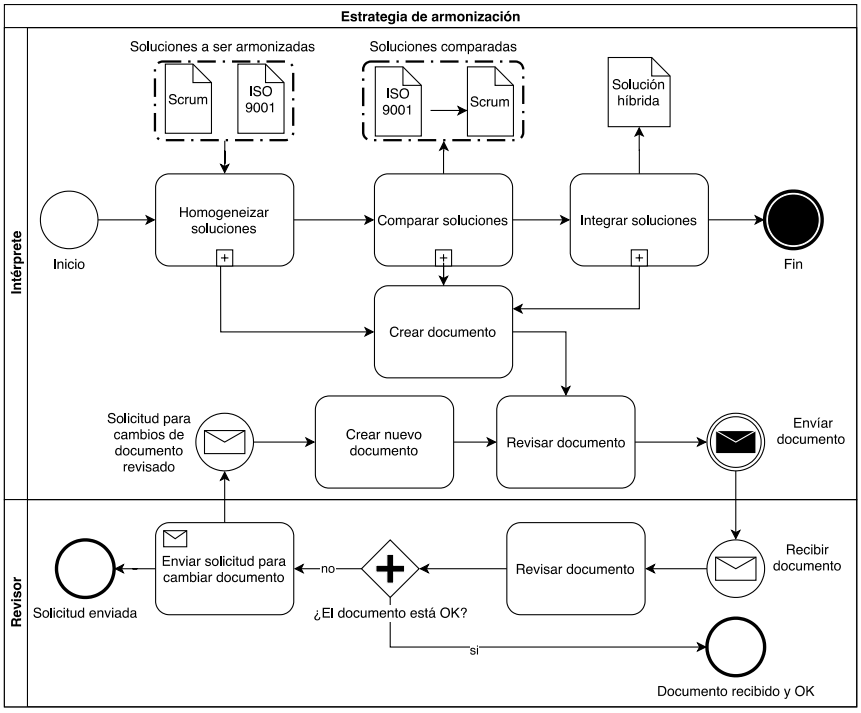


Figura 1 – Resumen del proceso en BPMN de la estrategia de armonización definida

3.2. Llevando a cabo la homogeneización

ISO 9001:2015 es un estándar internacional para la implementación de un SGC que abarca aspectos transversales de una organización, desde los deseos y necesidades de los clientes y sus partes interesadas, hasta la medición, evaluación y control del desempeño de los procesos internos, la naturaleza genérica del estándar permite su aplicación en la industria de desarrollo de software. Para obtener una certificación, es necesario demostrar la conformidad de los procesos de software generando la evidencia necesaria del cumplimiento de los requisitos. El costo de implementación de ISO 9001 va de la mano de capacitaciones de personal, consultores, auditores internos y por supuesto, el costo de la certificación y las consecuentes re certificaciones (ISO, 2015).

Antes de llevar a cabo la comparación de las dos soluciones, fue necesario armonizarlas a través de HoMethod y la estructura de elementos de proceso común descrita en (César Pardo et al., 2016), en donde se pudo organizar los elementos de cada solución y prepararlos para su posterior comparación. En este sentido, es posible observar que el estándar está conformado por capítulos, que a su vez contienen requisitos, los requisitos contienen literales y en algunos casos los literales contienen numerales. En total, son 10 capítulos sin contar la introducción (capítulo 0), de los cuales desde el 1 hasta el

3 se describe: el objeto, campo de aplicación, referencias normativas y los términos y definiciones. Los requisitos del estándar están comprendidos desde el capítulo 4, referente al *Contexto de la Organización* hasta el capítulo 10, referente a la *Mejora*. Los requisitos comprendidos entre los capítulos 4 y 7 están relacionados con la organización y el SGC de una organización.

La generalidad del estándar permite que pueda ser utilizado por cualquier organización que genere bienes o servicios, sin embargo, la armonización se enfocará únicamente en los capítulos: 8 (Operación), 9 (Evaluación del desempeño) y 10 (Mejora), ya que son los que se relacionan directamente con el ciclo de producción, evaluación y mejora de productos/servicios. Por otra parte, al analizar la sintaxis del estándar, en algunos de los requisitos se pueden observar verbos que hacen referencia al nivel de cumplimiento que el estándar requiere, por ejemplo: el verbo ***deber,*** en tiempo presente "debe", es común a través de los requisitos del estándar, lo mismo ocurre en los diferentes literales y parágrafos del requisito que implica una obligatoriedad en el resto del predicado. Así mismo, se identifica la palabra ***salida***, que corresponde a un producto de trabajo resultante de las actividades obligatorias. En la Tabla 1 se puede observar un resumen de la sintaxis utilizada para analizar las descripciones de los requisitos del estándar y la identificación de requisitos, esta sintaxis es sugerida por el método de homogeneización presentado en (César Pardo et al., 2013). En la Tabla 1 también es posible observar un extracto a manera de ejemplo del requisito 8.1 del estándar, las palabras que aparecen resaltadas son aquellas que implican una acción específica de acuerdo a la sintaxis utilizada.

Sintaxis	**Descripción**
[El/La/Los] [sujeto] [debe/no debe] [verbo], [verbo] ... [verbo] [para] [objetivo], [objetivo] ... [objetivo]	Implica que es obligatorio que el sujeto, que puede ser la organización o uno varios roles; cumplan con las tareas, actividades, procedimientos y/o objetivos descritos por el requisito.
Requisito	"Necesidad o expectativa establecida, generalmente implícita u obligatoria" (ISO, 2015)**.**
Proceso	"Un proceso es un conjunto de actividades interrelacionadas que interactúan para transformar entradas en salidas" (ISO, 2015)**.**
Entrada	Producto de trabajo necesario para un proceso.
Salida	Producto de trabajo obtenido de un proceso.
Ejemplo del análisis de sintaxis del requisito 8.1 del estándar ISO 9001:2015 (ISO, 2015)	
La organización ***debe*** planificar, implementar y controlar los ***procesos*** (véase 4.4) necesarios para ***cumplir*** los ***requisitos*** para la provisión de ***productos y servicios***, y para implementar las acciones determinadas en el capítulo 6, mediante: La ***salida*** de esta planificación ***debe*** ser adecuada para las operaciones de la organización. La organización ***debe*** controlar los cambios planificados y revisar las consecuencias de los cambios no previstos, tomando acciones para mitigar cualquier efecto adverso, según sea necesario. La organización ***debe*** asegurarse de que los procesos contratados externamente estén controlados (véase 8.4)."	

Tabla 1 – Sintaxis para identificar los requisitos en ISO 9001:2015

Por otra parte, Scrum es un enfoque ágil para la gestión de proyectos, se utiliza en el desarrollo de software y se centra principalmente en: la interacción individual, la producción de valor empresarial y la satisfacción del cliente (Rubin, 2012). A diferencia de ISO 9001, Scrum solo cubre la gestión del equipo y del producto software en una organización, no hay una compañía específica y responsable de mantener Scrum, los creadores Ken Schwaber y Jeff Sutherland decidieron que Scrum fuera libre para cualquier organización que quisiera implementarlo. Además, los costos de implementación de Scrum son relativamente bajos, y según (Schwaber & Sutherland, 2017) y (Rubin, 2012), aunque Scrum es fácil de entender, puede ser difícil de aplicar, en el sentido que cambia paradigmas y la cultura de la organización.

Al llevar a cabo el análisis de estructura y terminología de la guía de Scrum (Schwaber & Sutherland, 2017), se pueden identificar las siguientes secciones del documento: propósito, definición, usos, teoría, valores, equipo, eventos y artefactos. Además, fue posible también concluir que la guía no tiene una sintaxis estándar o común para identificar lo que es un requisito para el cumplimiento de Scrum, aunque hay secciones que dan a entender que algo se debe hacer de cierta manera, la sintaxis no se repite para todos los elementos del enfoque. Lo único que sugieren como obligatorio, es que la implementación de Scrum debe ser completa, y que su implementación parcial no puede considerarse Scrum.

Partiendo del análisis de la estructura realizada en ISO 9001 y Scrum, es posible encontrar elementos comunes, esto facilita su comparación e integración. La correspondencia de los elementos del proceso se soportó con la ayuda de la experiencia y el criterio de investigadores expertos, lo que permitió disminuir la subjetividad de las relaciones identificadas, las cuales fueron revisadas una a una por los expertos, si surgía alguna sugerencia, esta era documentada y notificada para luego ser realizada y actualizar el trabajo realizado. La Tabla 2 muestra la comparación de alto nivel de los elementos encontrados en el análisis de las estructuras de ambas soluciones.

#	Elementos sugeridos	ISO 9001:2015	Scrum
1	**Grupo de procesos**	Capítulos a través de numerales	*-No describe-*
2	**Procesos**	*-No describe-*	*-No describe-*
3	**Actividades**	Inherente en el texto	Describe las actividades como eventos
4	**Tareas**	*-No describe-*	*-No describe-*
5	**Roles**	*-No describe-*	Claros en el texto
6	**Herramientas**	*-No describe-*	*-No describe-*
7	**Artefactos**	*-No describe-*	Claros en el texto
8	**Objetivos**	Claros en el texto	Claros en el texto
9	**Métricas**	*-No describe-*	Claros en el texto

Tabla 2 – Estructuras de Scrum e ISO 9001:2015

3.3. Realizando la comparación

Para llevar a cabo la comparación, se utilizó el método CoMethod definido en (Pino et al., 2010), que fue adaptado para que permitiera efectuar una comparación desde una perspectiva de alto nivel entre los elementos comunes (ver Tabla 2) de ambas soluciones, y poder definir si una solución contiene los elementos de otra. Las entidades a ser comparadas entre las soluciones, son las actividades que se deben llevar a cabo según los enunciados de los numerales de los capítulos 8, 9 y 10 de ISO 9001, y las actividades definidas en los eventos de Scrum, llevadas a cabo por uno o más roles y cuyas salidas son artefactos de Scrum.

Para establecer el grado de relación/conformidad entre ISO 9001 y Scrum, se tuvo en cuenta que un capítulo del estándar (CH_i) está compuesto por un número de requisitos (R), y cada requisito tiene un conjunto de actividades (R_j). Asimismo, la comparación o relación debe llevarse a cabo a nivel de los capítulos del estándar vs todos los elementos de Scrum, es decir, un requisito de un capítulo de ISO 9001:2015 vs evento, rol y/o artefacto de Scrum. De tal manera que; si al analizar las actividades de un requisito en el estándar se determina que éstos pueden estar relacionados o ser equivalentes a un evento, rol y/o artefacto de Scrum, el requisito tendrá una relación de 1 a 1 con Scrum, es decir, que la conformidad de Scrum (*Cs*) con respecto al requisito R_j será igual a 1. Por otro lado, si las actividades del requisito R_j no está relacionada a un evento, rol y/o artefacto de Scrum, *Cs* con respecto al requisito R_j será igual a 0 (cero). Posteriormente, la sumatoria de las conformidades de cada requisito del capítulo CH_i será dividida entre el total de requisitos R del capítulo CH_i del estándar; por lo tanto, como se muestra en la Tabla 3, el grado de relación entre un capítulo del estándar y Scrum será expresado porcentualmente como C_s con respecto al capítulo CH_i. Además, para expresar la conformidad de Scrum respecto a un capítulo del estándar (C_sCH_i), se utilizará una escala de comparación discreta de cuatro puntos como se muestra en la Tabla 3.

No.	Ecuación para el cálculo del grado de relación de la comparación	
1	$C_sCH_i = \frac{\sum_{j=1}^{R} C_sR_j}{R}$	Donde: C_sCH_i es la conformidad de Scrum con respecto al capítulo *i* de ISO 9001:2015, C_sR_j es la conformidad de Scrum con respecto al requisito *j, y R* es la cantidad de requisitos del capítulo *i*.

Escala discreta de comparación	
Escala cualitativa	**Descripción del rango cuantitativo**
Conformidad total	Scrum soporta entre 85.1% y el 100% de los numerales del capítulo.
Conformidad alta	Scrum soporta entre el 50.1% y el 85% de los numerales del capítulo.
Conformidad parcial	Scrum soporta entre el 15.1% y el 50% de los numerales del capítulo.
No conforme	Scrum soporta entre el 0% y el 15% de los numerales del capítulo.

Tabla 3 – Ecuación y escala de comparación

Teniendo en cuenta que ISO 9001 es un enfoque tradicional que define 36 requerimientos de obligatorio cumplimiento, y que Scrum es un enfoque ágil compuesto de eventos, roles y artefactos que generalmente son adaptados por las organizaciones para su implementación, se tomó como modelo base a ISO 9001, esto, con el objetivo de buscar conformidad con los requerimientos en los diferentes elementos de Scrum. La identificación del grado de relación permitirá que organizaciones Scrum e ISO 9001 puedan ampliar sus certificaciones y aprovechar los beneficios de ambas soluciones.

3.4. Llevando a cabo la comparación

La comparación se llevó a cabo de acuerdo a su diseño, en donde el *Intérprete* certificado como Scrum Master y auditor interno de ISO 9001:2015 llevó a cabo el análisis de las 396 posibles relaciones entre los 36 requisitos de los capítulos 8, 9 y 10 del estándar ISO y los 11 elementos de Scrum, donde se identificó un total de 133 relaciones, que fueron a su vez validadas y depuradas por 2 *revisores*, 1 de ellos es un supervisor con amplia experiencia en la armonización y generación de modelos híbridos, y la segunda, es una *revisora* experta en implementar SGC con base en el estándar ISO 9001. La Tabla 4 presenta las relaciones 1 a 1 entre los requisitos de ISO 9001:2015 (filas) y Scrum (columnas) a través de la etiqueta x_i, donde *i* es un consecutivo único para cada relación.

		Eventos, roles y artefactos de Scrum											
ISO 9001:2015		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	
Capítulos	**Numeral**	**Sprint Planning**	**Daily Sscrum**	**Sprint Review**	**Sprint Retrospective**	**Product Owner**	**Scrum Master**	**Development Team**	**Product Backlog**	**Sprint Backlog**	**Increment**	**Definition of Done**	**#**
Capítulo 8: Operación	*8.1*	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	-	x_{10}	1
	8.2.1	-	-	x_{11}	-	x_{12}	-	-	x_{13}	-	x_{14}	-	2
	8.2.2	-	-	-	-	-	-	-	-	-	-	-	3
	8.2.3.1	-	-	-	-	-	-	-	-	-	-	-	4
	8.2.3.2	-	-	-	-	-	-	-	-	-	-	-	5
	8.2.4	x_{15}	-	x_{16}	-	x_{17}	x_{18}	x_{19}	x_{20}	-	-	-	6
	8.3.1	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}	x_{31}	7
	8.3.2	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	x_{37}	x_{38}	x_{39}	x_{40}	x_{41}	x_{42}	8
	8.3.3	x_{43}	-	-	-	x_{44}	-	-	x_{45}	x_{46}	x_{47}	x_{48}	9
	8.3.4	x_{49}	-	x_{50}	-	x_{51}	x_{52}	x_{53}	x_{54}	x_{55}	x_{56}	x_{57}	10
	8.3.5	-	-	x_{58}	-	x_{59}	x_{60}	x_{61}	x_{62}	x_{63}	x_{64}	x_{65}	11
	8.3.6	x_{66}	-	x_{67}	-	x_{68}	-	-	x_{69}	-	-	-	12
	8.4.1	-	-	-	-	-	-	-	-	-	-	-	13
	8.4.2	-	-	-	-	-	-	-	-	-	-	-	14
	8.4.3	-	-	-	-	-	-	-	-	-	-	-	15
	8.5.1	-	-	-	-	-	-	-	-	-	-	-	16
	8.5.2	-	-	x_{70}	-	x_{71}	x_{72}	x_{73}	x_{74}	x_{75}	x_{76}	-	17
	8.5.3	-	-	-	-	-	-	-	-	-	-	-	18
	8.5.4	-	-	x_{77}	-	x_{78}	x_{79}	x_{80}	x_{81}	x_{82}	x_{83}	-	19
	8.5.5	-	-	-	-	-	-	-	-	-	-	-	20
	8.5.6	-	-	-	-	-	-	-	-	-	-	-	21
	8.6	-	-	x_{84}	-	x_{85}	x_{86}	-	x_{87}	x_{88}	x_{89}	x_{90}	22
	8.7.1	x_{91}	-	x_{92}	-	x_{93}	x_{94}	x_{95}	x_{96}	x_{97}	-	x_{98}	23
	8.7.2	-	-	x_{99}	-	x_{100}	x_{101}	x_{102}	x_{103}	x_{104}	-	-	24

		Eventos, roles y artefactos de Scrum											
ISO 9001:2015		1	2	3	4	5	6	7	8	9	10	11	
Capítulos	**Numeral**	**Sprint Planning**	**Daily Sscrum**	**Sprint Review**	**Sprint Retrospective**	**Product Owner**	**Scrum Master**	**Development Team**	**Product Backlog**	**Sprint Backlog**	**Increment**	**Definition of Done**	**#**
Capítulo 9: Evaluación del desempeño	*9.1.1*	-	-	-	-	-	x_{105}	x_{106}	x_{107}	x_{108}	-	-	25
	9.1.2	-	-	x_{109}	x_{110}	-	-	-	-	-	-	-	26
	9.1.3	-	-	x_{111}	x_{112}	x_{113}	x_{114}	x_{115}	x_{116}	x_{117}	-	-	27
	9.2.1	-	-	-	-	-	-	-	-	-	-	-	28
	9.2.2	-	-	-	-	-	-	-	-	-	-	-	29
	9.3.1	-	-	-	-	-	-	-	-	-	-	-	30
	9.3.2	-	-	x_{118}	-	x_{119}	x_{120}	-	x_{121}	x_{122}	-	-	31
	9.3.3	-	-	-	-	-	-	-	-	-	-	-	32
Capítulo 10: Mejora	*10.1*	-	-	-	-	-	-	-	-	-	-	-	33
	10.2.1	x_{123}	-	-	-	x_{124}	x_{125}	x_{126}	x_{127}	x_{128}	x_{129}	-	34
	10.2.2	-	-	-	x_{130}	x_{131}	-	x_{132}	x_{133}	-	-	-	35
	10.3	-	-	-	-	-	-	-	-	-	-	-	36

Tabla 4 – Relaciones identificadas en la comparación

En la Tabla 5 se calcula la conformidad de Scrum con respecto a los capítulos 8, 9 y 10 del estándar según la ecuación no. 1 presentada en la Tabla 3. En este sentido, y a manera de ejemplo, en el capítulo 8 hay un total de 24 requisitos (*R*), de los cuales, sólo 14 de ellos están soportados por Scrum (C_s) y 10 no presentan ninguna relación (R_j), respecto a éstos últimos significa que el C_s es cero. Dado que hay 14 requisitos en los que sí se presentaron relaciones, el C_s para ellos es 1, en ese sentido $C_sCH_8 = [(10*0) + (14*1)]/24 = 14/24 = 58\%$, es decir que Scrum soporta con una conformidad alta los requisitos del capítulo 8 del estándar ISO, respecto a los capítulos 9 y 10, los soporta con una conformidad parcial del 50%, respectivamente. Aunque las soluciones describen prácticas diferentes, no son opuestas, sino más bien complementarias, incluso, es posible encontrar similitudes en sus descripciones con niveles de detalle diferente, lo que sugiere que pueden armonizarse e integrarse, e impactar positivamente en: el costo, el tiempo y los recursos asociados a la implementación de una solución híbrida, en vez de la implementación por separado e individual de cada una de ellas.

Capítulo	Ecuación	Porcentaje de conformidad	Calificación
Capítulo 8	$C_sCH_8 = \frac{\sum_{j=1}^{24} C_sR_j}{24} = \frac{(10*0)+(14*1)}{24} = \frac{14}{24}$	$C_sCH_{8=}58\%$	Conformidad alta
Capítulo 9	$C_sCH_9 = \frac{\sum_{j=1}^{8} C_sR_j}{8} = \frac{(4*0)+(4*1)}{8} = \frac{4}{8}$	$C_sCH_{9=}50\%$	Conformidad parcial
Capítulo 10	$C_sCH_{10} = \frac{\sum_{j=1}^{4} C_sR_j}{4} = \frac{(2*0)+(2*1)}{4} = \frac{2}{4}$	$C_sCH_{10=}50\%$	Conformidad parcial

Tabla 5 – Conformidad de Scrum respecto a los capítulos 8, 9 y 10 de ISO 9001

3.5. Analizando la no relación en la comparación

A continuación, en la Tabla 6 se analizan los resultados de la comparación donde no se encontró relación entre ISO 9001 y Scrum. Asimismo, se presentan algunas consideraciones y alternativas para lograr la conformidad.

Capítulo	Requisito	Ausencia de relación desde Scrum	Cómo lograr conformidad
Capítulo 8: Operación	8.2.2, 8.2.3.1 y 8.2.3.2	Scrum no define eventos para la definición de los requisitos del producto, asume que existen y alimentan el *Product Backlog*.	La organización debe definir un proceso para la definición y revisión de los requisitos de sus productos y servicios, aquí puede hacer uso de historias de usuario y criterios de aceptación.
	8.4.1, 8.4.2 y 8.4.3	Scrum no aborda el tema de servicios externos.	La organización debe definir los procesos necesarios para controlar los productos y servicios suministrados externamente.
	8.5.1	Aunque en Scrum un *Sprint* proporciona entregables funcionales de producto/ servicio en el artefacto *Incremento*, éste no sugiere la forma en que se realiza la provisión del servicio a los clientes, tampoco describe los elementos relacionados a: infraestructura, recursos para el seguimiento y medición, personal, actividades de liberación y posteriores a la entrega.	La organización debe definir un proceso adecuado para la gestión de la provisión del servicio.
	8.5.3	ISO 9001 abarca mucho más que Scrum con relación a la propiedad perteneciente a los clientes o proveedores externos, incluyendo la identificación, verificación y protección de la propiedad del cliente durante la provisión y prestación del servicio.	La organización debe definir un proceso que permita garantizar la seguridad de la información de la propiedad perteneciente a los clientes o proveedores externos durante la provisión y prestación del servicio.
	8.5.5 y 8.5.6	Scrum se enfoca en la construcción del producto más no en el mantenimiento y prestación del servicio.	La organización debe definir un proceso para la gestión de la provisión del servicio.

Capítulo	Requisito	Ausencia de relación desde Scrum	Cómo lograr conformidad
Capítulo 9: Evaluación del desempeño	9.2.1, 9.2.2, 9.3.1 y 9.3.3	Aunque Scrum provee insumos relacionados con la entrega de valor como medida principal de la satisfacción del cliente, no tiene eventos que realicen evaluaciones de desempeño y satisfacción del cliente de manera específica.	La organización debe definir los procesos necesarios con relación a la realización de auditorías internas y revisiones por la dirección.
Capítulo 10: Mejora	10.1 y 10.3	Aunque el evento *Retrospectiva del Sprint* aborda la autoevaluación y mejora continua del proceso de diseño y desarrollo, se requiere un proceso formal para determinar y seleccionar las oportunidades de mejora del SGC.	La organización debe definir un proceso para determinar y seleccionar las oportunidades de mejora del SGC.

Tabla 6 – Nomenclatura de comparación

3.6. Integración de Scrum e ISO 9001:2015

A partir de los resultados obtenidos, se diseñó una propuesta de proceso que integra la gestión de proyectos de software basado en Scrum y los requisitos comparados del estándar ISO 9001:2015. El proceso definido sugiere 4 procesos: Diseño y Desarrollo, Planeación del Sprint, Ejecución del Sprint y Fabricación del Código. Para organizar la información, se utilizó una versión ajustada de la plantilla de proceso de COMPETISOFT presentado en (Piattini et al., 2007), que permite describir de manera detalla la información de los procesos, así como: *propósito, descripción del proceso, objetivos, indicadores, subprocesos, roles involucrados, actividades, entradas, salidas y roles relacionados a las actividades*. Asimismo, se ha utilizado la notación de modelamiento de procesos de negocio (BPMN) para diseñar los diagramas de flujo de cada uno de los procesos. En las descripciones de los procesos se conservan algunos nombres originales en inglés de los elementos de proceso, debido a que son ampliamente utilizados y entendidos. La versión completa que permite observar los procesos y sus descripciones está disponible en: https://bit.ly/2V4CEwM. Debido al límite de espacio, a continuación, en la Tabla 7 sólo se presenta un extracto al proceso propuesto: *Ejecución del Sprint*.

Descripción general del proceso			
Proceso	**Ejecución del Sprint**		
Propósito	El propósito de este proceso es construir el Incremento del Producto utilizando las practicas ágiles de SCRUM y la rigurosidad de ISO 9001:2015 para garantizar la calidad y la satisfacción del cliente.		
Descripción	Este proceso define un flujo general para la ejecución de un sprint, que dura por lo general entre 2 y 4 semanas. Contiene actividades de seguimiento como la reunión diaria, solución de impedimentos, pruebas de caja negra y además contiene al sub proceso Fabricación del Código que encapsula todo lo relativo al desarrollo del producto software.		
Objetivos	Definidos implícitamente	***Indicadores***	Velocidad de equipo
Responsabilidad	SCRUM Team	***Subprocesos***	Fabricación del Código

Descripción general del proceso		
Entradas	**Nombre**	**Fuente**
	Sprint Backlog	Proceso de Planeación del Sprint
Salidas	**Nombre**	**Fuente**
	Incremento del Producto	Proceso de Diseño y Desarrollo
Roles involucrados	**Rol**	**Descripción**
	SM - SCRUM Master, PO - Product Owner, DT – Development Team	Ver el link: https://bit.ly/2K1JbSE
Actividades		
Rol	**Descripción**	
A1. Participar/asistir a reunión diaria.		
Entradas	Ninguna	
SM, PO, DT	El Product Owner, el SCRUM Master y el Development Team se reúnen cada día durante 15 minutos. El Product Owner puede asistir, pero no debe participar, esto para mantener al equipo libre de presiones y enfocado totalmente en el Sprint Backlog. En cada reunión los miembros del equipo responden a las preguntas ¿Qué hiciste desde la última reunión? ¿Qué harás hasta la próxima reunión? ¿Qué impedimentos has encontrado? El SCRUM Master documenta la reunión en el Acta de reunión.	
Salidas	Acta de reunión Impedimentos	

Tabla 7 – Extracto del proceso de Ejecución del Sprint

3.7. Limitaciones

Se han identificado algunas limitaciones después de aplicar la estrategia de armonización a Scrum y la norma ISO 9001:2015, algunas de ellas se describen a continuación:

La exactitud en la interpretación de los resultados desde el análisis teórico. Un riesgo relacionado a este aspecto es la subjetividad involucrada en el manejo de eventos y datos desarrollados por el rol *Intérprete*, y su conocimiento y comprensión de las soluciones armonizadas. Hemos intentado minimizar esta amenaza incorporando un rol adicional de "Revisor", que continuamente cubrió y monitoreó los métodos y el proceso de armonización. Sin embargo, esto puede no ser suficiente. Por lo tanto, tenemos la intención de tener más de una persona cubriendo cada rol para disminuir la subjetividad, por ejemplo: dos intérpretes y dos revisores podrían trabajar de forma independiente para evitar la influencia o el sesgo, y después de completar sus actividades, pueden fusionar sus resultados.

Revisión por expertos. Es posible que algunas relaciones se pierdan o incluso sean incorrectas. En este sentido, y para enriquecer el trabajo realizado, se incluyó la opinión y experiencia de expertos con consultas acerca de las relaciones e integraciones realizadas, en un futuro, las revisiones podrían incluir consultas reiteradas aplicadas a través de métodos formales como el método Delphi, entre otros, aplicado a través de un método empírico.

Criterios predefinidos para apoyar la toma de decisiones durante la integración. Aunque se utilizó un conjunto de criterios para apoyar la toma de decisiones durante la integración de las prácticas de las soluciones armonizadas, es posible que se requiera un criterio más detallado que permita abordar situaciones de integración específicas, por ejemplo: qué hacer y qué elementos deben integrarse si el contenido de una práctica de la solución A es más detallado que una práctica relacionada con la solución B, pero el contenido de B es mejor que el de A. En este sentido, será necesario abordar otras situaciones posibles, esto podría facilitar la toma de decisiones, incluso también reducir el tiempo involucrado en la integración. Sin embargo, esta reducción en el tiempo es actualmente solo una hipótesis que espera el rechazo o la confirmación en nuestro trabajo futuro.

4. Conclusiones

En este artículo se ha presentado la estrategia de armonización diseñada para soportar la combinación del estándar ISO 9001:2015 y el enfoque ágil Scrum. Para llevar a cabo la armonización de estas soluciones, se configuró una estrategia que estuvo conformada por tres métodos: homogeneización, comparación e integración, éstos permitieron principalmente: (i) solucionar las diferencias estructurales; (ii) comparar los elementos que describen las soluciones; e (iii) integrar los elementos equivalentes para soportar el cumplimiento del estándar ISO 9001 y Scrum.

Tanto Scrum como ISO 9001:2015 describen requisitos y prácticas para solucionar necesidades específicas, por un lado, Scrum provee un conjunto de sugerencias y elementos a tener en cuenta para mejorar la gestión de proyectos a través de un enfoque ágil, por su parte, el estándar ISO describe los requisitos a tener en cuenta para establecer un SGC en las organizaciones. Con la armonización se ha podido observar que existe una relación alta y parcial entre Scrum y el estándar ISO 9001, este es del 58% para el capítulo 8, y del 50% para el capítulo 9 y 10, respectivamente. Esto significa que hay 133 relaciones en las que Scrum ofrece algún tipo de soporte para los requisitos de ISO 9001:2015. Esto sugiere que si bien, las soluciones definen prácticas para diferentes necesidades, éstas no son totalmente diferentes y, por lo tanto, es posible encontrar una relación cercana que puede aprovecharse para disminuir costos en su implementación y disfrutar los beneficios de ambas soluciones en los proyectos de una organización. Las relaciones conceptuales establecidas entre las dos soluciones se han identificado bajo los criterios y la experiencia de un *intérprete* responsable del análisis y la comparación de los modelos. Asimismo, se presenta la manera en la que se pueden abordar aquellos requisitos del estándar en donde no se encontró ninguna relación con Scrum. Por otra parte, los resultados obtenidos de la armonización, permitió definir un proceso híbrido que da soporte a la gestión de proyectos de desarrollo de software tomando como base elementos tanto de Scrum como los requisitos de los capítulos 8, 9 y 10 del estándar ISO.

Teniendo en cuenta lo anterior, la armonización de ISO 9001:2015 y Scrum realizada en este trabajo, puede ser de gran beneficio práctico para las organizaciones certificadas en el estándar cuando buscan adoptar los elementos definidos en Scrum, o viceversa, cuando una organización usa Scrum y desea adoptar prácticas para implementar un SGC, por lo menos, en lo que respecta a los capítulos 8, 9 y 10 del estándar ISO 9001.

Como trabajo futuro, se espera llevar a cabo un estudio empírico que permita realizar la comparación de las soluciones utilizando la opinión de varios expertos y/o profesionales involucrados en el uso de Scrum e ISO 9001 en algunas organizaciones. Esta validación permitiría analizar en un dominio la correspondencia o relación entre estas dos soluciones, no solo desde un punto de vista teórico, sino también desde un punto de vista empírico y práctico.

5. Agradecimientos

El profesor César Pardo agradece la contribución de la Universidad del Cauca, donde trabaja como profesor Asociado.

Referencias

Calabrese, J., Esponda, S., Boracchia, M., & Pesado, P. (2019). Scrum Towards IRAM-ISO 9001:2015. Integrating Documentation Required. In P. Pesado & C. and Aciti (Eds.), *Computer Science – CACIC 2018,* (pp. 183–196). Cham: Springer International Publishing.

Fernandez, D. J., & Fernandez, J. D. (2008). Agile Project Management —Agilism versus Traditional Approaches. *Journal of Computer Information Systems*, *49*(2), 10–17.

Grimes, K. (2002). *ISO 9001:2000 A Practical Quality Manual Explained*. New York, NY: Quality Press.

ISO. (2015). *ISO 9001:2015 Quality management systems requirements* (5th edition). ISO editors.

Jakobsen, C., & Sutherland, J. (2009). *Scrum and CMMI going from good to great*. Retrieved from: https://doi.org/10.1109/AGILE.2009.31

Kuhrmann, M., Diebold, P., Münch, J., Tell, P., Garousi, V., Felderer, M., Trektere, K., McCaffery, F., Linssen, O., Hanser, E., & Prause, C. R. (2017). Hybrid Software and System Development in Practice: Waterfall, Scrum, and Beyond. In *Proceedings of the 2017 International Conference on Software and System Process*, (pp. 30–39). New York, NY, USA: Association for Computing Machinery.

Larman, C. (2003). *Agile and Iterative Development: A Manager's Guide* (1st edition). New York, NY: Addison-Wesley Professional.

Nawrocki, J. R., Jasiński, M., Walter, B., & Wojciechowski, A. (2002). Combining Extreme Programming with ISO 9000. In *EurAsia-ICT 2002: Information and Communication Technology*, (pp. 786–794). Heidelberg, Berlin: Springer.

Pardo, C, Garcia, F., Piattini, M., Pino, F. J., Lemus, S., & Baldassarre, M. T. (2016). Integrating multiple models for definition of IT governance model for banking ITGSM. *International Business Management*, 10(19), 4644-4653.

Pardo, César, García, F., Piattini, M., & Pino, F. (2012). *A Framework to Support the Harmonization between Multiple Models and Standards*. (Doctoral Thesis in Computer Sciences), Computer's department, University of Castilla-La Mancha, Spain.

Pardo, César, Pino, F., & García, F. (2016). Towards and integrated management system (IMS), harmonizing the ISO/IEC 27001 and ISO/IEC 20000-2 standards. *International Journal of Software Engineering and Its Applications*, 10(9), 217–230.

Pardo, César, Pino, F. J., Garcia, F., Baldassarre, M. T., & Piattini, M. (2013). From chaos to the systematic harmonization of multiple reference models: A harmonization framework applied in two case studies. *Journal of Systems and Software*, 86(1), 125–143.

Petersen, K., & Wohlin, C. (2010). The effect of moving from a plan-driven to an incremental software development approach with agile practices. *Journal of Empirical Software Engineering*, 15(6), 654–693.

Piattini, M., Pino, F., Alquicira, C., & Ruiz, F. (2007). Software Process Improvement: The COMPETISOFT Project. *IEEE Computer Journal*, 40(10), 21–28.

Pino, F. J., Baldassarre, M. T., Piattini, M., & Visaggio, G. (2010). Harmonizing maturity levels from CMMI-DEV and ISO/IEC 15504. *Journal of Software Maintenance and Evolution: Research and Practice*, 22(4), 279–296.

Popovic, T. (2015). Getting ISO 9001 Certified for Software Development Using Scrum and Open Source: A Case Study. *Tehnicki Vjesnik Journal*, 22(1), 1633–1640.

Qasaimeh, M., & Abran, A. (2013). An Audit Model for ISO 9001 Traceability Requirements in Agile-XP Environments. *Journal of Software*, 8(7), 1556–1567.

Qasaimeh, M., & Abran, A. (2010). Investigation of the Capability of XP to Support the Requirements of ISO 9001 Software Process Certification. In *Proceedings of the Eighth ACIS International Conference SERA*, 239–247. Montreal, QC, Canada: IEEE Computer Society Press.

Royce, W. W. (1987). Managing the Development of Large Software Systems: Concepts and Techniques. In *Proceedings of the 9th International Conference on Software Engineering*, 328–338. Monterey, California, USA: IEEE Computer Society Press.

Rubin, K. (2012). *Essential Scrum: A Practical Guide to the Most Popular Agile Process* (1st edition). New York, NY: Addison-Wesley Professional.

Schwaber, K., & Sutherland, J. (2017). *The Scrum Guide*. The Definitive Guide to Scrum: The Rules of the Game. Retrieved from: https://bit.ly/2RHk64q

Shen, B., & Ruan, T. (2008). A Case Study of Software Process Improvement in a Chinese Small Company. *2008 International Conference on Computer Science and Software Engineering*, (pp. 609–612). Hubei, China: IEEE Computer Society Press.

Silva, J. G. A. O. B. da, & Cunha, P. R. da. (2006). Reconciling the Irreconcilable? A Software Development Approach that Combines Agile with Formal. *Proceedings of the 39th Annual Hawaii International Conference HICSS'06*, 216b-216b.

Stålhane, T., & Hanssen, G. K. (2008). The Application of ISO 9001 to Agile Software Development. In International Conference on Product Focused Software Process Improvement - *PROFES 2008*, (pp. 371–385). Berlin, Heidelberg: Springer.

Vriens, C. (2003). Certifying for CMM Level 2 and IS09001 with XP@Scrum. In *Proceedings of the Agile Development Conference*, (pp. 120–124). Salt Lake City, UT, USA: IEEE Computer Society Press.

Wright, G. (2003). Achieving ISO 9001 Certification for an XP Company. In *Extreme Programming and Agile Methods - XP/Agile Universe*, (pp. 43–50). Berlin, Heidelberg: Springer.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 21/08/2020
Aceitação/Acceptance: 19/10/2020

SCMOnto: Una ontología para soportar la gestión de la configuración de software

Carlos Orozco[1], César Pardo[1], Sebastián Vásquez[1]

carlosorozco@unicauca.edu.co; cpardo@unicauca.edu.co; jsvasquez@unicauca.edu.co

[1] Grupo de investigación GTI, Universidad del Cauca, Carrera 2º Calle 15n Esquina Campus Tulcán 19001, Popayán, Colombia.

Pages: 75-90

Resumen: Actualmente, existe un amplio portafolio de soluciones para la gestión de la configuración de software (GCS) así como: IEEE 828-2012, CMMI-DEV, ISO 10007:2003 e ISO/IEC 29110, entre otros. Sin embargo, las empresas han realizado esfuerzos de mejora sin tener una terminología común que permita eliminar la confusión y conflictos terminológicos que afectan la adopción e implementación de estas soluciones. Este artículo presenta la ontología SCMOnto, la cual fue definida utilizando el formalismo REFSENO y cuyo objetivo es sugerir una terminología común y coherente que permita compartir el conocimiento generado en torno a la GCS en los procesos de software de una empresa de manera genérica, formal y sin ambigüedad. Asimismo, se presenta la evaluación satisfactoria de SCMOnto en tres casos de aplicación. La ontología propuesta podría ser de gran utilidad para las empresas y consultores que planean llevar a cabo el análisis, implementación o evaluación del proceso de GCS.

Palabras-clave: Gestión de configuración de Software (GCS); CMMI; IEEE 828-2012; ISO 10007:2003; ISO/IEC 29110.

SCMOnto: An ontology to support software configuration management

Abstract: Currently, although there is a wide portfolio of solutions for software configuration management (SCM) as well as: IEEE 828-2012, CMMI-DEV, ISO 10007:2003 and ISO/IEC 29110, among others. However, companies have made improvement efforts without having a common terminology that eliminates the confusion and terminological conflicts that affect the adoption and implementation of these solutions. This article presents the SCMOnto ontology, which was defined using the REFSENO formalism and whose objective is to suggest a common and coherent terminology that allows sharing the knowledge generated around the SCM in the software processes of a company in a generic, formal and unambiguously way. Likewise, the satisfactory evaluation of SCMOnto in three application cases is presented. The proposed ontology could be very useful for organizations and consultants who plan to carry out the analysis, implementation or evaluation of the SCM process.

Keywords: Software configuration management (GCS); CMMI; IEEE 828-2012; ISO 10007:2003; ISO/IEC 29110.

1. Introducción

Actualmente, las empresas de software deben afrontar múltiples desafíos relacionados a los cambios constantes que suceden durante la ejecución de los proyectos que llevan a cabo, por lo cual, es necesario adoptar soluciones que faciliten el control y seguimiento sistemático de las actividades relacionadas a la gestión de los cambios en los artefactos involucrados durante el proceso de desarrollo con el propósito de mantener su integridad y trazabilidad, así como: cambios en la documentación de análisis y diseño, control de las versiones generadas durante la fase de implementación y pruebas del software, gestión y control de los cambios durante el despliegue y el mantenimiento de la solución, entre otros (Perera et al, 2013). En este sentido, la adopción de buenas prácticas que faciliten la gestión de la configuración en los proyectos software se hace necesaria y crucial (Ben Menachem, 1995).

Teniendo en cuenta lo anterior, actualmente existen soluciones que permiten soportar al proceso de GCS, algunas de ellas son: IEEE 828-2012 (IEEE, 2012), CMMI-DEV 2.0 (CMMI Institute, 2020), ISO 10007:2003 (ISO 10007, 2003) e ISO 33000 (ISO, 2020), estas soluciones (en adelante modelos de referencia - MR) son aplicadas por las empresas como el medio para mejorar su proceso de desarrollo, alcanzar un grado de capacidad o llegar a un nivel madurez exigido por auditores externos. Las empresas con suficiente capital y recursos humanos optan por aplicar modelos como IEEE 828-2012 o CMMI-DEV 2.0. Por otro lado, las micro, pequeñas y medianas empresas (Mincomercio, 2020) desarrolladoras de software (MiPyMEs_DS) que son limitadas en aspectos relacionados a: su capacidad económica, talento humano y tiempo (Regalado Hernández, 2007), lo hacen mediante la aplicación de modelos de facto más ligeros como el estándar ISO/IEC 29110 (ISO/IEC 29110-2-1, 2015), o el uso de enfoques ágiles que no describen practicas específicas para la GCS y que deben ser complementados, por ejemplo: Scrum, Kanban, XP, SAFe, LeSS, entre otros.

La gran variedad emergente de MR contribuye a que las empresas de software puedan evaluar e institucionalizar nuevos y/o mejorados procesos, aumentando su competitividad y produciendo productos de mayor calidad. Adicionalmente, esto les permite elegir un modelo en particular para cubrir un problema específico o seleccionar varios modelos para abordar diferentes necesidades (Ortega, Pardo & Pino, 2019). No obstante, actualmente las empresas buscan mejorar continuamente mediante la adopción de uno o varios modelos con el objetivo de obtener un grado de capacidad y madurez que resulten en una mejora de sus procesos internos. Según (Pardo, Pino, García, Piattini & Baldassarre, 2011), los factores que pueden persuadir a un empresa para considerar la necesidad de trabajar con más de un modelo son: (i) nichos de mercado con modelos específicos, (ii) mejora de procesos legados, (iii) posicionamiento comercial, (iv) apalancamiento o fusión corporativa, (v) búsqueda sistemática de la capacidad de sus procesos, (vi) crecimiento del negocio, entre otros. Sin embargo, las empresas han experimentado dificultades para trabajar con más de un modelo al mismo tiempo y a menudo hacen un gran esfuerzo para interpretarlos, esto, debido a que cada modelo se ha definido a partir de diferentes: opiniones y puntos de vista, grupos de trabajo (culturales y políticos), intereses y organismos de GCS (Wang & King, 2000). En este sentido, cada uno de los modelos de referencia propone su propia terminología, estructura y elementos a instanciar, por lo que existe cierto grado de heterogeneidad y

ambigüedad en las definiciones, conceptos y relaciones, perspectiva de calidad, alcance, orientación, propósito y otras características presentadas por cada MR con respecto a los demás (Ortega et al, 2019).

Teniendo en cuenta lo anterior, el objetivo de este artículo es presentar una solución al problema a través de una ontología llamada SCMOnto, que describe de manera genérica los conceptos y las relaciones comunes a considerar en el proceso de GCS, en este sentido, la ontología puede ser utilizada independientemente del MR que utilice o desee adoptar una empresa, reduciendo la ambigüedad de los conceptos y términos en ese dominio, y permitiendo que los interesados puedan reutilizar y compartir el conocimiento en el área. Asimismo, SCMOnto promueve y facilita la evaluación del cumplimiento del proceso de software de GCS a través de la identificación de las relaciones entre los elementos propuestos. Por otro lado, SCMOnto usa, amplía y adapta algunos de los conceptos de diferentes MR para la GCS, esto con el objetivo de presentar una solución unificada, homogénea y sin ambigüedad, lo que resulta útil y facilita su adopción, esto se ha logrado mediante la armonización de los elementos propuestos por diferentes MR relacionados con la GCS, entre ellos: IEEE 828-2012, CMMI-DEV 2.0, ISO 10007:2003, ISO 33000 y algunas sub-ontologías: ontología de GCS propuesta por (Arantes, Falbo & Guizzardi, 2007), la ontología para MR de procesos PrMO (Pardo, García, & Piattini, 2014) y la ontología para la armonización de MR presentada en (Pardo, Pino, García, Piattini & Baldassarre, 2012).

Además de la presente introducción, el documento está organizado de la siguiente manera. La Sección 2 muestra los trabajos relacionados. La Sección 3 presenta el método para definir ontologías en la ingeniería de software - REFSENO. La Sección 4 presenta la ontología para la GCS – SCMOnto a nivel de sus conceptos, relaciones y discusión de algunos de sus elementos. La Sección 5 presenta la evaluación de SCMOnto en tres casos de aplicación: a través de la narración de un ejemplo teórico, una instanciación en Protégé usando prácticas de CMMI-DEV 2.0 y la validación cualitativa por expertos profesionales en GCS. Las conclusiones y trabajos futuros se presentan en la Sección 6.

2. Trabajos relacionados

Con el análisis de los trabajos relacionados, fue posible observar propuestas relacionadas al uso de ontologías para soportar la GCS, algunos de ellos fueron: estudio que presenta una ontología genérica de modelos de referencia de procesos denominado PrMO, que define una estructura común de elementos de procesos como medio para apoyar la armonización de diferencias estructurales entre múltiples procesos, entre ellos la gestión de la configuración de software (Pardo et al, 2014), una ontología para soportar la GCS (Arantes et al, 2007). Finalmente, un estudio en el cual se presenta una herramienta que busca facilitar la integración de la GCS durante el proceso de desarrollo de software soportado en una proto-ontología (Nunes & Falbo, 2006). Otros estudios se han centrado principalmente en el desarrollo de ontologías para representar los elementos de dominios particulares: ontologías para representar los modelos ISO y CMMI, por ejemplo: CMMI-SW (Soydan & Kokar, 2008); CMMI e ISO/IEC 15504 (Liao & Qu, 2005) (Salviano & Figueiredo, 2008) e ISO 9001 y CMMI-SW (Ferchini, Bigand & Lefebvre, 2008) y en (Malzahn, 2009) se define una ontología para vincular las similitudes entre varios

modelos de software. Además, en (Mendes & Abran, 2004) se presenta una ontología para el dominio de la ingeniería basado en SWEBOK y otros modelos.

Teniendo en cuenta la situación descrita anteriormente, es posible ver que la mayoría de los estudios se han centrado esencialmente en el desarrollo de ontologías para representar y/o apoyar los elementos clave de dominios particulares, así como comprender la estructura de los enfoques de calidad basados en procesos. Si bien se encontró una propuesta para GCS en (Nunes & Falbo, 2006), esta está desactualizada. Por otra parte, fue posible encontrar que otros estudios también se centran en ontologías de desarrollo para apoyar la integración de procesos de negocio, pero este tema va más allá del alcance de este artículo. En este sentido, en contraste con los trabajos relacionados analizados, la propuesta aquí presentada tiene la intención de proporcionar un nivel más detallado para la GCS.

3. REFSENO como método para la definición de ontologías en SW

Para el diseño y definición de ontologías existen varias metodologías, por ejemplo: Methontology, un enfoque de traducción para las especificaciones de ontologías (Fernández, Gómez & Juristo, 1997), un modelo para la gestión de conocimiento basado en ontologías (Gruber, 1993), el formalismo de representación para ontologías de ingeniería de software - REFSENO (Tautz & Von Wangenheim, 1998), entre otros. Después de realizar la revisión de dichas propuestas y estudiar su posible aplicación para la definición de SCMOnto, se decidió utilizar REFSENO (Tautz & Von Wangenheim, 1998) tomando en cuenta los siguientes criterios: (i) REFSENO, a pesar de estar basada en Methontology, es una metodología ampliamente utilizada para la definición de ontologías de diferentes campos del conocimiento, propone una adaptación específica de aplicación para el diseño de ontologías de ingeniería del software, (ii) REFSENO define diferentes mecanismos de análisis y presentación de la solución a nivel de conceptos, atributos y relaciones. Esto es posible debido a que REFSENO propone tres tablas para representar dichos elementos: glosario de conceptos, atributos y relaciones y (iii) a diferencia de otros enfoques, REFSENO provee numerosas técnicas para el análisis de la consistencia de la ontología e instancias a nivel de implementación. Además, REFSENO permite realizar una caracterización entre los niveles de conocimiento conceptual y de contexto específico; de esta manera, REFSENO representa una alternativa más intuitiva para lectores que no estén familiarizados con lógica de predicado de primer orden o similares.

4. SCMOnto: Una ontología para la GCS

Después de realizar un análisis exhaustivo de los MR existentes, se evidenció que existe heterogeneidad y ambigüedad en sus definiciones, en este sentido, se establecieron los siguientes objetivos para la definición de SCMOnto: (i) ubicación e identificación de sinónimos, homónimos, inconsistencias y conflictos terminológicos e (ii) integración de los conceptos encontrados en la literatura. Estos objetivos se logran con SCMOnto, la cual es una ontología común que representa el dominio de la GCS, y además proporciona: conceptos, términos con definiciones claras y concisas, y sus respectivas relaciones. Estos elementos también se pueden utilizar como base para respaldar los requisitos de valoración y evaluación del proceso de GCS de una empresa independiente del MR que ésta use o desee adoptar.

De acuerdo con la clasificación propuesta por (Ruiz & Hilera, 2006), las ontologías en el área de ingeniería de software se pueden agrupar en dos categorías: (i) ontologías de dominio, aquellas que describen el conocimiento de un dominio particular y (ii) ontologías como artefactos de software, que sirven como artefactos durante el proceso de desarrollo de software. De acuerdo con lo anterior, SCMOnto, es una ontología de dominio dado que permite organizar el conocimiento relacionado con los conceptos y las relaciones en torno a la GCS. A continuación, se presenta una aproximación general de la ontología en términos de su propósito, el glosario de conceptos y su representación gráfica.

4.1. Propósito

SCMOnto permite que aquellos interesados en el desarrollo de software conozcan los conceptos y las relaciones que deben estar presentes en el proceso de GCS, independientemente del enfoque particular que se adopte. SCMOnto se puede usar desde las siguientes perspectivas: (i) *académico*, aquellos interesados en aprender o investigar sobre este tema pueden usar la ontología para conocer los términos y relaciones utilizados en la GCS e (ii) *industrial*, las empresas de software pueden utilizar la ontología para instanciar sus procesos de software y saber en qué medida están presentes los conceptos y las relaciones básicas que se producen en el proceso de GCS. Asimismo, puede usarse como base para apoyar la valoración y evaluación de la GCS de una empresa, por otra parte, SCMOnto también se puede usar como herramienta conceptual para facilitar la armonización de múltiples modelos relacionados con GCS.

4.2. Integración de SCMOnto con otros modelos de GCS

La definición de SCMOnto se llevó a cabo mediante la armonización de los conceptos de MR para GCS, entre ellos: IEEE 828-2012 (IEEE, 2012), CMMI-DEV 2.0 (CMMI

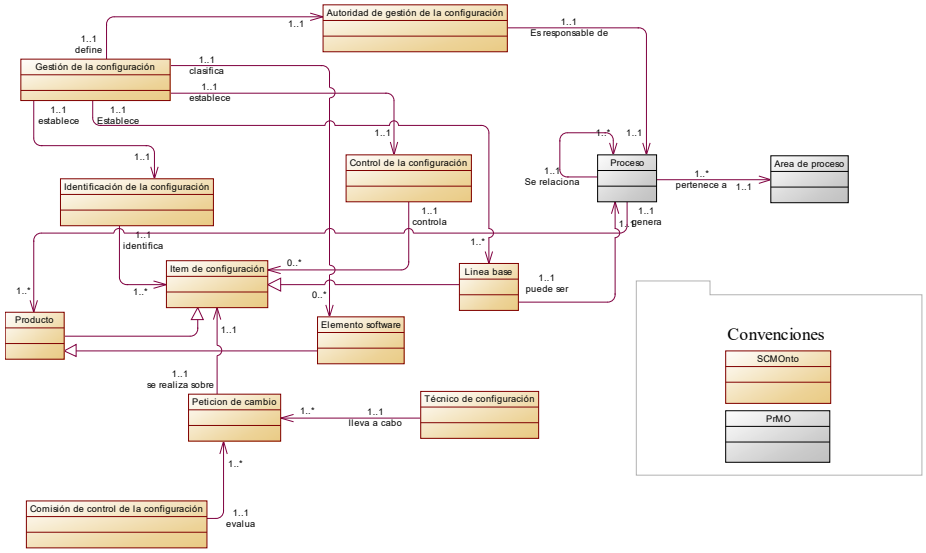


Figura 1 – Extracto de SCMOnto. Fuente propia

Institute, 2020), ISO/IEC 29110 (ISO/IEC 29110-2-1, 2015) y una sub-ontología de GCS que presenta la evolución y actualización de la ontología definida en (Nunes & Falbo, 2006) llamada EvSCMO. Además de los modelos de referencia, SCMOnto utiliza dos sub-ontologías adicionales que aportan conceptos y relaciones entre conceptos y que están estrechamente relacionados a la GCS desde la perspectiva de área de proceso, ellas son: la ontología de GCS (Arantes et al, 2007), la ontología para MR de procesos PrMO (Pardo et al, 2014) y la ontología para la armonización de múltiples estándares y modelos presentada en (Pardo et al, 2012). La Figura 1 presenta los conceptos y las relaciones de SCMOnto mediante la notación de diagramas UML, la convención de colores especifica los conceptos utilizados de otras ontologías, por limitaciones de espacio se presenta un extracto de la ontología, el diagrama completo está disponible en https://bit.ly/2VYN7KH y el código fuente en: https://bit.ly/3barhbL

4.3. Conceptos de SCMOnto

Las definiciones precisas de los conceptos incluidos en SCMOnto son presentados en la Tabla 1, la cual ha sido organizada en cuatro columnas: en la primera y segunda columna se indica el término que será descrito y su súper concepto en caso de ser una especialización de otro concepto definido en SCMOnto, la tercera y cuarta columna muestran la definición del término y la fuente relacionada de donde se extrajo en caso de no ser una definición propia de la propuesta. La Tabla 2 presenta las relaciones entre los conceptos presentados en la Tabla 1.

4.4. Discusión de algunos conceptos y sus relaciones

Algunos de los términos que conforman la propuesta de SCMOnto merecen un análisis especial para clarificar su contexto de utilización e inclusión dentro de esta ontología, tales como: (i) *Producto***,** aunque en (IEEE, 2012) un producto está definido como el conjunto de artefactos a ser desarrollados, entregados y mantenidos en un proyecto, la definición utilizada en SCMOnto ha sido extraída y adaptada de CMMI-DEV 2.0 (CMMI Institute, 2020), debido a que su aplicación dentro del contexto de GCS pudiera brindar una mayor utilidad explicitando que un producto puede ser un entregable a cliente o puede ser utilizado como insumo interno del proyecto, (ii) *Técnico de configuración* es una especialización del concepto *Persona*. Se ha definido como un rol especial para todo aquel miembro de proyecto que se vea involucrado en la modificación, actualización y/o alteración de un *ítem de configuración* tras la aprobación de una *petición de cambio*, (iii) *Autoridad de gestión de la configuración*, a nivel de proyecto, es el rol responsable de la correcta planeación e implementación del proceso de gestión de la configuración. Este rol personifica al experto, es visto como el mentor y orientador del proceso de GCS, y no como la persona encargada de la toma de decisiones por sí mismo y (iv) *Comisión de control de la configuración***,** es otra de las especializaciones del concepto *Persona* que se han definido en SCMOnto. Este rol es desempeñado por un grupo de personas encargadas del estudio de cada una de las *peticiones de cambio* que surjan para un *ítem de configuración* dentro del proyecto; son los responsables de aprobar o rechazar dichas peticiones y de asegurar su realización en caso de resultar aprobadas.

Términos utilizados en SCMOnto			
Término	**Súper concepto**	**Definición**	**Fuente**
Auditoría	Concepto	Revisión objetiva de un producto de trabajo.	(CMMI Institute, 2020)
Auditoría de configuración	Auditoría	Evaluación que garantiza que una línea base cumple con los criterios establecidos por el proceso.	Adaptación (CMMI Institute, 2020)
Auditoría de configuración física	Auditoría de configuración	Auditoría realizada para verificar que un ítem de configuración corresponde a la documentación técnica que lo define.	Adaptación (IEEE, 2012)
Auditoría de configuración funcional	Auditoría de configuración	Verificación que garantiza que el desarrollo de un ítem de configuración ha terminado de manera satisfactoria.	Adaptación (IEEE, 2012)
Autoridad de gestión de la configuración	Persona	Grupo que asegura que las actividades de gestión de la configuración se llevan a cabo correctamente.	(IEEE, 2012)
Base de datos de gestión de la configuración	Repositorio	Repositorio que almacena toda la información relacionada al proceso de gestión de la configuración.	Adaptación (IEEE, 2012)
Comisión de control de la configuración	Persona	Grupo que debe evaluar los cambios en los ítems de configuración.	(IEEE, 2012)
Construcción	Concepto	Versión operacional de un sistema.	Adaptación (IEEE, 2012)
Contabilidad de estado de configuración	Concepto	Registro y reporte de información necesaria para gestionar efectivamente el cambio en un ítem de la configuración.	(CMMI Institute, 2020)
Control de la configuración	Concepto	Evaluación, coordinación, e implementación de cambios a los ítems de configuración.	(CMMI Institute, 2020)
Elemento software	Concepto	Componente identificable de un producto software.	Adaptación (IEEE, 2012)
Evaluación objetiva	Concepto	Validación de las actividades y productos de trabajo siguiendo un conjunto de criterios.	Adaptación (CMMI Institute, 2020)
Gestión de entrega de software	Concepto	Gestión de las actividades alrededor de la entrega de una o más versiones de software al cliente.	(IEEE, 2012)
Gestión de la configuración	Concepto	Proceso responsable de Identificar, controlar, registrar y reportar los cambios realizados a un artefacto.	(IEEE, 2012)
Identificación de la configuración	Concepto	Selección de los ítems de configuración para un producto, asignando identificadores únicos.	(CMMI Institute, 2020)
Ítem de configuración	Concepto	Producto de trabajo sensible a cambios.	(IEEE, 2012)
Ítem de configuración constituyente	Concepto	Elemento que hace parte de un ítem de configuración más grande.	(IEEE, 2012)

Términos utilizados en SCMOnto			
Término	**Súper concepto**	**Definición**	**Fuente**
Lanzamiento	Concepto	Versión entregada de una aplicación que se encuentra disponible para salida a producción.	Adaptación (CMMI Institute, 2020)
Línea base	Concepto	Especificaciones que han sido revisadas y aceptadas formalmente y que sirven como base para futuros desarrollos.	Adaptación (IEEE, 2012)
Línea de producto/ servicio	Concepto	Grupo de productos/servicios que comparten características en común.	Adaptación (CMMI Institute, 2020)
Persona	Concepto	Recurso con un rol determinado dentro de la empresa.	Definición propia
Petición de cambio	Concepto	Propuesta formal para la modificación de un ítem de configuración.	Definición propia
Plan de lanzamiento	Concepto	Plan que describe las partes de funcionalidad del sistema que serán implementadas en cada entrega.	Adaptación (IEEE, 2012)
Producto	Concepto	Producto de trabajo del equipo de proyecto.	Adaptación (CMMI Institute, 2020)
Repositorio	Concepto	Colección de todos los artefactos relacionados a software pertenecientes a un sistema.	(IEEE, 2012)
Requerimiento de cliente	Concepto	Representación formal de la necesidad de un cliente.	(CMMI Institute, 2020)
Técnico de configuración	Persona	Rol encargado de realizar los cambios aprobados a un artefacto sensible de configuración.	Definición propia

Tabla 1 – Términos utilizados en SCMOnto

Relaciones entre conceptos utilizadas en SCMOnto		
Nombre	**Conceptos**	**Descripción**
Almacena	Repositorio – Versión	Un repositorio **almacena** n versiones. Una versión **es almacenada** en un repositorio.
Almacenada en	Contabilidad de estado de configuración (CEC) – Base de datos de GC	La CEC es **almacenada en** una Base de datos de GC.
Almacenado en	Lanzamiento – Repositorio	Un lanzamiento **es almacenado en** un repositorio. Un repositorio puede **almacenar** n lanzamientos.
Basada en	Auditoría – Evaluación objetiva	Una auditoría **se basa en** una evaluación objetiva.
Clasifica	Identificación de la configuración – Elemento software	La identificación de la configuración **clasifica** n elementos software.
Controla	Control de la configuración – Ítem de configuración.	El control de la configuración **controla** los ítems de configuración.

Relaciones entre conceptos utilizadas en SCMOnto		
Nombre	**Conceptos**	**Descripción**
Define	Gestión de la configuración (GC) – Autoridad de gestión de la configuración	GC **define** una autoridad de GC. Una autoridad de CM es **definida** por el proceso de GC.
Define	GC – Plan de lanzamiento	GC **define** uno o más planes de lanzamiento. Uno o más planes de lanzamientos **son definidos** dentro del proceso de CM.
Es almacenada en	Auditoría – Base de datos de GC	Una auditoría **es almacenada** en una Base de datos de GC. Una base de datos de CM **puede almacenar** n auditorías.
Es responsable de	Autoridad de gestión de la configuración (AGC)– Proceso	La AGC es **responsable de** la correcta implementación del proceso. El proceso **es responsabilidad de** la AGC.
Establece	GC – Línea base	La GC **establece** una o más líneas base. Una línea base **es establecida** por la GC.
Establece	GC – Control de la configuración	La GC **establece** un control de la configuración. Un control de la configuración **es establecido** por la GC.
Establece	GC - Identificación de la configuración	La identificación de la configuración **es establecida** por la GC. La GC **establece** la identificación de la configuración.
Establece	GC – Contabilidad de estado de configuración (CEC)	La GC **establece** un área CEC. El área de CEC **es establecida** por la GC.
Evalúa	Comisión de control de la configuración (CCC) – Petición de cambio	La CCC **evalúa** de cero a muchas peticiones de cambio para aprobarlas o no. Una petición de **cambio es evaluada** por una CCC.
Genera	Plan de lanzamiento – lanzamiento	Un plan de lanzamiento **genera** lanzamientos. Un lanzamiento **es generado** por un plan de lanzamiento.
Genera	Proceso – Producto	Un proceso **genera** de uno a muchos productos. Un producto **es generado** por un y solo un proceso.
Gestiona	Gestión de entrega de software (GES) – Producto	La GES **gestiona** a uno o más productos. Los productos **son gestionados** por la GES.
Identifica	Identificación de la configuración – Ítem de configuración (IC)	La identificación de la configuración **identifica** a los IC.
Incluye	GC – Auditoría	La GC **incluye** una o más a auditorías dentro del proceso. Una o más auditorías **son incluidas** dentro de la GC.
Lleva a cabo	Técnico de configuración (TC) – Petición de cambio	Un TC **lleva a cabo** peticiones de cambio. Una petición de cambio **es llevada a cabo** por un TC.
Planea	Plan de lanzamiento – Gestión de entrega de software (GES)	Un plan de lanzamiento **planea** la GES. La GES **es planeada** por uno o más planes de lanzamiento.
Puede ser	Línea base – Proceso	Una línea base **puede ser** un proceso. Un proceso **puede ser** tomado como una línea base.
Se realiza sobre	Petición de cambio – Ítem de configuración (IC)	Una petición de cambio puede ser **realizada** sobre uno o más IC. Un IC puede **recibir** cero o más peticiones de cambio.

Relaciones entre conceptos utilizadas en SCMOnto		
Nombre	**Conceptos**	**Descripción**
Solicita	Persona – Petición de cambio	Una persona **solicita** una petición de cambio. Una petición de cambio es **solicitada** por una y solo una persona.
Soluciona	Producto – requerimiento de cliente	Un producto **soluciona** uno o más requerimientos del cliente. Un requerimiento del cliente es **solucionado** por uno o más productos.

Tabla 2 – Relaciones entre conceptos utilizados en SCMOnto

5. Evaluación de SCMOnto

Según (Saraiva de Almeida, Pardo & Mira-da Silva, 2018), la evaluación de la ontología debe centrarse principalmente en la evaluación del usuario en tres elementos relacionados con el rendimiento y la definición correcta: consistencia, integridad y concisión. La evaluación de la consistencia, si es posible inferir el conocimiento de la ontología, la evaluación de la integridad, demostrando la inexistencia de lo incompleto, determinando si las definiciones individuales están bien establecidas y si todo lo que se supone que debe establecerse en la ontología está presente o puede inferirse. Además, si la ontología propuesta no es redundante y tiene definiciones inútiles, entonces la ontología puede considerarse concisa.

La ontología resumida y presentada en este documento fue implementada en el lenguaje de ontologías web OWL (OWL, 2012), usando Protégé (Protégé, 2020) y el razonador Hermit para analizar y evaluar la consistencia de la estructura ontológica de SCMOnto, el código de SCMOnto se puede consultar en el siguiente enlace: https://bit.ly/3barhbL. Una vez que se realizó el análisis, fue posible verificar que SCMOnto no presentaba inconsistencias en su estructura. A continuación, se presenta un resumen de cómo SCMOnto fue evaluada teóricamente a través de un ejemplo de aplicación, una instanciación de un proceso basado en los elementos propuestos para GCS de CMMI-DEV 2.0 y una evaluación cualitativa por medio de una entrevista realizada a un grupo de expertos profesionales en GCS de la industria.

5.1. Narración de un ejemplo teórico de aplicación

A continuación, se presenta un ejemplo teórico de aplicación de algunos de los conceptos incluidos en SCMOnto, los conceptos se podrán apreciar en letra cursiva. Una MiPyME_DS interesada en incrementar la calidad de sus *productos*, ha fijado su atención en la inclusión de mejores prácticas en su *proceso* de desarrollo de software, esto bajo la motivación de llevar a cabo procesos más controlados, organizados y de acuerdo a *modelos de calidad* reconocidos internacionalmente que brinden características de madurez a su empresa; de esta manera ha decidido adoptar prácticas propias de *gestión de la configuración de software*. Teniendo en cuenta lo anterior, la empresa deberá identificar: (i) elementos de proceso primordiales que constituyan a la gestión de la configuración,

tales como *identificación de la configuración* y *control de la configuración*, y (ii) aquellos artefactos en los que se basará dicha área de proceso, así como: *líneas base*, *ítems de configuración*, y sus respectivos *repositorios*, entre otros. Una MiPyME_DS que haya implementado un proceso de GCS, muy probablemente estará interesada en auditar y evaluar constantemente el grado de cumplimiento que sus artefactos, para llevarlo a cabo, la principal herramienta que la empresa tendrá serán los dos tipos de *auditoría de configuración* que la GCS incluye: la *auditoría de configuración física* y la *auditoría de configuración funcional*, ambas son utilizadas con el propósito de verificar el estado y proceso de evolución y desarrollo de los artefactos involucrados durante el proceso de construcción de software y el grado de correspondencia con respecto a su documentación.

5.2. Soportando la instanciación de control de la configuración

La Figura 2 muestra un extracto de la instancia de SCMOnto utilizando el plugin gráfico GraphViz (GraphViz, 2020), base para la visualización de ontología en Protégé, también fue necesario complementar dicha representación con el plugin gráfico OntoGraph (OntoGraph, 2020), responsable de la creación visual de las entidades, así como de sus relaciones. En el extracto se presenta el concepto de *control de la configuración*, uno de los principales componentes de la GCS que se encarga de establecer las actividades necesarias para garantizar la integridad de los artefactos sensibles a cambio durante el proceso de desarrollo de un producto. En la Figura 2 es posible observar cómo algunos de los conceptos representados tienen influencia directa o indirecta sobre otros conceptos mediante la correcta representación de las relaciones presentadas en

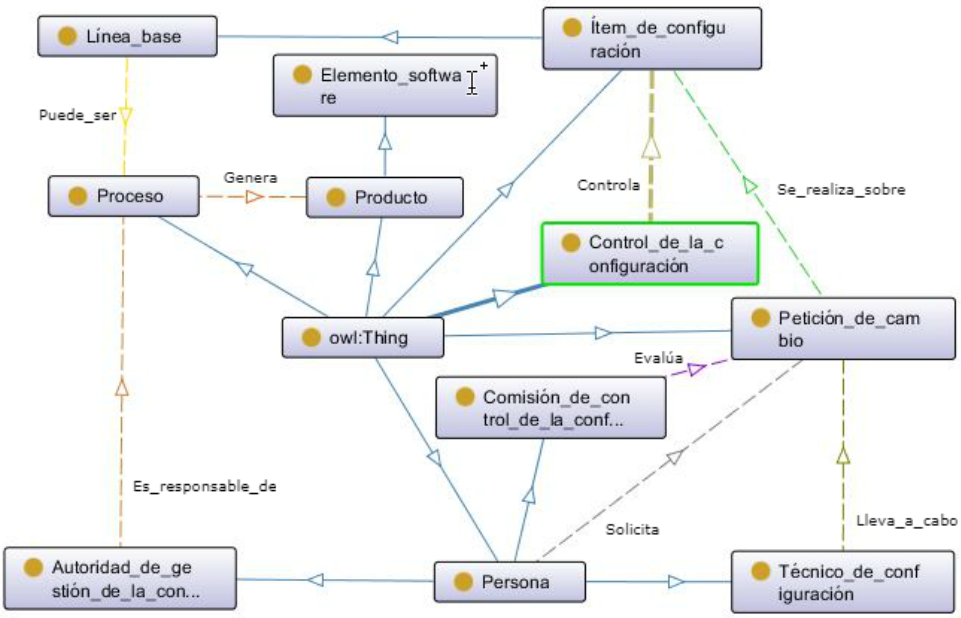


Figura 2 – Extracto de la instancia de SCMOnto

la Tabla 2, lo que permite apreciar que la ontología definida provee un correcto soporte a la representación de los conceptos utilizados en este dominio, más concretamente en el universo del discurso relacionado a la GCS. Las líneas punteadas representan las relaciones entre conceptos, asimismo, las líneas continuas representan herencia entre conceptos, por ejemplo: el concepto *Línea Base* es una especialización del concepto *Ítem de Configuración*.

Uno de los aspectos más importantes de la instanciación de la ontología es buscar información relevante que brinde información adicional sobre el proceso a través de consultas a la ontología que devuelvan información de interés. Gracias a ello, fue posible realizar un mapeo general de todos los conceptos y relaciones entre conceptos con el objetivo de encontrar aquellos que son más recurrentes en la GCS. Como resultado de realizar esta actividad se logró identificar que los términos *Identificación de la configuración, Comisión de control de la configuración, Auditoria de configuración, Ítem de configuración, Línea base y Gestión de la configuración* son los conceptos más significativos de la ontología, lo cual es consecuente con la terminología propuesta por los modelos existentes que definen la GCS a partir de su Identificación, Control, Seguimiento y Auditoría (CMMI Institute, 2020), describiendo de manera exitosa los aspectos más relevantes del área de proceso de forma transparente a cualquier modelo o estándar.

5.3. Evaluación cualitativa a través de entrevistas con profesionales

Se realizó una evaluación cualitativa con una entrevista en la que participaron un total de seis profesionales especialistas en el campo del desarrollo de software y el proceso de GCS que fueron parte de un grupo focal. El procedimiento para llevar a cabo la evaluación sigue los lineamientos definidos en (Kontio, Bragge & Lehtola, 2008). Para la evaluación, los participantes recibieron la documentación correspondiente a un proceso formal para soportar la GCS en el cual se describen los conceptos más relevantes asociados a la terminología propuesta en SCMOnto. Posteriormente, se realizó una evaluación cualitativa en la que los participantes explicaron en sus propias palabras los términos utilizados en el proceso y evaluaron en una escala discreta de 1 a 5 si las definiciones presentadas en SCMOnto corresponden con su propia definición. Los participantes también evaluaron las definiciones de SCMOnto a nivel de su: completes, consistencia y concisión. La Tabla 3 presenta el extracto de la terminología identificada y su calificación, la escala discreta utilizada corresponde a: valor 1: Completamente En Desacuerdo (CD) con la definición del concepto, valor 2: En Desacuerdo (ED) con la definición del concepto, valor 3: el Concepto es Ambiguo (CA), valor 4: Parcialmente de Acuerdo (PA) con la definición del concepto y valor 5: Totalmente de Acuerdo (TA) con la definición del concepto. En general, se pudo observar que los participantes estuvieron de acuerdo con la mayoría de los conceptos evaluados y declararon que SCMOnto es un recurso de gran ayuda que proporciona una visión más clara y completa sobre la GCS, así como un referente a tener en cuenta para adelantar procesos de valoración o evaluación de los procesos software en lo referente a la GCS. Sin embargo, en la Tabla 3 se puede observar que 2 conceptos no fueron calificados de manera favorable: línea base (No. 7) y comisión de control (No. 11), razón por la cual se realizaron acciones de mejora sobre ellos. Asimismo, como resultado del grupo focal, surgieron algunas sugerencias que fueron consideradas en la versión aquí presentada.

No.	Nombre	Calificación					Completes		Consistencia		Concisión	
		CD	ED	CA	PA	TA	SI	NO	SI	NO	SI	NO
1	Producto	0	0	0	5	1	5	1	6	0	6	0
2	Proceso	0	0	0	0	6	6	0	6	0	6	0
3	Modelo de calidad	0	0	0	0	6	6	0	6	0	4	2
4	Gestión de la configuración de software	0	0	0	0	6	6	0	6	0	6	0
5	Identificación de la configuración	0	0	0	4	2	4	2	6	0	6	0
6	Control de la configuración	0	0	0	4	2	6	0	6	0	5	1
7	Línea base	0	0	2	2	2	3	3	4	2	4	2
8	Ítem de configuración	0	0	0	5	1	6	0	4	2	6	0
9	Repositorio	0	0	0	0	6	6	0	6	0	6	0
10	Solicitud de cambios	0	0	0	0	6	6	0	6	0	6	0
11	Comisión de control	0	2	4	0	0	4	2	5	1	3	3
12	Monitoreo	0	0	0	4	2	6	0	6	0	6	0
13	Evaluación	0	0	0	0	6	6	0	6	0	6	0
14	Auditoría física	0	0	0	5	1	5	1	5	1	5	1
15	Auditoría funcional	0	0	0	5	1	6	0	6	1	6	0

Tabla 3 – Conformidad de los participantes respecto a los conceptos identificados

6. Conclusiones y trabajos futuros

En este documento, se ha presentado el resumen de una ontología semi-formal para el dominio de la GCS llamada SCMOnto. El vocabulario descrito en SCMOnto permite organizar el conocimiento relacionado con el desarrollo de software y específicamente el relacionado con la GCS de manera genérica y formal, reduciendo los problemas de inconsistencia, integridad y ambigüedad de los conceptos y términos presentes en este campo de investigación. Además, SCMOnto ha sido definida usando el formalismo REFSENO, representado en UML y construida en el lenguaje OWL para facilitar su uso por la comunidad académica y las empresas de software (el código fuente está disponible en https://bit.ly/3barhbL).

Por otro lado, el objetivo principal de SCMOnto es proporcionar una base para la discusión de los términos, conceptos y relaciones identificados en este campo de investigación, y que ha sido aplicada con éxito en tres casos de implementación: (i) una validación teórica como ejemplo de aplicación, (ii) la instanciación de algunos de los conceptos de

CMMI-DEV 2.0 que permiten apoyar la GCS; y (iii) la evaluación cualitativa con una entrevista en la que participaron un total de seis profesionales especialistas en el campo del desarrollo de software y GCS. Tanto el ejemplo narrativo, como la instanciación de SCMOnto y el grupo focal trajeron consigo un conjunto de beneficios comentados por los participantes, entre ellos: (i) los conceptos, y relaciones entre conceptos son homogéneos y no presentan ambigüedad con respecto a los modelos y estándares existentes, (ii) la instanciación del modelo permitió observar el comportamiento de SCMOnto en un entorno controlado y aclaró los conceptos que inicialmente eran difusos, y (iii) la evaluación cualitativa del modelo por los participantes permitió refinar aquellos conceptos que no eran completamente claros y podían presentar ambigüedad.

Los resultados obtenidos en el desarrollo de SCMOnto se utilizarán para abordar algunos trabajos futuros en el siguiente orden: (i) mejorar la evaluación del constructo de SCMOnto a través de la formulación de preguntas de competencia y el lenguaje de consulta SPARQL, (ii) se espera evaluar la ontología mediante la aplicación de estudios de casos en empresas de software que implementan GCS en sus procesos software y actualizar y ampliar la ontología en función de las recomendaciones obtenidas, (iii) utilizar la ontología como base conceptual para la definición de un marco que respalde la evaluación de la GCS y (iv) focalizar también los esfuerzos en la automatización de la etapa de evaluación, ya que el proceso de evaluación generalmente en las auditorías es actualmente un área manual.

7. Agradecimientos

El profesor César Pardo agradece la contribución de la Universidad del Cauca, donde trabaja como profesor Asociado.

Referencias

Ben Menachem, M. (1995). *Software Configuration Management Guidebook*. Retreived from: http://dl.acm.org/citation.cfm?id=541488

Ferchini, A. Bigand, M., & Lefebvre, H. (2008). An ontology for quality standards integration in software collaborative projects. In *proceedings of the 1st workshop on model driven interoperability for sustainable information systems*, Montepellier (pp. 17-30).

Fernández, M., Gómez, A., & Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In *AAAI-97 Spring Symposium Series*, 97(6), 33–40. doi: 10.1109/AXMEDIS.2007.19.

GraphViz. (2020). Graphviz-Graph visualization software. Retrieved from https://www.graphviz.org/

Gruber, T. R. (1993). A translation approach to portable ontology specifications. In *Knowledge Acquisition*, 5(2), 199-200. doi: 10.1006/knac.1993.1008.

IEEE. (2012). IEEE 828-2012, IEEE Standard for Configuration Management in Systems and Software Engineering IEEE Computer Society. Retreived from: https://ieeexplore.ieee.org/document/6170935

ISO/IEC 29110-2-1. (2015). Software engineering -- Lifecycle profiles for Very Small Entities (VSEs). Retreived from: https://www.iso.org/standard/62712.html

ISO. (2020). ISO 33000: Familia de normas para la evaluación y mejora de la capacidad y madurez de procesos. Retreived from: https://www.iso33000.es/

ISO 10007. (2003). Quality management systems. Guidelines for configuration management. Retreived from: goo.gl/qHsRqR

Jyrki, K., Bragge, J., & Lehyola, L. (2008). The Focus Group Method as an Empirical Tool in Software Engineering. In S. Forrest, S. Janice, & B. Dag I. K (Eds). *Guide to Advanced Empirical Software Engineering,* (pp. 93-116). London, UK: Springer London.

Malzahn, D. (2009). Assessing-Learning-Impriving an integrated approach for self assessment and process improvement systems. In *Proceedings of the 4th international conferenec on systems. IEEE Computer Society*. Guadeloupe (pp. 126-130).

Mendes, O., & Bran, A. (2004). Software engineering ontology: A development methodology. *Metrics News,* 9(1), 68-76.

Mincomercio. (2020). Definición tamaño empresarial Micro, Pequeña, Mediana o Grande empresa. Retreived from: https://bit.ly/2Wd8bol

Liao, L., & Qu, H. (2005). A software process ontology and its application. In *Proceedings of the 4th International Semantic Web Conference, Springer lecture notes in conputer science*. doi: 10.3233/978-1-61499-370-4-207

OntoGraph. (2020). OntoGraph - Protege. Retrieved from: https://protegewiki.stanford.edu/wiki/OntoGraf

Ortega, W., Pardo, C., & Pino, F. (2019). OntoAgile: An ontology for agile software development processes. *DYNA*, 86(209), 79–90.

OWL 2 Web Ontoogy Languaje. (2012). Retreived from: https://www.w3.org/TR/owl2-overview/

Pardo, C., García-rubio, F., & Piattini, M. (2014). Una ontología de referencia para la armonización de modelos de referencia de procesos. *Revista Facultad de Ingeniería Universidad de Antioquia*. 73(1), 29–42. doi: 10.19053/*issn*.0121-1129.

Pardo, C., Pino, F. J., García, F., Piattini, M., & Baldassarre, M. T. (2011). Trends in Harmonization of Multiple Reference Models. In *5th International Conference ENACE 2010,* 230(1), 61-73. doi: 10.1007/978-3-642-23391-3.

Pardo, C., Pino, F. J., García, F., Piattini, M., & Baldassarre, M. T. (2012). An ontology for the harmonization of multiple standards and models. In *Computer Standards and Interfaces*. 34(1), 48-59. doi: 10.1016/j.csi.2011.05.005.

Perera, P. N. G., Atapattu, A. R. D. C., Dias, H. T., Liyanage, N. T., Silva, R. O. C., Rupasingha, P. S., Fernando, S. G. S., & Manawadu, C. D. (2013). The impact of effective configuration management usage in software development firms in Sri Lanka. In *Proceedings of the 8th International Conference on Computer Science and Education*. doi: 10.1109/ICCSE.2013.6553997.

Protége (2020). The Protégé Ontology Editor and Knowledge Acquisition System. Retrieved from goo.gl/vQvpvo.

Regalado Hernández, R. (2007). Las MiPyMEs en Latinoamérica. Retreived from: goo.gl/6MiTxe.

Ruiz, F., & Hilera, J. (2006). Using ontologies in software engineering and technology. In *Ontologies for Software Engineering and Software Technology*. (pp. 126-130). doi: 10.1007/3-540-34518-3_2.

Salviano, C., & Figueiredo, A. (2008). Unified basic concepts for procees cabability models. In *Proceedings of the conference on SEKE*. San Francisco, USA. (pp. 173-178).

Saraiva de Almeida, R., Pardo, C., & Mira-da Silva, M. (2018). An ontology-based model for ITL process assessment using TIPA for ITIL. *Software process improvement and capability determination*. New York, NY. Springer editors.

Software Engineering Institute. (2010). CMMI for Development, Version 2.0. *Software Engineering Institute*. Retreived from: goo.gl/bbyt4z.

Soydan, G., & Kokar, M. (2013). An OWL ontology for representing the CMMI-SW model. Retreived from: http://km.aifb.kit.edu/ws/swese2006/final/soydan-full.pdf

Tautz, C., & Von Wangenheim, C. G. (1998). REFSENO: A Representation Formalism for Software Engineering Ontologies. Retreived from: https://bit.ly/3fudno7.

Wang, Y., & King, G. (2000). *Software Engineering Processes: Principles and Applications (1st Ed)*. Cleveland, Ohio. CRC Press.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 02/08/2020
Aceitação/Acceptance: 06/10/2020

Bloques de primos para acelerar el algoritmo Quadratic Sieve

Juan Manuel Campo Revelo[1], Jhon Alejandro Melo Zúñiga[2],
Carlos Alberto Trujillo Solarte[3], Siler Amador Donado[4]

**jmcamporevelo@unicauca.edu.co; alejandromelo@unicauca.edu.co;
trujillo@unicauca.edu.co; samador@unicauca.edu.co**

[1,3] Facultad de Ciencias Naturales, exactas y de la Educación, Universidad del Cauca, Carrera 2 #4N-140, 190003, Popayán- Cauca, Colombia

[2,4] Facultad de Ingeniería Electrónica y Telecomunicaciones, Universidad del Cauca, Carrera 2 #4N-140, 190003, Popayán- Cauca, Colombia

Pages: 91-102

Resumen: Se presenta una variación en el proceso de cribado tradicional del algoritmo Quadratic Sieve, la cual disminuye el tiempo de recolección de números suaves y por tanto impacta en el tiempo total de factorización de llaves públicas generadas con el algoritmo RSA.

Palabras-clave: Cribado; criptoanálisis; factorización; números suaves; RSA.

Prime blocks to speed up the Quadratic Sieve algorithm

Abstract: We presented a variation in the traditional screening process of the Quadratic Sieve algorithm, which reduces the collection time of soft numbers and therefore impacts the total factorization time of public keys generated with the RSA algorithm.

Keywords: Screening; cryptanalysis; factoring; soft numbers; RSA.

1. Criptoanálisis de RSA y el problema de factorización

Con el advenimiento de los canales de comunicación actuales, uno de los problemas más importantes en la seguridad de la información está relacionado con los algoritmos criptográficos, ya que son objetos de estudio en cuanto a sus fortalezas y especialmente en sus debilidades. Al hablar de la seguridad de un algoritmo criptográfico, está generalmente asociada a algún argumento de reducción desde encontrar un mapeo inverso de la función de cifrado hasta un problema matemático cuya complejidad se conjetura es de tipo exponencial con respecto a la magnitud de la instancia del problema a resolver, dos de los ejemplos más usados en la actualidad son: el problema del logaritmo discreto en un grupo cíclico finito y el problema de factorización de números enteros.

Este artículo supone que criptoanalizar RSA (Rivest, 1983) es al menos tan difícil como factorizar un número entero compuesto producto de dos primos distintos. Sin embargo, para ilustrar un poco esta reducción, se presenta el proceso de generación de llaves pública y privada con RSA junto con la relación del problema de factorización y la inversión de la función de cifrado.

1.2. Generación de llaves publica y privada con RSA

Se inicia por elegir dos números primos p,q y se calcula:

$$N = pq,$$

$$\varphi(N) = (p-1)(q-1),$$

donde φ es la función de Euler que cuenta la cantidad de enteros positivos, menores que *N*, que son primos relativos con *N*.

Ahora se selecciona un entero *e*, $1 \leq e \leq \varphi(N)$, tal que

$$mcd\big(e, \varphi(N)\big) = 1.$$

Esto con el fin de que el texto plano cifrado con RSA pueda descifrarse. Por último se resuelve la ecuación diofántica:

$$ex \equiv 1 \ mod \ \varphi(N),$$

cuya solución es la llave privada que permitirá descifrar el texto plano. Sí $x = d$ se obtiene el par de llaves, la pública (e, N) y la privada d.

1.3. Criptoanálisis del algoritmo RSA

Dada la llave pública (e, N) se desea calcular la llave privada; por tanto, se debe calcular la solución a la ecuación

$$ex + \varphi(N)k = 1,$$

que se resuelve de manera rápida a nivel computacional, usando el algoritmo extendido de Euclides, siempre y cuando se conozcan los coeficientes $e, \varphi(N)$; el coeficiente e se conoce como parte de la llave pública; el coeficiente $\varphi(N)$ no se conoce, pero si el entero N; por tanto, se debe encontrar el entero $\varphi(N)$ a partir de N y como $\varphi(N)$ está en función de los números primos cuyo producto es igual a N, se puede suponer que criptoanalizar el algoritmo RSA es “al menos”, tan difícil como factorizar el entero N. Es muy importante enfatizar en la expresión “al menos”, puesto que si bien es muy usual pensar que la

seguridad de un protocolo como RSA está basada completamente en el problema de factorización de enteros, los análisis sobre los argumentos reduccionistas nos dicen lo contrario como en el caso de (Neal Koblitz, 2007).

2. Algoritmos de factorización

Dado un entero compuesto producto de dos números primos, por ejemplo:

114381625757888867669235779976146612010218296721242362562561842935706935245733897830597123563958705058980751475992900268795435 41 (RSA-129)

¿Qué tan difícil puede ser encontrar los números primos que lo componen? Cuando se hace referencia a "difícil", se debe pensar en que si se desarrolla algún tipo de procedimiento para obtener la respuesta y se implementa en un ordenador, ¿cuánto tiempo tardaría éste en encontrar el par de números primos? Si el entero se hace cada vez más grande, ¿cuál es la relación que existe entre la longitud del entero que se investiga y el tiempo que tardaría el ordenador en encontrar una respuesta? Para el problema general de factorización de enteros, que es la base de la seguridad proporcionada por las llaves generadas con RSA, esta relación se describe mediante una función exponencial; por tanto, una llave de 1024 bits podría ser segura aún para ser atacada por un ordenador tradicional bajo un ataque basado en factorización.

2.1. Fermat y la diferencia de cuadrados en los enteros

Uno de los primeros algoritmos de factorización creados en la historia fue el del matemático francés Pierre de Fermat, este es el origen de los algoritmos de factorización actuales. Sin embargo, el corazón de los mismos lo proporcionaría el matemático Belga Kraitchik, cuya idea se considera en la siguiente sección.

Para comenzar se tiene el siguiente polinomio:

$$Q(x_i) = x_i^2 - N,$$

donde N es el entero que se desea factorizar y $x_i = i + \lceil \sqrt{N} \rceil$ con i un contador que recorre los enteros positivos hasta encontrar un cuadrado perfecto; es decir, si para algún $i = k$ fijo existe un entero positivo y tal que:

$$y^2 = x_k^2 - N,$$

despejando N, se obtiene que:

$$\begin{aligned} N &= x_k^2 - y^2, \\ &= (x_k - y)(x_k + y). \end{aligned}$$

Por tanto, se tiene que probablemente,

$$p = (x_k - y),$$

$$q = (x_k + y).$$

El problema de este método, es que su tiempo de ejecución está relacionado con la distancia entre los primos p, q; por tanto, se comienza a ver que los primos empleados por RSA deben cumplir con ciertas condiciones como las que se describen en (Ronald L Rivest, 1999)

2.2. Kraitchik y la diferencia de cuadrados módulo N

El siguiente gran paso lo dio Maurice Kraitchik en (Kraïtchik, 1926),el matemático Belga que extendió el método de Fermat, debilitando la condición de encontrar un cuadrado en los enteros a encontrar un cuadrado en los enteros módulo N, esto es en $\mathbb{Z}_N = \{0,1,2, \dots, N-1\}$. La propuesta de Kraitchik se describe como sigue.

Considere $x, y \in \mathbb{Z}$, si se cumple que:

$$x^2 \equiv y^2 \ mod \ N, \tag{1}$$

$$x \not\equiv y \ mod \ N, \tag{1}$$

entonces, hay un 66.7% de probabilidad de que $mcd(x \pm y, N)$ sea un factor no trivial de *N*. Encontrar ese par de enteros es el "corazón" de los actuales algoritmos de factorización.

3. Algoritmo Quadratic Sieve

Los tres algoritmos más potentes que hoy existen son:

- Quadratic Sieve (QS)
- Multiple Polynomial Quadratic Sieve (MPQS)
- General Number Field Sieve (GNFS)

La modificación que se realizó fue en el Quadratic Sieve, un algoritmo creado por Carl Pomerance en 1981 (Pomerance, 1996), cuyo proceso de cribado no difiere mucho de los otros dos, pero si usa conjuntos y polinomios diferentes. Para entender mejor el algoritmo QS a continuación se describe.

Primero se crea una "base" *B* de números primos, los cuales cumplen que:

$$\left(\frac{N}{p}\right) = 1,$$

para cada primo *p* en *B*. Donde $\left(\frac{N}{p}\right)$ es el simbolo de Legendre, el cual para un entero *n* y un primo *p* dados, se define como:

$$\left(\frac{n}{p}\right) = \begin{cases} 0, & Si\ p\ didive\ a\ n, \\ 1, & si\ n\ es\ residuo\ cuadrático\ módulo\ p, \\ -1, & si\ n\ no\ es\ residuo\ cuadrático\ módulo\ p. \end{cases}$$

En segundo lugar está el "proceso de cribado", justo aquí se concentra el estudio para entender en donde dio la modificación a la implementación. Para comenzar este proceso, es importante notar la siguiente relación:

$$Q(x_i) \equiv x_i^2 mod\ N.$$

Se considera una sub-sucesión de enteros $x_{i_1}, x_{i_2}, x_{i_3}, \ldots, x_{i_j}$ tales que:

$$\prod_{t=1}^{j} Q(x_{i_t}) = y^2\ , y \in \mathbb{Z}, \tag{3}$$

ahora, tomando:

$$\prod_{t=1}^{j} x_{i_t} = x,$$

tenemos que :

$$x^2 = \prod_{t=1}^{j} {x_{i_t}}^2 = \prod_{t=1}^{j} Q(x_{i_t}) = y^2\ mod\ N,$$

por lo tanto, se ha resuelto la ecuación (1). El proceso de cribado, se encarga de encontrar dicha sub-sucesión $\{x_{i_t}\}_{t=1}^{j}$ que cumpla la condición (3). Para ello, si se genera una lista de enteros x_i la sub-sucesión de enteros $\{x_{i_t}\}_{t=1}^{j}$ cumplirá que la sub-sucesión $\{Q(x_{i_t})\}_{t=1}^{j}$ se factoriza completamente con los primos en la base *B*. Hasta aquí hay varios aspectos que son importantes identificar, primero ¿cuántos primos debe tener la base *B*?, segundo ¿cuántos enteros x_i *se* deben generar?, y ¿de qué tamaño debe ser la sub-sucesión $\{x_{i_t}\}_{t=1}^{j}$? Por fortuna estos aspectos ya han sido ampliamente estudiados y se encuentran en la referencia (Landquist, 2001). Donde los valores estimados son:

$$|B| = \left(e^{\sqrt{\ln(N)\ln(\ln(N))}}\right)^{\sqrt{2}/4}$$

la cantidad de x_i que se deben generar es aproximadamente $|B|^3$ y la longitud de la sub-sucesión está alrededor de $|B| + 1$.

3.1. Cribado tradicional

En el lenguaje usual se podría decir que se busca un producto de números de la forma $Q(x_{i_t})$ que sea un cuadrado perfecto y aunque hasta aquí no se ha explicado como ayudan estos enteros que se factorizan completamente con primos en la base *B*, se puede decir que en la literatura se les conoce como números *B*-suaves, los cuales se definen como sigue.

Dada una base de primos $B = \{p_1, p_2, \dots, p_k\}$, decimos que un entero *n* es *B*-suave si todos los factores primos presentes en la descomposición canónica de nn son menores o iguales que el mayor de los primos contenidos en *B*.

Por ejemplo, el entero 15 es *B*-suave para cada una de las bases *B= {3,11}, B= {2,5,19}, B= {3,5,17}*. Encontrar este tipo de números en una lista de longitud exponencial es un problema en el cual aún muchos científicos están trabajando. Se ha estimado en (Kazumaro Aoki, 2004) (Wambach, 1995), que el porcentaje de tiempo empleado en esta fase es alrededor del 70%, por tanto vale la pena dirigir esfuerzos a mejores formas de realizar dicho procedimiento.

El cribado tradicional hace uso de la función de von Mangoldt, la cual se define como:

$$\Lambda(n) = \begin{cases} \text{Log } p, & si\ n = p^k\ para\ p\ primo\ y\ k \geq 1, \\ 0, & en\ otro\ caso. \end{cases}$$

La idea general de este recurso teórico está basada en pre-calcular para cada primo $p \in B$, el número $Log\ p$. Aúnque si un número es *B* - suave entonces el valor del logaritmo de este número, por la identidad anterior, debe ser $\sum_{p \in B} Log\ p$, aún mejor, se puede aproximar un valor general para el logaritmo de cada entero que se desea saber si es suave, es decir aproximar el valor de $Log\left(Q(x_i)\right) = T$ y después comparar este valor con $\sum_{p \in B} Log\ p$, para cada $p\ divisor\ de\ Q(x_i)$. Ésta es la idea tras el proceso de cribado, el valor de *T* se ha estimado como $Log\left(\sqrt{2NM}\right) - \delta$, donde δ es un parámetro que estima el error, debido al redondeo en el cálculo de los logaritmos de los números primos.

Posterior a esta fase para probar que cada uno de estos enteros es *B*-suave, se hacía con un procedimiento de divisiones sucesivas sugerido como en (Pomerance, 2008); las divisiones por ensayo y error suelen ser un buen candidato para el reconocimiento de números *B*-suaves. Sin embargo, si las divisiones se hacen de una en una, se puede generar un alto gasto computacional, justamente es aquí donde tiene lugar el algoritmo propuesto en este trabajo.

4. Algoritmo de cribado por bloques

Una de las ideas propuestas en (Bard, 2009), que parecía tener una ventaja considerable respecto a los métodos tradicionales, consiste en tomar una base de primos, por ejemplo $B = \{2,5,11,13,19\}$, y el producto de todos los primos en *B*, esto es el entero $B_1 = 27170$. Si al considerar el entero N=15884, se toma el máximo común divisor entre *N* y B_1 se obtiene 418, este número aporta una información muy especial, pues contiene los primos en común entre la base *B* y el entero *N*, evitando tener que seleccionar primos en la base que posiblemente no estén y finalmente solo se procede a dividir, en este caso:

$$\frac{n}{gcd(n, B_1)} = 38.$$

Note que esto ha reducido de manera considerable el tamaño del número que se desea probar, un nuevo cálculo del máximo común divisor permitirá reducir de manera sucesiva el número hasta llegar a 1 en el caso de que sea *B*-suave o a un entero distinto de 1 en caso contrario.

Esta idea parece ser prometedora. Sin embargo, para implementarse en el algoritmo QS necesita un par de ajustes. Por una parte, en la práctica las bases generadas para factorizar un entero grande tienen un tamaño alrededor de los cientos, miles o millones, dependiendo del propósito del proyecto. Para el segundo aspecto se debe introducir una definición que permitirá visualizar otro inconveniente de la técnica anterior.

4.1. Vector exponente

Pese a que se ha hecho énfasis en que se deben encontrar números suaves, no se ha dicho como ellos permiten construir un cuadrado perfecto y es aquí donde los vectores exponentes juegan un papel crucial.

Considere un entero *n* que sea *B*-suave, este se puede escribir como:

$$n = \prod_{i=1}^{|B|} p_i^{\alpha_i}.$$

A este entero se le asocia un vector:

$$v(n) = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_{|B|}).$$

Sí además cada componente se reduce módulo 2, y se denota mediante $\mathbb{Z}(B)$ al conjunto de todos los enteros *B*-suaves, es muy conveniente definir una relación en que cada par de enteros n_1, n_2 y se considera relacionados, escrito $n_1 \sim^2 n_2$ si y solo sí $n_1 n_2 = m^2$, es decir el producto de los dos enteros $n_1 n_1$, $n_2 n_2$ es un cuadrado perfecto. Esta relación es simétrica, reflexiva por la conmutatividad de la multiplicación y transitiva; afirmaciones que pueden probarse con la ayuda de propiedades básicas de divisibilidad. Lo importante de este hecho es que esta relación es de "equivalencia" y por lo tanto, induce una partición. La "función" vector exponente envía elementos de $\mathbb{Z}(B)$ a elementos del espacio vectorial binario $F_2^{|B|}$. Un ejemplo que ayuda a ilustrar dicha asignación.

Sea $B = \{2,5,11,13,19\}$, un elemento de $\mathbb{Z}(B)$ es 15884 , ya que:

$$15884 = 2^2 \text{x} 11 \text{x} 19^2,$$

entonces,

$$v: \mathbb{Z}(B) \longrightarrow F_2^{|B|}$$

$$v(15884) \longrightarrow (0,0,1,0,0).$$

Se puede probar, además que esta función convierte productos en sumas, es decir:

$$v(nm) = v(n) + v(m).$$

Esto permite factorizar enteros por partes, recuperando sus vectores exponentes para al final sumarlos y encontrar el vector exponente asociado al número inicial.

Regresando al ejemplo, si se toma el número $n = 15884$, basta con factorizar $418 = 2 * 11 * 19$ y contar cuantas veces este número divide a *n*, como es una vez, se asigna el número 1 a las posiciones correspondientes de los primos que están presentes en 418, obteniendo el vector $(1,0,1,0,1)$; ahora se calcula el segundo máximo común divisor entre $\frac{n}{gcd(n,B_1)}$ y *B* y repitiendo el proceso anterior se obtiene el vector $(1,0,0,0,1)$, sumando estos vectores se obtiene el vector exponente del número original:

$$v(n) = (1,0,1,0,1) + (1,0,0,0,1) = (0,0,1,0,0).$$

5. Algoritmo QS y la generación de cuadrados perfectos

Finalmente se explican las últimas dos etapas del algoritmo QS. El álgebra lineal juega un papel fundamental en la penúltima etapa del algoritmo, puesto que si tenemos una lista de $|B| + 1$ vectores, al agruparlos como las columnas de una matriz A, el sistema:

$$Ax = 0$$

siempre tiene solución, lo que permite encontrar un producto de enteros que es un cuadrado perfecto puesto que, por la relación de equivalencia expuesta en 4.1, el núcleo de esta función son los enteros que cumplen esta propiedad (solución del sistema lineal homogéneo arriba mencionado). Finalmente la raíz de este producto es y; los respectivos x_{i_t} conforman el producto que produce el valor de *x*. Con estos elementos ya se puede realizar el cálculo $\gcd(x \pm y, N)$ para obtener un factor no trivial de *N*.

6. Implementación y resultados

Cuando se trabaja en un algoritmo con complejidad de tipo exponencial o sub exponencial, se debe estar dispuesto a realizar cálculos con números de grandes longitudes en bits, este va a ser uno de los primeros problemas a los que se enfrentan las personas que deseen implementar el algoritmo QS, debido a que para poder factorizar enteros grandes, las funciones, rutinas y algoritmos deben ejecutarse sobre tales números y las librerías tradicionales pueden colapsar para estos propósitos.

Por ejemplo, uno de los problemas es la magnitud de los números que aparecen en el polinomio de Fermat, los cuales crecen conforme la variable de evaluación; por tanto,

si se está cribando sobre un gran intervalo, evaluar este polinomio puede volverse un problema a nivel de implementación, funciones aritméticas como el símbolo de Legendre pueden ser programadas sin tomar en cuenta la magnitud de los números primos que pueden surgir al factorizar un entero grande e incluso la implementación del máximo común divisor en el algoritmo por bloques o en el paso final del algoritmo pueden no estar pensadas para ejecutar cálculos sobre enteros de estas magnitudes: como primer paso se debe escoger un lenguaje de programación adecuado, con librerías robustas que permitan realizar este tipo de operaciones.

Dentro de la revisión bibliográfica, en la cual se incluyó la búsqueda de software de factorización, una de las librerías más usadas para lograr este objetivo fue la librería GMP del lenguaje C. GMP, ha sido usada para implementación de procesos criptográficos, cálculo de una cantidad grande de cifras de números irracionales, álgebra computacional, entre otros, por ello para el desarrollo de este proyecto se usó GMP, básicamente porque no impone límites a la precisión de ciertos cálculos de interés, salvo la capacidad en memoria de la máquina. GMP tiene tres grandes grupos de funciones: funciones aritméticas para números enteros (mpz), funciones aritméticas para números racionales (mpq), y funciones aritméticas para números de punto flotante (mpfr); en particular se usaron muchas funciones de la categoría (mpz). Para más información acerca de la librería puede consultarse https://gmplib.org/.

6.1. Construcción de llaves

Una vez desarrollado el algoritmo, se deben tener enteros adecuados para someterlo a pruebas, pese a que se podría pensar que cualquier par de números primos pueden servir para construir una llave fuerte, se debe tener cuidado porque esto no es así, se mencionó alguna propiedad especial en la descripción del algoritmo de factorización realizado por Fermat. Un software muy usado para implementar protocolos criptográficos es OpenSSL, este software libre tiene implementadas herramientas tales como: algoritmos de cifrado, funciones hash y, en particular, algoritmos de cifrado con llaves públicas que utilizan curvas elípticas y el de nuestro particular interés, RSA. Entonces, se utilizó OpenSSL como software para generar llaves de ciertas longitudes en bits para colocar a prueba las implementaciones del algoritmo QS y el algoritmo QS modificado.

6.2. Álgebra lineal desde CADO NFS

Para la solución del sistema $Ax = 0$ se hace uso de uno de los softwares de factorización más usados y que encabeza la lista de los más rápidos hasta el momento CADO NFS; aquí se ha hecho un trabajo robusto a nivel de matemáticas e implementación, en particular en el algoritmo “Block of Wiedemann”. Para resolver este tipo de sistemas se utilizan algoritmos de tipo aleatorio en lugar de usar los deterministas tradicionales, como la reducción gaussiana, esto debido a que la factorización de enteros grandes produce matrices de gran tamaño, para las cuales el número de operaciones ejecutadas en un algoritmo tradicional no es razonable. Los algoritmos de factorización en la actualidad usan dos algoritmos “Block Lanczos” o “block of Wiedemann” mostrando un mejor comportamiento “block of Wiedemann” (Thomé, 2002), debido a que presenta menos inconvenientes al distribuir dicho algoritmo. En este sentido, muchos problemas

relacionados con criptografía se reducen a sistemas de ecuaciones lineales sobre campos finitos, el algoritmo de Wiedemann (Wiedemann, 1986) resuelve en general matrices dispersas con estas características.

Esta es la única parte que no fue desarrollada de manera propia, sino que se usó el módulo desarrollado por CADO para este tipo de problemas.

6.3. Comparación de tiempos

En la tabla 1 se presentan algunos de los números que representan el conjunto de llaves públicas factorizadas.

Número	Cantidad de bits	Dígitos decimales
N=2885024642365232115893	72	22
N=6498531341628375087427	76	23
N=1037629337240609549795357	80	25
N=1524619533651447201728396 9	84	26
N=1892596886561418941611763 69	88	27
N=3711206406576858345168225797	92	28
N=6080300895909956409435784 1711	96	29

Tabla 1 – Llaves generadas con OpenSSL

En la tabla 2,se muestran los tiempos que tardaron en ser factorizadas las llaves públicas, comparando los tiempos totales entre el algoritmo QS y el algoritmo QS modificado.

Cantidad de bloques creados	Tiempo De QS	Tiempo de QS modificado	Primos
10	27,62 s	12,04 s	p=52638391627 q=54808373759
10	63,41 s	24,66 s	p=257810647163 q=252066057517
10	290,9 s	56,35 s	p=1009603002269 q=1027759757953
10	101,55 s	45,42 s	p=4114279168553 q=3705678373273
10	192,30 s	58,20 s	p=13845184828807 q=13669711960967
10	643,43 s	128,58 s	p=62669233635271 q=59218953085907
10	1798,25 s	286,31 s	p=245932543121911 q=247234498481801

Tabla 2 – QS vs QS modificado

Se observa que hay una diferencia considerable entre los tiempos, conforme aumenta el orden de magnitud de las llaves públicas, con esta leve modificación al algoritmo se ha logrado acelerar el proceso de cribado y además se ha ajustado para poder construir los vectores exponentes de los números suaves asociados.

7. Conclusiones

El aspecto más importante de esta investigación consiste en que este método de cribado por bloques de primos, representa una mejora en el tiempo total de factorización respecto al método tradicional que emplea divisiones sucesivas; pese a que se puedan diseñar experimentos para números más grandes, es necesario realizar una demostración analítica del ahorro en operaciones aritméticas que presenta esta implementación con el fin de poder afirmar, con mayor precisión, que el método modificado es computacionalmente más eficiente que el tradicional.

Otra observación que se hizo consiste en que la longitud de los bloques creados, si bien en este caso se usó una longitud constante de 10, se realizaron experimentos donde el aumento de la longitud representaba una disminución en tiempo al acercarse a la longitud total de la base. Sin embargo, para bases muy grandes hay puntos donde la diferencia de tiempo entre un aumento de base y otro no es significativo, es como si llegáramos a un punto de estancamiento donde la ganancia no es mucha.

También es necesario encontrar una mejor forma de calcular las veces que divide un bloque al número de prueba, puesto que esto garantiza, y acelera, la construcción del vector exponente. Este tipo de problemas son evidencia de que debe formalizarse este proceso para poder encontrar una mejor conexión entre las rutinas que involucra y sobre todo saber si este comportamiento de reducción del tiempo se mantendrá al llegar a longitudes más grandes.

Agradecimientos

Agradecimientos a la Universidad del Cauca en especial al semillero SEC (Security, Encryption & Cybersecurity) del grupo de investigación GTI y al grupo de investigación Altenua-Matdis.

Referencias

Bard, G. V. (2009). *Algebraic Cryptanalysis*. Springer-Verlag .

Kazumaro Aoki, H. U. (2004). Sieving Using Bucket Sort. *Lecture Notes in Computer Science, 3329*(https://doi.org/10.1007/978-3-540-30539-2_8).

Kraïtchik, M. (1926). Théorie des nombres. *Tomo II*.

Landquist, E. (2001). *The Quadratic Sieve Factoring Algorithm*. https://www.cs.virginia.edu/crab/QFS_Simple.pdf, MATH 488: Cryptographic Algorithms.

Neal Koblitz, A. J. (2007). Another Look at "Provable Security". *Springer Link - Journal of Cryptology volume*, 20-27.

Pomerance, C. (1996). A tale of two sieves. *Notices of the AMS, 1473-1485*. https://www.ams.org/notices/199612/pomerance.pdf.

Pomerance, C. (2008). Smooth numbers and the quadratic sieve. *Algorithmic Number Theory, 44*(http://library.msri.org/books/Book44/files/03carl.pdf), 2008.

Rivest, R. S. (1983). A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications Of The ACM, 26. https://doi: 10.1145/357980.358017*, 96-99.

Ronald L Rivest, R. D. (1999). Are 'Strong' Primes Needed for RSA'. (https://people.csail.mit.edu/rivest/pubs/RS01.version-1999-11-22.pdf), 11-22.

Thomé, E. (05 de 2002). Subquadratic Computation of Vector Generating Polynomials and Improvement of the Block Wiedemann Algorithm. *Journal of Symbolic Computation, 33*(https://doi.org/10.1006/jsco.2002.0533), 757-775.

Wambach, G. W. (1995). Block Sieving Algorithms. *http://e-archive.informatik.uni-koeln.de/190/*.

Wiedemann, D. (01 de 1986). Solving sparse linear equations over finite fields. *IEEE Transactions on Information Theory, 32*(https://doi:10.1109/TIT.1986.1057137), 54-62.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 13/08/2020
Aceitação/Acceptance: 26/10/2020

Mapeo Sistemático sobre la Web Semántica de las Cosas y el Cuidado de Adultos Mayores

Andrea Pabón-Guerrero[1], Miguel-Ángel Niño-Zambrano[1], Lider-Julian Rojas-Bolaños[1]

andreapabon@unicauca.edu.co; manzamb@unicauca.edu.co; jlr@unicauca.edu.co

[1] Universidad del Cauca, Calle 5 # 4-70, 190001, Popayán, Colombia.

Pages: 103-116

Resumen: A nivel mundial se evidencia una tendencia hacia el incremento de la población de adultos mayores, muchos de ellos viven solos con los correspondientes riesgos de salud propios de la edad. Colombia no es ajena a este fenómeno, por lo cual, crear mecanismos que permitan brindarles una vejez activa y saludable es indispensable. Las tecnologías relacionadas a la Web de las Cosas y la vida asistida por el entorno (AAL) permiten monitorear el estado de salud de los adultos mayores y brindarles servicios que ayuden a su comodidad y bienestar, teniendo en cuenta sus necesidades y preferencias particulares. Este articulo presenta un análisis sobre las soluciones propuestas para personalizar los servicios en escenarios AAL utilizando técnicas de la Web Semántica de las Cosas. Los resultados muestran que, aunque se han realizado esfuerzos para personalizar servicios, aún no han logrado incorporar completamente las necesidades y preferencias específicas de esta población.

Palabras-clave: Adultos Mayores, Web Semántica de las Cosas, Vida Asistida por el Entorno.

Systematic Mapping of the Semantic Web of Things and the Care of Elderly

Abstract: Globally, there is a trend towards an increase in the population of older adults, a lot of them living alone with the corresponding health risks of age. Colombia is no stranger to this phenomenon, therefore, creating mechanisms that allow them to provide them with an active and healthy old age is essential. Technologies related to the Web of Things and Ambient Assisted Living (AAL) make it possible to monitor the health status of older adults and provide them with services that help their comfort and well-being, taking into account their particular needs and preferences. This article presents an analysis of the solutions proposed to personalize services in AAL scenarios using techniques from the Semantic Web of Things. The results show that, although efforts have been made to personalize services, they have not yet fully incorporated the specific needs and preferences of this population.

Keywords: Elderly, Semantic Web of Things, SWoT, Ambient Assisted Living.

1. Introducción

La Internet de las Cosas (Internet of Things – IoT) ha permitido que los objetos cotidianos puedan obtener y procesar información de su entorno, proporcionando servicios y comunicándose por medio de protocolos web con otros objetos, convirtiéndose así en objetos inteligentes (Estrada-Martinez & Garcia-Macias, 2013). Cuando estos objetos tienen una representación en una plataforma común y ampliamente utilizada como lo es la web, aparece el concepto de la Web de las Cosas (Web of Things – WoT) (Mathew, Atif, Sheng, & Maamar, 2011).

Teniendo en cuenta la heterogeneidad de los dispositivos de la IoT, se han incorporado técnicas de la Web Semántica para que los datos compartidos cuenten con un significado bien definido, de manera que sean comprensibles tanto para las máquinas como para los humanos (Berners-Lee,Hendler & Lassila, 2001). Lo anterior ha permitido construir servicios de información inteligentes, que suplen las necesidades de los usuarios, a ello se le ha denominado la Web Semántica de las Cosas (Semantic Web of Things – SWoT) (Scioscia & Ruta, 2009). El hecho de tener datos comprensibles, permite a la SWoT tener capacidades de razonamiento, infiriendo implícitamente consecuencias lógicas y nuevos hechos sobre los datos. Para codificar el conocimiento de dominio y las reglas de inferencia se usan los lenguajes de ontologías (Nagib & Hamza, 2016).

Uno de los campos de aplicación de la IoT (Internet of Things – IoT) es la creación de objetos inteligentes con Inteligencia Ambiental, donde los objetos registran cambios en el ambiente y ofrecen servicios ajustados a sus usuarios (Dohr, Modre-Osprian, Drobics, Hayn, & Schreier, 2010). En este sentido, la IoT está fuertemente relaciona con los sistemas AAL (Ambient Assisted Living), cuyo propósito es monitorear tanto las condiciones ambientales, como el comportamiento del adulto mayor (personas mayores de 60 años). Un entorno AAL cuenta con un sistema de sensores desplegados en la casa inteligente, con el fin de apoyar al adulto mayor en las actividades diarias, mantener su autonomía y alertar a sus familiares sobre situaciones de riesgo (Al-Shaqi,Mourshed & Rezgui, 2016; Fides-Valero, Freddi, Furfari, & Tazari, 2008).

Una preocupación respecto a los adultos mayores surge en torno a la pregunta de investigación: ¿cómo apoyar a los adultos mayores en sus actividades diarias para que puedan vivir de una manera activa e independiente en un ambiente familiar para ellos? La preocupación anterior cobra relevancia teniendo en cuenta que son una población en aumento tanto en Colombia como el mundo y en la mayoría de los casos viven solos, por ello, se han realizado una serie de investigaciones en las que la IoT ha desempeñado un papel importante, por ejemplo, en el campo de la salud, monitoreando sus actividades para reaccionar de manera oportuna ante anomalías y mejorar su calidad de vida (Sundmaeker, Guillemin, Friess, & Woelfflé, 2010; Wu & Tsai, 2018), ayudarles a encontrar los objetos importantes que muchas veces olvidan debido a su edad (Hung,Chao & Chen, 2018), detectar el cambio de comportamiento, predecir depresión (Chen, Chen, Chang, & Tang, 2018), detectar caídas (Galvão, Albuquerque, Fernandes, & Valença, 2017), y realizar seguimiento de los patrones del sueño para detectar crisis de apnea del sueño (Yacchirema, Sarabia-Jácome, Palau, & Esteve, 2018), entre otras. Otras investigaciones pretenden modelar por medio de tecnologías de la web semántica, específicamente por medio de ontologías, las actividades de la vida diaria (Ni,Pau & García Hernando, 2016), la información médica de los adultos

mayores, los datos ambientales, preferencias relacionadas a condiciones ambientales específicas (Kurnianingsih, Nugroho, Widyawan, Lazuardi, & Non-alinsavath, 2015), modelar las discapacidades y condiciones de salud utilizando estándares internacionales y localización dentro y fuera de casa de manera no intrusiva (Almeida et al., 2019).

A pesar de lo anterior, en la mayoría de las soluciones propuestas, se presta poca atención a la representación de los objetos inteligentes y los servicios ofrecidos por los mismos para suplir la carencia de personalización[1] de servicios en la IoT. La personalización se logra incluyendo mecanismos semánticos necesarios para aprovechar los datos obtenidos y aprender del usuario con el fin de ofrecerles servicios ajustados a sus preferencias o necesidades actuales, el cual es un campo aún en investigación (Hussein, Han, Lee, Crespi, & Bertin, 2017; Vallée et al., 2016; Wang & Chow, 2016).

En este sentido, este mapeo sistemático se centra en analizar los estudios que incorporen técnicas de Web Semántica en entornos AAL para personalizar los servicios que son ofrecidos a los adultos mayores, y de esta manera brindarles independencia y una mejor calidad de vida.

Este estudio se estructura de la siguiente manera: En la Sección 2 se presenta la metodología seguida para realizar el mapeo sistemático. En la Sección 3 se muestran los resultados obtenidos y se concluye sobre las preguntas de investigación, finalmente se presentan las conclusiones y trabajo futuro en la Sección 4.

2. Metodología

El presente estudio establece como procedimiento guía los pasos definidos por (Petersen, Feldt, Mujtaba, & Mattsson, 2008), sin embargo, se complementa con el procedimiento de investigación documental propuesto por (Hoyos, 2000; Serrano, 2003), tomando de estos últimos las plantillas descriptivas con el fin de estructurar el conocimiento científico relevante encontrado en los diferentes estudios. El mapeo sistemático se realizó en tres fases: Preparatoria, Ejecución y Documentación, como se muestra en la Figura 1.

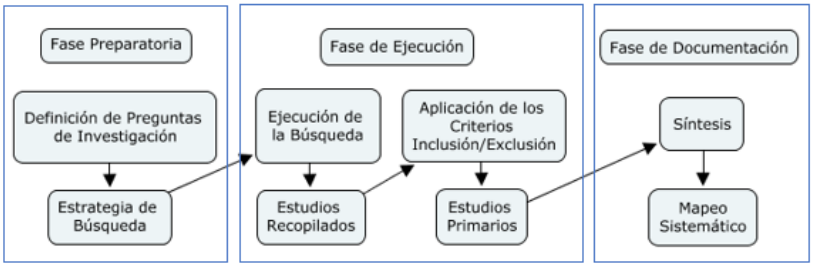


Figura 1 – Fases Mapeo Sistemático

[1] Personalización: la capacidad de proporcionar contenidos y servicios adaptados a las personas en función del conocimiento sobre sus necesidades, expectativas, preferencias, limitaciones y comportamientos"(Berkovsky & Freyne, 2015)

2.1. Fase Preparatoria.

El presente mapeo sistemático tiene como objetivo proveer una visión general del conocimiento sobre la personalización de servicios en entornos AAL basado en la Web Semántica de las Cosas para el cuidado de los adultos mayores. Se toma como ventana de tiempo desde el año 2015 hasta la actualidad.

Pregunta de investigación.

Las preguntas de investigación presentadas en la Tabla 1 permiten conocer los trabajos que han incorporado la Web Semántica para la personalización de servicios para el cuidado de adultos mayores en entornos AAL. Además, permiten la identificación de técnicas utilizados y de las brechas de investigación existentes.

No.	Pregunta
PI1	¿Cuáles son las investigaciones desarrollados alrededor de Sistemas AAL que utilizan técnicas de la Web Semántica de las Cosas para el cuidado de los adultos mayores?
PI2	¿Qué técnicas de la Web Semántica de las Cosas han sido utilizan en los sistemas AAL para el cuidado de los adultos mayores?
PI3	¿Qué soluciones han sido propuestas para personalizar los servicios ofrecidos por los sistemas AAL?
PI4	¿Las soluciones propuestas brindan recomendaciones o guía al usuario para mejorar su calidad de vida?

Fuente de Datos y Estrategia de Búsqueda.

Tabla 1 – Preguntas de Investigación

La cadena de búsqueda se construye en base a los pasos definidos por (Kitchenham, 2007), para la cual se considera el uso los operadores lógicos AND y OR. En la Tabla 2 se describen los términos usados en esta investigación para la construcción de la cadena de búsqueda. La búsqueda se realizó en los motores: Scopus, Science Direct, ACM Library e IEEExplore, considerando los estudios realizados entre el 2015 y 2020.

Términos Principales	Términos Relacionados
Elderly	60 and over age OR aged OR aging population OR older adult OR older people OR senior
Personalization	Customization OR Service Personalization OR User Preferences
AAL	Internet of things OR web of things OR ambient assisted living OR Assistive living systems OR smart home
Semantic Web	Semantic Web Technology OR ontology OR Semantic Web of Things

Tabla 2 – Términos de Búsqueda

Criterios de Inclusión y Exclusión.

Los estudios fueron delimitados aplicando criterios de inclusión y exclusión. En cuanto a los criterios de inclusión se tomaron en cuenta: (1) publicaciones académicas indexadas

realizadas en conferencias, revistas y libros, (2) artículos publicados entre 2015 - 2020 y (3) artículos que incorporen los términos considerados en la cadena de búsqueda.

Fueron excluidos los artículos que: (1) no utilizaron técnicas de la web semántica de las cosas, (2) no fueron enfocados a la personalización de servicios para escenarios de la AAL enfocados en el cuidado de los adultos mayores.

Método de Síntesis.

La información de los estudios seleccionados, se extrajo y estructuró adaptando las fichas bibliográficas propuestas por (Hoyos, 2000; Serrano, 2003): Identificación (título, año de publicación, autores), resumen, enfoque de solución, enfoque metodológico, conclusiones y brechas.

2.2. Fase de Ejecución.

Luego de realizar la búsqueda en los motores definidos, el número de artículos encontrados fue 72, distribuidos según el motor de búsqueda de la siguiente manera: ACM: 17, IEEE: 7, Science Direct: 26, Scopus: 22, de los cuales 7 estaban duplicados. Después de leer el título, resumen y palabras claves, fueron aplicados los criterios de inclusión y exclusión, de ser necesario los artículos fueron descargados y revisados a mayor detalle hasta tomar una decisión. Finalmente, 14 artículos fueron categorizados como estudios primarios debido a su aporte y relación con la temática de estudio.

2.3. Fase de Documentación.

A continuación se presentan los resultados obtenidos a cada una de las preguntas de investigación definidas en la etapa de Planificación.

Pregunta PI1: ¿Cuáles son las investigaciones desarrollados alrededor de Sistemas AAL que utilicen técnicas de la Web Semántica de las Cosas para el cuidado de los adultos mayores?

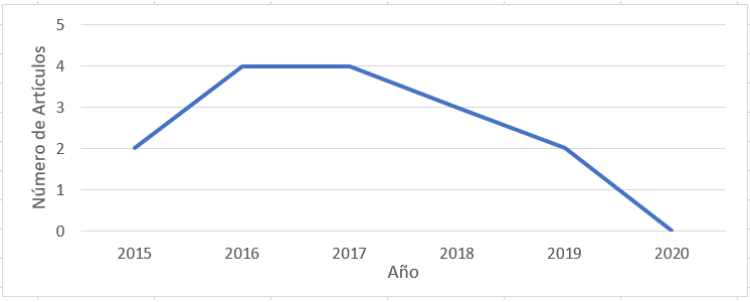


Figura 2 – Evolución Temporal de las Publicaciones

Las investigaciones encontradas se preocupan por mejorar la calidad de vida del adulto mayor, promoviendo su independencia y monitoreando constantemente su estado de

salud para generar alertas oportunas. En cuanto a la personalización de servicios, como se muestra en la Figura 2, el periodo de tiempo comprendido entre los años 2015 y 2017 fue el más relevante en este sentido, sin embargo, después de este periodo las investigaciones decaen, debido a que se enfocan en otros aspectos como por ejemplo en los protocolos de comunicación más que en la personalización.

Pregunta PI2: ¿Qué técnicas de la Web Semántica de las Cosas se utilizan en los sistemas AAL para el cuidado de los adultos mayores?

Con el fin de clasificar las técnicas utilizadas por las diferentes propuestas, fue utilizado el ciclo de vida del contexto propuesto por (Perera, Zaslavsky, Christen, & Georgakopoulos, 2014), el cual se compone de las siguientes fases: i) adquisición, se refiere a la forma de obtener el contexto, ii) modelado, o la forma de representar el contexto, iii) razonamiento, donde se obtiene información del modelo que se generó en la fase anterior y iv) distribución, la cual proporciona métodos para entregar el contexto a los consumidores.

De acuerdo a lo anterior, en la Tabla 3 sobresale la utilización de ontologías como medio para modelar el contexto, las mismas son creadas a partir de alineaciones a SSN (Semantic Sensor Network Ontology), DUL (DOLCE+DnS Ultralite ontology) y FOAF (Friend Of A Friend), ontologías que son ampliamente citadas en el contexto de la IoT debido a su interoperabilidad, reutilización y extensibilidad. El razonamiento se realiza por medio de reglas SWRL, pero también existen propuestas con enfoques basados en lógica difusa (Chiang & Liang, 2015), emparejamiento híbrido (Stavrotheodoros,Kaklanis & Tzovaras, 2018) o clasificadores Naives Bayes (Baig & Jilani, 2017) para ofrecer servicios adecuados o predecir situaciones de riesgo. En cuanto a los estudios que utilizan la técnica de distribución, estos se basan en servicios RESTFull, Web Sockets y Publicador/Suscriptor. Estudios como los realizados por (de Vargas & Pereira, 2017; Moguel et al., 2018) se enfocan en presentar la ontología.

	Adquisición			**Modelado**	**Razonamiento**	**Distribución**
Artículo	**Sensada**	**Derivada**	**Manual**			
(Chiang & Liang, 2015)	X	X	X	Ontología	Reglas Fuzzy	
(Ni et al., 2016)	X		X	Ontología	Reglas SWRL	
(Stavrotheodoros et al., 2018)	X		X	Ontología	Reglas SWRL, enfoque de emparejamiento híbrido	
(Yilmaz, 2019)	X	X	X	Ontología	Basado en reglas	RESTFull API
(Baig & Jilani, 2017)	X	X		Ontología	Clasificador Naive Bayes	RESTFull API
(Kurnianingsih et al., 2015)	X	X		Ontología	Basado en reglas	

(Baldassini, Colombo, Spoladore, Sacco, & Arlati, 2017)	X	X		Ontología	Reglas SWRL	
(Merkle & Zander, 2016)	X			Ontología	Reglas SWRL	Web Sockets
(de Vargas & Pereira, 2017)	X	X	X	Ontología	Basado en Reglas	
(Loreti et al., 2019)	X	X		Ontología	Cálculo de Eventos Reactivos (REC) Reglas CEP	Publicar/Suscribir. WebSockets
(Moguel et al., 2018)	X		X	Ontología		
(De Backere, Bonte, Verstichel, Ongenae, & De Turck, 2017)	X	X		Ontología	Por medio de los razonadores Hermit, Pellet y consultas SPARQL	
(Machado, Maran, Augustin, Wives, & de Oliveira, 2017)	X	X		Ontología	Reglas SWRL	
(Lunardi et al., 2018)	X	X		Ontología	Redes Lógicas de Marcov	

Tabla 3 – Técnicas de la Web Semántica utilizadas por los estudios encontrados

Pregunta PI3: ¿Qué soluciones han sido propuestas para personalizar los servicios ofrecidos por los sistemas AAL?

Los estudios monitorean los signos vitales, condiciones de salud, localización, comportamiento y actividades realizadas por el adulto mayor, con el fin de generar alertas tanto a familiares como a personal de salud si alguna anomalía se presenta. Las propuestas presentadas por (Baig & Jilani, 2017; Chiang & Liang, 2015; Moguel et al., 2018; Stavrotheodoros et al., 2018) tienen parámetros ya definidos para generar las alertas o ya tienen definido qué servicios ejecutar de acuerdo a la situación. Por su parte, propuestas como (de Vargas & Pereira, 2017; Kurnianingsih et al., 2015; Merkle & Zander, 2016; Moguel et al., 2018; Yilmaz, 2019), personalizan los servicios de acuerdo a las preferencias del usuario en cuanto a nivel de luz, humedad, etc., además, permiten que los usuarios ingresen sus propias reglas al sistema.

Pregunta PI4: ¿Las soluciones propuestas brindan recomendaciones o guía al usuario para mejorar su calidad de vida?

Los estudios que utilizan técnicas de la Web Semántica de las Cosas en entornos AAL se enfocan en monitorear al adulto mayor para detectar anomalías y generar alertas, además de emitir recordatorios programados por el mismo usuario. Según las condiciones de salud pueden sugerir servicios (por ejemplo, si para alguna condición de salud es contraindicado temperaturas bajas, se ajustará la temperatura de la habitación), sin embargo, no brindan recomendaciones respecto a la promoción de actividad física, alimentación saludable o manejo del estrés.

3. Discusión

Los estudios encontrados utilizan las ontologías para modelar el contexto y ofrecer servicios adecuados a los usuarios. Para ello, incorporan los datos personales del usuario(Baldassini et al., 2017; Chiang & Liang, 2015; de Vargas & Pereira, 2017; Kurnianingsih et al., 2015; Ni et al., 2016; Stavrotheodoros et al., 2018; Yilmaz, 2019), estado de salud (Baldassini et al., 2017; Chiang & Liang, 2015; de Vargas & Pereira, 2017), preferencias en cuanto a luz, humedad, temperatura, entre otras, (de Vargas & Pereira, 2017; Ni et al., 2016). Algunos estudios tienen en cuenta las discapacidades y limitaciones funcionales del adulto mayor basándose en estándares de salud aceptados por la comunidad médica (de Vargas & Pereira, 2017; Stavrotheodoros et al., 2018).

En cuanto a los servicios ofrecidos, se modela su descripción, parámetros y detalles de conexión (Chiang & Liang, 2015; de Vargas & Pereira, 2017; Merkle & Zander, 2016; Stavrotheodoros et al., 2018). Algunos trabajos reutilizan la ontología SSN (Semantic Sensor Network Ontology) tomando conceptos como Sensor, Observación, Evento, Medida y Propiedad (Merkle & Zander, 2016; Ni et al., 2016; Stavrotheodoros et al., 2018). Para sugerir servicios adecuados, algunos trabajos modelan la casa del adulto mayor (Baldassini et al., 2017; Kurnianingsih et al., 2015; Merkle & Zander, 2016; Ni et al., 2016). El aspecto temporal es tenido en cuenta en investigaciones como (de Vargas & Pereira, 2017; Ni et al., 2016; Yilmaz, 2019), lo cual permite realizar inferencias sobre el comportamiento y las actividades que lleva a cabo el adulto mayor, con el fin de detectar anomalías y alertar de manera oportuna a familiares o personal de salud.

Los anteriores estudios evidencian los conceptos importantes para construir un modelo de contexto adecuado, que permita a un sistema AAL ofrecer servicios personalizados usando técnicas de la web semántica.

En cuanto al razonamiento, no todas las propuestas tienen en cuenta las preferencias de usuario para ejecutar los servicios ofrecidos:

- (Chiang & Liang, 2015) utiliza razonamiento basado en reglas de lógica difusa para clasificar los signos vitales del adulto mayor, comparándolos con parámetros definidos por organizaciones médicas, para ofrecer servicios de emergencia (llamar a familiares o profesionales médicos), servicios de comodidad (ajustar la temperatura, luz, humedad) o servicios de mejoramiento autónomo (por ejemplo, enciende el humidificador para prevenir la artritis). Este estudio se preocupa por los adultos mayores con dificultades funcionales, por ello propone una interfaz que detecta el movimiento de las manos (por ejemplo, levantar la mano permite registrarse en el sistema). Al utilizar una cámara para detectar los gestos del usuario, es un poco intrusivo.
- (Stavrotheodoros et al., 2018) combina reglas SWRL con un enfoque de emparejamiento híbrido para detectar situaciones de emergencia y seleccionar el servicio adecuado, el usuario decide si ejecuta o no el servicio y emite una calificación para que el sistema mejore sus recomendaciones. Las reglas no tienen en cuenta las preferencias del usuario.
- (Baldassini et al., 2017) propone un sistema para que el adulto mayor realice ejercicio de manera segura, para ello monitorea las condiciones de salud y las condiciones del ambiente, y por medio de reglas SWRL infiere si son adecuadas.

Si alguna condición puede poner en riesgo al adulto mayor, genera servicios para mejorarla (por ejemplo, si hay poca luz, abre las cortinas).

- (Merkle & Zander, 2016) dota a los objetos de etiquetas QR para su identificación. El usuario debe usar gafas de seguimiento ocular (Eye Tracking Glasses). Cuando el adulto mayor enfoca un objeto, el sistema le sugiere servicios que le puede ofrecer basado en reglas SWRL. El usuario decide si ejecuta o no el servicio. Como trabajo futuro, propone buscar una solución menos intrusiva, y disminuir la intervención del usuario. Debido a la necesidad de utilizar gafas especiales para leer las etiquetas QR, esta solución es costosa.
- (Moguel et al., 2018) proponen perfiles virtuales que representan la información del adulto mayor, información de los dispositivos y sistemas con los cuales interactúa. Además de información sobre las habilidades del usuario, la cual es obtenida a través de una evaluación que realizan los profesionales de la salud. Lo anterior permite adaptar los servicios (por ejemplo, si el adulto mayor presenta problemas de movilidad y cuenta con una silla inteligente, esta se debe adaptar para que la persona se pueda levantar fácilmente). Si el sistema detecta anomalías, se genera una alarma.

Algunas propuestas utilizan el razonamiento para localizar al usuario y evitar situaciones no deseadas:

- (Baig & Jilani, 2017) dotan al adulto mayor de sensores wearables, y despliegan Beacons dentro de la casa para rastrear su posición. Los datos arrojados por los sensores son clasificados con Naive Bayes, y dependiendo si son críticos o no, se envía una alerta al centro médico más cercano.
- (De Backere et al., 2017) monitorea las condiciones de salud del adulto mayor para detectar comportamientos anómalos y generar alertas al cuidador de ser necesario, además, pregunta al adulto mayor si necesita ayuda para realizar actividades.
- (Lunardi et al., 2018) incorporan una Red Lógica de Marcov para realizar un razonamiento probabilístico para detectar la actividad que está realizando el adulto mayor.
- (Machado et al., 2017) propone actuar de una manera proactiva ante situaciones de emergencia, percibiendo con antelación eventos no deseados, generando acciones para prevenirlos y alertando al adulto mayor sobre las posibles consecuencias.
- Por su parte (Ni et al., 2016) no se enfoca en generar servicios, sino en detectar las actividades que está realizando el usuario, para ello utiliza reglas SWRL.

Por otro lado, aunque se generan alertas ante anomalías, es importante que los adultos mayores reciban sugerencias sobre promoción de actividad física, alimentación saludable o manejo del estrés, que les permita llevar un estilo de vida saludable e independiente, impactando en su calidad de vida.

Finalmente, las siguientes investigaciones incorporan en su propuesta las preferencias de usuario para ofrecer los servicios

- (Kurnianingsih et al., 2015) propone una ontología que incluye las condiciones médicas y los dispositivos con los que interactúa el adulto mayor. Utiliza el

enfoque de razonamiento basado en reglas para ajustar el entorno de acuerdo a las condiciones de salud y preferencias del usuario (nivel de luz, humedad y temperatura). Si bien modela el registro médico y los signos vitales, no utiliza un estándar para ello. No especifica el mecanismo para obtener las preferencias del usuario.

- (de Vargas & Pereira, 2017) proponen una ontología que modela los problemas funcionales del usuario y sus condiciones de salud utilizando estándares internacionales, lo cual hace esta ontología reutilizable. Tiene en cuenta las preferencias del usuario (temperatura, tamaño de texto). Dichas preferencias son generadas automáticamente, por ejemplo, si el usuario presenta limitaciones visuales, se genera una preferencia sobre el tamaño del texto. El mecanismo de razonamiento no es presentado.
- (Yilmaz, 2019) propone un sistema para adultos mayores con demencia temprana, donde la persona a cargo del cuidado del adulto mayor define zonas seguras y reglas que se adecúen a las necesidades del adulto mayor. El sistema rastrea al adulto mayor, y si está fuera de estas zonas, emite una alerta de emergencia. Debido a que se enfoca en los adultos mayores con demencia, solo ofrece servicios limitados de localización y generación de alertas. En cuanto a la interfaz ofrecida al adulto mayor, permite reproducir por medio del Smartphone indicaciones para que el adulto mayor pueda volver a casa.

4. Conclusiones

Los resultados encontrados muestran que la utilización de técnicas de la web semántica para el cuidado de los adultos mayores en entornos AAL, permite modelar tanto al usuario como los dispositivos y servicios proporcionados, permitiendo razonar sobre este modelo para obtener servicios adecuados a las necesidades del usuario. Aunque se han presentado varias propuestas de ontologías, aún no se tiene una ontología estándar que integre los conceptos necesarios para entornos AAL y que tenga en cuenta las necesidades y preferencias del adulto mayor.

En cuanto al razonamiento, las investigaciones no se centran en reglas SWRL, sino que también han mostrado la factibilidad de incorporar diferentes enfoques como razonamiento basado en lógica difusa.

La personalización de servicios aún necesita de la intervención del usuario, y aunque las investigaciones han tenido en cuenta la condición de salud y algunas preferencias, hace falta incorporar los intereses personales, el estado de ánimo y otras consideraciones que permitirían generar servicios adecuados para la población de adultos mayores. Por ende, la construcción de un modelo que tenga en cuenta lo mencionado anteriormente, podría mejorar su calidad de vida y fomentar su independencia. Además, es importante que los servicios generados tengan mecanismos no intrusivos para que hagan parte natural de sus propias actividades.

Por otro lado, estos sistemas carecen de recomendaciones orientadas en forma de hábitos, lo que sería de gran ayuda para que los adultos mayores lleven un estilo de vida saludable. Es de resaltar que en los estudios se evidencia la necesidad de desarrollar

sistemas que ayuden a los adultos mayores a tener una mejor calidad de vida, además, estas propuestas ayudarían a reducir costos en cuanto al sistema de salud.

Referencias

Al-Shaqi, R., Mourshed, M., & Rezgui, Y. (2016). Progress in ambient assisted systems for independent living by the elderly. *SpringerPlus*, *5*(1), 624. https://doi.org/10.1186/s40064-016-2272-8

Almeida, A., Mulero, R., Rametta, P., Urošević, V., Andrić, M., & Patrono, L. (2019). A critical analysis of an IoT—aware AAL system for elderly monitoring. *Future Generation Computer Systems*, *97*, 598-619. https://doi.org/10.1016/j.future.2019.03.019

Baig, M. M. B., & Jilani, M. T. (2017). An iBeacon based Real-time context-aware e-healthcare system. In *2017 First International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT),* Karachi (pp. 1-5). IEEE.

Baldassini, D., Colombo, V., Spoladore, D., Sacco, M., & Arlati, S. (2017). Customization of domestic environment and physical training supported by virtual reality and semantic technologies: A use-case. In *2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI),* Modena (pp. 1-6). IEEE.

Berkovsky, S., & Freyne, J. (2015). Web Personalization and Recommender Systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* Sydney (pp. 2307–2308). Association for Computing Machinery.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, *284*(5), 34-43.

Chen, W., Chen, L., Chang, W., & Tang, J. (2018). An IoT-based elderly behavioral difference warning system. In *2018 IEEE International Conference on Applied System Invention (ICASI),* Chiba (pp. 308-309). IEEE.

Chiang, T. C., & Liang, W. H. (2015). A Context-Aware Interactive Health Care System Based on Ontology and Fuzzy Inference. *Journal of Medical Systems*, *39*(9). https://doi.org/10.1007/s10916-015-0287-2

De Backere, F., Bonte, P., Verstichel, S., Ongenae, F., & De Turck, F. (2017). The OCarePlatform: A context-aware system to support independent living. *Computer Methods and Programs in Biomedicine*, *140*, 111-120. https://doi.org/10.1016/j.cmpb.2016.11.008

de Vargas, M. F., & Pereira, C. E. (2017). Ontological User Modeling for Ambient Assisted Living Service Personalization. In *System Level Design from HW/SW to Memory for Embedded Systems,* Cham (pp. 3-14). Springer International Publishing.

Dohr, A., Modre-Osprian, R., Drobics, M., Hayn, D., & Schreier, G. (2010). The Internet of Things for Ambient Assisted Living. In *2010 Seventh International Conference on Information Technology: New Generations,* Las Vegas (pp. 804-809). IEEE.

Estrada-Martinez, P. E., & Garcia-Macias, J. A. (2013). Semantic interactions in the Internet of Things. *Int. J. Ad Hoc Ubiquitous Comput.*, *13*(3/4), 167-175. https://doi.org/10.1504/ijahuc.2013.055464

Fides-Valero, Á., Freddi, M., Furfari, F., & Tazari, M.-R. (2008). The PERSONA Framework for Supporting Context-Awareness in Open Distributed Systems. In In E. Aarts, J. L. Crowley, B. de Ruyter, H. Gerhäuser, A. Pflaum, J. Schmidt, & R. Wichert, *Ambient Intelligence* Berlin, Heidelberg (pp. 91-108). Springer Berlin Heidelberg.

Galvão, Y. M., Albuquerque, V. A., Fernandes, B. J. T., & Valença, M. J. S. (2017). Anomaly detection in smart houses: Monitoring elderly daily behavior for fall detecting. In *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI),* Arequipa (pp. 1-6). IEEE.

Hoyos, C. (2000). *Un modelo para investigación documental : guía teórico-práctica sobre construcción de Estados del Arte con importantes reflexiones sobre la investigación*. Señal Editora.

Hung, L.-P., Chao, Y.-H., & Chen, C.-L. (2018). A Hybrid Key Item Locating Method to Assist Elderly Daily Life Using Internet of Things. *Mobile Networks and Applications*, *24*, 786-795. https://doi.org/10.1007/s11036-018-1083-2

Hussein, D., Han, S. N., Lee, G. M., Crespi, N., & Bertin, E. (2017). Towards a dynamic discovery of smart services in the social internet of things. *Computers & Electrical Engineering*, *58*(Supplement C), 429-443. https://doi.org/10.1016/j.compeleceng.2016.12.008

Kitchenham, B. (2007). Procedures for Performing Systematic Literature Reviews in Software Engineering. *Keele University & Durham University, UK*.

Kurnianingsih, Nugroho, L. E., Widyawan, Lazuardi, L., & Non-alinsavath, K. (2015). Ontology-based context aware for ubiquitous home care for elderly people. In *2015 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE),* Semarang (pp. 454-459). IEEE.

Loreti, D., Chesani, F., Mello, P., Roffia, L., Antoniazzi, F., Cinotti, T. S., Paolini, G., Masotti, D., & Costanzo, A. (2019). Complex reactive event processing for assisted living: The Habitat project case study. *Expert Systems with Applications*, *126*, 200-217. https://doi.org/10.1016/j.eswa.2019.02.025

Lunardi, G. M., Machado, G. M., Machot, F. A., Maran, V., Machado, A., Mayr, H. C., Shekhovtsov, V. A., & Oliveira, J. P. M. d. (2018). Probabilistic Ontology Reasoning in Ambient Assistance: Predicting Human Actions. In *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA),* Krakow (pp. 593-600).

Machado, A., Maran, V., Augustin, I., Wives, L. K., & de Oliveira, J. P. M. (2017). Reactive, proactive, and extensible situation-awareness in ambient assisted living. *Expert Systems with Applications*, *76*, 21-35. https://doi.org/10.1016/j.eswa.2017.01.033

Mathew, S. S., Atif, Y., Sheng, Q. Z., & Maamar, Z. (2011). Web of Things: Description, Discovery and Integration. In *Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing,* (pp. 9-15).

Merkle, N., & Zander, S. (2016). Improving the Utilization of AAL Devices through Semantic Web Technologies and Web of Things Concepts. *Procedia Computer Science*, *98*, 290-297. https://doi.org/10.1016/j.procs.2016.09.045

Moguel, E., Berrocal, J., Murillo, J. M., García-Alonso, J., Mendes, D., Fonseca, C., & Lopes, M. (2018). Enriched elderly virtual profiles by means of a multidimensional integrated assessment platform. *Procedia Computer Science*, *138*, 56-63. https://doi.org/10.1016/j.procs.2018.10.009

Nagib, A. M., & Hamza, H. S. (2016). SIGHTED: A Framework for Semantic Integration of Heterogeneous Sensor Data on the Internet of Things. *Procedia Computer Science*, *83*, 529-536. https://doi.org/10.1016/j.procs.2016.04.251

Ni, Q., Pau, I., & García Hernando, A. (2016). A foundational ontology-based model for human activity representation in smart homes. *Journal of Ambient Intelligence and Smart Environments*, *8*(1), 47-61. https://doi.org/10.3233/AIS-150359

Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Context Aware Computing for The Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials*, *16*(1), 414-454. https://doi.org/10.1109/SURV.2013.042313.00197

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, *17*.

Scioscia, F., & Ruta, M. (2009). Building a Semantic Web of Things: Issues and Perspectives in Information Compression. In *Semantic Computing, 2009. ICSC '09. IEEE International Conference on,* Berkeley (pp. 589-594). IEEE.

Serrano, C. E. (2003). *Un Modelo Integral para un Profesional en Ingeniería*. Universidad del Cauca.

Stavrotheodoros, S., Kaklanis, N., & Tzovaras, D. (2018). A personalized cloud-based platform for AAL support to cognitively impaired elderly people. In *IFMBE Proceedings,* Singapore (pp. 87-91). Springer Singapore.

Sundmaeker, H., Guillemin, P., Friess, P., & Woelfflé, S. (2010). *Vision and Challenges for Realizing the Internet of Things*. Publications Office of the European Union. https://doi.org/10.2759/26127

Vallée, T., Sedki, K., Despres, S., Jaulant, M. C., Tabia, K., & Ugon, A. (2016). On Personalization in IoT. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI),* Las Vegas (pp. 186-191). IEEE.

Wang, E., & Chow, R. (2016). What can i do here? IoT service discovery in smart cities. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops),* Sydney (pp. 1-6). IEEE.

Wu, H.-T., & Tsai, C.-W. (2018). A home security system for seniors based on the beacon technology. *Concurrency and Computation: Practice and Experience*, *30*(15), e4496. https://doi.org/10.1002/cpe.4496

Yacchirema, D. C., Sarabia-Jácome, D., Palau, C. E., & Esteve, M. (2018). A Smart System for Sleep Monitoring by Integrating IoT With Big Data Analytics. *IEEE Access*, *6*, 35988-36001. https://doi.org/10.1109/ACCESS.2018.2849822

Yilmaz, Ö. (2019). An ambient assisted living system for dementia patients. *Turkish Journal of Electrical Engineering and Computer Sciences*, *27*(3), 2361-2378. https://doi.org/10.3906/elk-1806-124

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 25/08/2020
Aceitação/Acceptance: 28/10/2020

Clasificación de noticias criminales basada en procesamiento del lenguaje natural y algoritmos de aprendizaje automático

Camilo Ernesto Sarmiento Torres[1], Néstor Diaz[1] y Rubiel Vargas Cañas[1].

csarmiento@unicauca.edu.co; nediaz@unicauca.edu.co; rubiel@unicauca.edu.co

[1] Universidad del Cauca, Calle 5 # 4-70, 190001, Popayán, Colombia.

Pages: 117-129

Resumen: En este trabajo se desarrolló un sistema de clasificación de noticias criminales reportados en diferentes medios de prensa digital apoyado en técnicas de procesamiento del lenguaje natural y algoritmos de aprendizaje de máquina. Inicialmente se construyó un conjunto de datos de noticias criminales donde se identificaron ocho tipos de delitos. Posteriormente se realizó un preprocesamiento de los documentos, se eliminaron las palabras vacías, se aplicó una lematización y una representación de los documentos con el modelo de bolsa de palabras donde se estimó el coeficiente de frecuencia de término–frecuencia inversa de documento (tf-idf). Además, se construyeron ocho diccionarios de palabras acordes a los tipos de delitos y se implementaron para estimar el rendimiento de cinco algoritmos supervisados de clasificación. El algoritmo de bosques aleatorios obtuvo el mejor rendimiento con exactitud del 97,22%, precisión del 98,36%, sensibilidad del 98,35%, puntaje F1 del 98,32% y MCC del 0,97% en la prueba realizada.

Palabras-clave: noticia criminal; procesamiento del lenguaje natural; aprendizaje automático; minería de texto.

Criminal news classification based on natural language processing and machine learning algorithms

Abstract: In this work, a classification system of criminal news was developed from different digital press media, supported by natural language processing techniques and machine learning algorithms. Initially, a criminal news data set was constructed where eight types of crime were identified. Subsequently, the documents were pre-processed, the stop words were eliminated, a lemmatization was applied, and a representation of the documents with the bag of words model, where the coefficient of term frequency-inverse document frequency (tf-idf) was estimated. In addition, eight-word dictionaries were built according to the types of crimes and implemented to estimate the performance of five supervised classification algorithms. The random forest algorithm obtained the best performance with 97.22% of accuracy, 98.36% of precision, 98.35% of sensitivity, F1 score of 98.32%, and MCC of 0.97% in the test performed.

Keywords: criminal news, natural language processing, machine learning, text mining.

1. Introducción

Las noticias criminales son relatos que describen las características de los sucesos delictivos y los hechos que van en contra de lo establecido por la Ley. En ellas se describen: las víctimas, personas implicadas, los daños, la propiedad comprometida, una descripción de la causa, el relato cronológico, las acciones legales relacionadas con la investigación y el proceso judicial (Vega, 2014). Donde, el consolidado oficial de delitos de impacto establecidos por la Policía Nacional de Colombia en el año 2019 registró un total de 38.281 lesiones en accidentes de tránsito, 303.040 hurtos a personas, 44.404 hurtos a residencias, 34.248 hurtos a motocicletas, 10.113 hurtos a automotores, 128 hurtos a entidades financieras, 57.425 hurtos a entidades comerciales, 12.558 homicidios y 34.583 delitos sexuales (Estadística delictiva, 2019). Además, la mayoría de estos hechos y sucesos fueron reportados e informados por los diferentes medios de comunicación, especialmente por los noticieros, medios de prensa y medios de difusión a través de la internet.

Durante los últimos años los medios de prensa han digitalizado la mayor parte de su contenido y lo han dejado disponible a través de la web; este hecho, ha creado un gran volumen de datos que pueden ser aprovechados para dar una visión de los tipos de delitos que se cometen en una región específica o de los modus operandi de los delincuentes. Dicha información puede ser recopilada, almacenada y procesada usando técnicas de minería de texto, procesamiento de lenguaje natural y aprendizaje automático para ayudar a cuantificar o evaluar las actividades delictivas de una región en particular.

Se han desarrollado algunos trabajos en la temática criminal apoyados de técnicas de procesamiento de texto (Ananyan, 2004; Chau et al., 2020; Helbich et al., 2013; Iqbal et al., 2012; Keyvanpour et al., 2011; Nokhbeh Zaeem et al., 2017) para extraer patrones importantes de informes policiales con el fin de establecer quién comete los delitos, además de sistemas que ayudan a predecir y georreferenciar las zonas con altas tasas de crímenes (Kianmehr & Alhajj, 2008; Po & Rollo, 2018), sin embargo, pocos trabajos se centran en identificar cuáles son los delitos que se están cometiendo y el grado de similitud entre dichos crímenes. Otra de las dificultades es que la mayoría de los sistemas de apoyo en el área criminal fueron diseñados para procesar texto en el idioma inglés, por lo cual es fundamental la implementación de este tipo de herramientas en el idioma español y en este trabajo se buscó reducir dicha brecha.

En este trabajo en particular, se abordó la clasificación automática de diferentes tipos de delitos de noticias criminales reportados en medios de prensa digital a través del texto en la región del Cauca, Colombia entre los años 2018-2019. Para realizar esta tarea, se propuso la construcción de un conjunto de datos de noticias criminales, el cual fue procesado a través del modelo de bolsa de palabras y la extracción del coeficiente de frecuencia de término–frecuencia inversa de documento (tf-idf), además, se comparó cada documento para construir un modelo de clasificación capaz de distinguir entre ocho tipos de delitos que fueron identificados en el conjunto de datos e incluye, accidentes de tránsito, delitos sexuales, homicidios, hurtos, incendios, secuestros, tráfico de droga y

minería ilegal. Dichas categorías fueron establecidas en (Chen et al., 2004), debido a su grado de daño al público en general, sin embargo, se identificó la minería ilegal como una categoría adicional no reportada y que tiene un crecimiento en el contexto caucano y colombiano.

El resto del documento se encuentra organizado de la siguiente manera: En la sección 2, se presentan los trabajos relacionados con el uso de técnicas de procesamiento de texto y aprendizaje de máquina relacionadas con la temática criminal. La sección 3, describe la metodología implementada en el desarrollo de este trabajo. En la Sección 4, se muestran los resultados y discusiones de la de investigación. Finalmente, la sección 5 describe las conclusiones y el trabajo futuro.

2. Trabajos relacionados

Las técnicas de minería de datos, minería de texto y aprendizaje automático han dado un impulso notable en el campo de la criminalística, puesto que han permitido extraer conocimiento, resolver preguntas de investigación, desarrollar sistemas de apoyo y herramientas para identificar las características de la delincuencia. En este contexto, se han propuesto trabajos como el realizado por (Keyvanpour et al., 2011), quienes implementaron un enfoque basado en técnicas de minería de datos para extraer entidades importantes de informes narrativos policiales. Al utilizar este enfoque, los datos sobre delitos pueden ingresarse automáticamente en las bases de datos de las agencias policiales. Además, implementaron un método de agrupación con un mapa autoorganizado (SOM) en el ámbito del delito y técnicas de agrupamiento jerárquico para realizar un proceso de correspondencia de delito.

En la investigación desarrollada por (Ananyan, 2004), los autores analizaron reportes históricos de un departamento de policía de Virginia, Estados Unidos, donde se identificaron patrones históricos e indagaron la relación entre tipo de delito y la ubicación del incidente, además de los tipos de arma utilizados; para ello implementaron técnicas de minería de texto, con el fin de prevenir, identificar y facilitar el análisis de los delitos. Como consecuencia de este desarrollo, las decisiones que se toman en el departamento policial se basan en el análisis de todos los datos disponibles, incluyendo la porción de texto con mayor cantidad de información, en lugar de un análisis subjetivo que basado principalmente en los analistas con experiencia.

La predicción de la ubicación de los puntos con mayor cantidad de crímenes es una tarea importante para la seguridad pública, ya que puede proporcionar información útil para mejorar las actividades relacionadas con la prevención de problemas de seguridad, por lo que se han desarrollado trabajos como el de (Kianmehr & Alhajj, 2008), donde se implementó un algoritmo de máquina de soporte vectorial (SVM) para predecir la ubicación de crímenes utilizando diferentes bases de datos en las que se reporta la tasa de criminalidad y las variables relacionadas con la localización como la latitud, longitud, entre otras.

En el trabajo desarrollado en (Alatrista-Salas et al., 2020), los autores plantearon una metodología para diferenciar entre noticias relacionadas con el crimen y no relacionadas. Además, asignan a cada noticia un tipo de delito, con el fin de responder

a la pregunta de cómo se cometen los delitos y que tipos de delitos se cometen para ofrecer un termómetro a los ciudadanos sobre la actividad criminal. En este trabajo se implementaron algoritmos de máquina de soporte vectorial SVM y una red neuronal perceptrón multicapa (MLP) obteniendo resultados de puntaje F1 del 84% logrado con el enfoque Doc2vec.

Dentro de los trabajos reportados se identificó que el uso de herramientas basadas en el procesamiento del lenguaje natural y aprendizaje automático presentan buenos resultados en la identificación y clasificación de delitos. Además el trabajo planteado por (Alatrista-Salas et al., 2020), obtuvo resultados prometedores para el idioma español, sin embargo, no se exploraron otros algoritmos supervisados de clasificación como los bosques aleatorios o k-vecinos más cercanos. Adicionalmente técnicas como la validación cruzada con k-pliegues para la elección de un modelo de clasificación, ni tampoco se abordó la elaboración de una lista de palabras vacías propia y adaptada a la temática criminal, por lo cual en este trabajo se presentan aportes de dichas brechas.

3. Metodología

Este trabajo fue desarrollado en cuatro etapas (Figura 1.), primero se construyó el conjunto de datos, luego, se desarrolló una etapa de preprocesamiento haciendo uso de técnicas de procesamiento del lenguaje natural para así obtener un conjunto de datos anotados, posteriormente, se realizó una etapa de modelado donde se implementan técnicas del aprendizaje automático y, finalmente, se evaluó el desempeño del modelo por medio de diferentes índices.

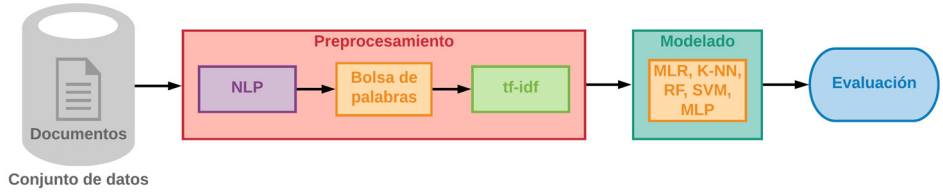


Figura 1 – Metodología

3.1. Construcción del conjunto de datos

El conjunto de datos fue construido a través de la recolección de noticias de índole criminal de dos medios de prensa con publicaciones web de la región del Cauca, Colombia durante los años 2018-2019. Cada una de estas noticias contiene el título y el relato, además fueron almacenadas de forma independiente en archivos de formato .txt en codificación utf-8. No se tuvieron en cuenta noticias repetidas entre los medios de prensa y se realizó un proceso de etiquetado manual en el cual se agruparon las noticias en relación al tipo de delito según lo establecido en (Chen et al., 2004), así: accidentes de tránsito, delitos sexuales, homicidios, hurtos, incendios, secuestros, tráfico de droga, además de la categoría de minería ilegal que no había sido reportada en trabajos anteriores.

3.2. Preprocesamiento

En esta etapa se realizó el preprocesamiento de los documentos, en donde se implementaron técnicas de procesamiento del lenguaje natural (NLP), el cual implica la aplicación de métodos estadísticos con o sin conocimientos de lingüística para comprender los textos (Rao & McMahan, 2019). En este trabajo se generaron representaciones computacionalmente utilizables por medio de trasformaciones a vectores o tensores. Inicialmente se implementó un proceso de tokenización, luego la eliminación de las palabras vacías (stop words), posteriormente una lematización y finalmente se implementó el modelo de bolsa de palabras con el cálculo de la frecuencia de término – frecuencia inversa de documento, así:

Tokenización: En este proceso el texto de cada documento es dividido en unidades denominadas tokens, estas unidades corresponden a cada una de las palabras y se almacenan en un vector. Para determinar las diferencias entre palabras se hace a través de los delimitadores que corresponden a signos de puntuación, espacios en blanco y caracteres que no son alfabéticos.

Eliminación de palabras vacías (stop words): Se eliminaron palabras "sin significado" como: palabras con más de veinte caracteres, símbolos, signos de puntuación, preposiciones, artículos, adverbios, pronombres, conjunciones, nombres propios, meses, días, números y conjugaciones de los verbos ser, estar haber, hacer, ir y tener.

Lematización: Cada palabra fue convertida en su lema. El lema es la palabra flexionada que engloba todas las formas de una palabra. Es decir, eliminar el plural, femenino y conjugación.

Bolsa de palabras: Se implementó un modelo de espacio vectorial o bolsa de palabras a todo el conjunto de documentos, en el cual, se representa cada documento ignorando el orden de las palabras en un espacio multidimensional a través de una matriz *M*(Ecuación 1) siguiendo lo establecido en (Zhang, 2010). En esta matriz, cada fila representa a un documento *d* y cada columna corresponde a un término *t*, además el valor asignado a cada termino refleja una medida de importancia o frecuencia ponderada que produce cada término.

$$M = \begin{pmatrix} d_1 = (w_{1,1}, \dots, w_{m,1}) \\ \vdots \\ d_j = (w_{1,j}, \dots, w_{m,j}) \end{pmatrix} \quad (1)$$

Donde,d_i es el documento y $w_{m,j}$ representa la importancia o frecuencia ponderada del término t_j. La medida de importancia implementada es el coeficiente de frecuencia de término – frecuencia inversa de documento (tf-idf) y expresa la relevancia de una palabra para un documento en un conjunto de documentos (Rao & McMahan, 2019). Este índice es el producto de dos medidas: la frecuencia de termino, $tf(t,d)$, que es la frecuencia bruta del término *t* en el documento *d* y la frecuencia inversa, $idf(t,D)$, de documento, que es una medida de cuan común es un término en la colección de documentos (Ecuación 2.).

$$idf(t,D) = \log \frac{|D|}{|\{d \in D : t \in d\}|} \quad (2)$$

Donde, $|D|$ es el número total de documentos de la colección y $|\{d \in D: t \in d\}|$ es el número de documentos donde aparece el término t. Luego:

$$tf{\cdot}idf(t,d,D) = tf(t,d) \times idf(t,D) \quad (3)$$

Posterior a la representación en bolsa de palabras, se seleccionaron las 100 palabras con los índices más elevados de $tf{\cdot}idf$ para cada uno de los tipos de delitos.

3.3. Modelado

En esta etapa, el conjunto de datos fue dividido en dos conjuntos uno para entrenamiento y otro para validación, dicha partición se realizó conservando el 70% para entrenamiento y el 30% restante para la etapa de prueba. El modelo de clasificación se construyó a partir de los datos de entrenamiento, se implementaron cinco algoritmos de clasificación supervisados para los cuales se ajustaron y variaron cada uno de sus parámetros para seleccionar el modelo con el mejor desempeño, de la siguiente manera:

- Regresión Logística Multinomial: Se implementó el método de Limited-memory-Broyden–Fletcher–Goldfarb–Shanno (LBFGS).
- K-vecinos más cercanos (K-NN): se aplicó la distancia euclidiana y se varió el número de k de 1 hasta 10.
- Bosques aleatorios (RF): se varió el número de árboles de 0 hasta 100.
- Máquina de soporte vectorial (SVM): se aplicaron tres kernels, primero el lineal, luego el polinomial variando el grado del polinomio de 2 hasta 5 y la función de base radial (RBF).
- Red neuronal perceptrón multicapa (MLP): se implementó el algoritmo LBFGS, con una sola capa oculta y se varió el número de neuronas de 1 hasta 100.

Además, se estimó la exactitud, precisión y sensibilidad de cada uno de los algoritmos utilizando el método de k-pliegues y validación cruzada con k=10, puesto que este método permite garantizar que los resultados son independientes de la partición de los datos además de arrojar indicios sobre la estabilidad de los modelos de clasificación (Raschka, 2018).

3.4. Evaluación

En esta etapa se evaluó el desempeño de los modelos en relación con diferentes índices, utilizando el conjunto de datos de prueba compuesto por los documentos no incluidos en la etapa de entrenamiento. Los indicadores de desempeño fueron derivados de la matriz de confusión y se expresan de la siguiente manera:

- Exactitud: número de documentos correctamente identificados dividido entre el número total de documentos.

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

  Donde el verdadero positivo (TP) representa el número total de segmentos normales detectados correctamente y verdadero negativo (TN) representa el número total de segmentos anormales detectados correctamente. Falso positivo (FP) y Falso negativo (FN) representa el número total de errores positivos y errores negativos.

- Sensibilidad: mide la proporción de positivos reales que se identifican correctamente como tales.

$$Se = \frac{TP}{TP + FN} \tag{5}$$

- Precisión: Número de instancias correctamente clasificados dividido entre el número de resultados devueltos.

$$Pr = \frac{TP}{TP + FP} \tag{6}$$

- Puntaje F1: Es la media armónica de la precisión y sensibilidad, donde una puntuación F1 alcanza su mejor valor en 1 (precisión y sensibilidad perfecta) y peor en 0.

$$F1 = \frac{2 \times Pr \times Se}{Pr + Se} \tag{7}$$

- Coeficiente de correlación de Matthews (CCM): Es una medida de la calidad de las clasificaciones binarias y multiclase. Tiene en cuenta los verdaderos y falsos, tanto positivos y negativos, y generalmente se considera una medida equilibrada que se puede utilizar incluso si las clases sonz de tamaños muy diferentes. El CCM es en esencia un valor de coeficiente de correlación entre -1 y +1. Donde un coeficiente de +1 representa una predicción perfecta, 0 una predicción aleatoria promedio y -1 una predicción inversa.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{8}$$

- Área bajo la curva ROC (AUC): El enfoque balanceado más popular para medir la sensibilidad y la especificidad es el análisis ROC (análisis de la característica operativas del receptor) que involucra una representación gráfica de la sensibilidad frente a la tasa de falsos positivos. La información sobre el rendimiento de clasificación en la curva ROC se puede resumir en una puntuación conocida como el área bajo la curva ROC (AUC).

$$AUC = \int_0^1 Pr[TP](v)dv \tag{9}$$

Donde $Pr[TP]$ es función de $v = Pr[FP]$.

4. Resultados y discusión

4.1. Descripción del conjunto de datos

El conjunto de datos recolectado contiene un total de 601 noticias de índole criminal de dos medios de prensa, además, fue dividido en ocho tipos de delitos (Tabla 1.).

Delitos	Medio A.	Medio B.	*Total*
Accidentes Tránsito	47	1	48

Delitos	Medio A.	Medio B.	*Total*
Delitos Sexuales	30	15	45
Homicidios	93	18	111
Hurtos	71	57	199
Incendios	20	2	22
Secuestros	27	10	37
Tráfico Drogas	111	76	187
Minería Ilegal	23	0	23
Total	422	179	601

Tabla 1 – Conjunto de datos de noticias criminales.



Figura 2 – Nube de palabras por categoría. A) Accidentes de tránsito. B) Delitos sexuales. C) Homicidios. D) Hurtos. E) Incendios. F) Minería ilegal. G) Secuestros. H) Tráfico de Drogas.

A partir del modelo de bolsa de palabras se seleccionaron 100 palabras con los índices de tf-idf más elevados para cada uno de los tipos de delitos, estas palabras fueron utilizadas para construir un diccionario por cada clase, dichos diccionarios son representados por los diagramas de nubes de palabras donde las etiquetas con los índices más elevados presentan un tamaño mayor (Figura 2).

4.2. Modelado

Se implementaron cinco algoritmos de clasificación supervisados utilizando el conjunto de datos de entrenamiento con el fin de clasificar ocho tipos de crímenes a través del modelo de bolsa de palabras con el coeficiente de tf-idf, adicionalmente se varió cada uno de los parámetros de los algoritmos de acuerdo con la sección 3.3. y se seleccionaron los parámetros que producían los índices más elevados, posteriormente, se procedió a seleccionar el modelo con el mejor desempeño (Tabla 2.).

Algoritmo	Exactitud	Precisión	Sensibilidad
MLR	92,73 ± 3,97%	91,22 ± 7,09%	83,59 ± 8,13%
KNN	87,11 ± 4,23%	84,37 ± 7,60%	80,59 ± 8,58%
RF	97,44 ± 1,58%	98,67 ± 0,86%	95,97 ± 3,64%
SVM	96,23 ± 1,72%	95,51 ± 4,47%	93,17 ± 4,81%
MLP	95,06 ± 2,59%	92,58 ± 4,93%	92,70 ± 4,37%

Tabla 2 – Evaluación de los algoritmos en la etapa de entrenamiento.

La evaluación y selección del modelo se hizo teniendo en cuenta los índices más elevados establecidos en la sección 3.4 y utilizando el conjunto de entrenamiento con el método de k-pliegues y validación cruzada, que permite verificar el rendimiento del algoritmo a lo largo de todo el conjunto de datos haciendo pliegues de entramiento y prueba. Además de dar una idea sobre la estabilidad del modelo y visualizar cómo este cambia con diferentes divisiones de conjuntos de entrenamiento. Es importante notar que la etapa de prueba fue implementada para estimar el rendimiento de los algoritmos frente a nuevos documentos.

El algoritmo de bosques aleatorios tuvo el mejor desempeño en la etapa de entrenamiento, con una exactitud de 97,44%, precisión de 98,67% y sensibilidad de 95,97%. Sin embargo, no se deben descartar los modelos obtenido de la máquina de soporte vectorial y la red neuronal perceptrón multicapa ya que también presentan resultados elevados. Otra de las ventajas que presentó el algoritmo de bosques aleatorios es su estabilidad a lo largo del conjunto de datos, puesto que la desviación estándar de la exactitud, precisión y sensibilidad fue, en todos los casos, inferior respecto de los demás algoritmos.

Con el conjunto de datos de entrenamiento se evaluó el comportamiento de los diferentes algoritmos respecto a documentos no implementados en la etapa de entrenamiento para verificar la no existencia de subajuste y sobreajuste. Además, se incluyó coeficientes equilibrados como el MCC y el área ROC (Tabla 3.).

Con el conjunto de datos de prueba el algoritmo de bosques aleatorios obtuvo el mejor desempeño con una exactitud del 97,22%, precisión del 98,36%, sensibilidad del 98,35%, Puntaje F1 del 98,32%, MCC del 0,97%, sin embargo, SVM y MLP presentaron valores elevados, es decir que tienen un buen comportamiento en la tarea de clasificar los documentos y que no presenta subajuste o sobreajuste. La matriz de confusión en la etapa de prueba muestra el buen comportamiento del algoritmo RF frente a nuevos documentos y un escaso error en la clasificación de delitos sexuales, hurtos y tráfico de drogas (Figura 3).

Algoritmo	Exactitud	Precisión	Sensibilidad	Puntaje F1	MCC
MLR	92,77	95,15	82,51	84,49	0,91
KNN	87,77	84,79	81,61	82,40	0,85
RF	97,22	98,36	98,35	98,32	0,97
SVM	97.22	96,51	96,07	96,25	0,97
MLP	96,66	93,44	97,16	94,93	0,96

Tabla 3 – Evaluación de los algoritmos en la etapa de prueba.

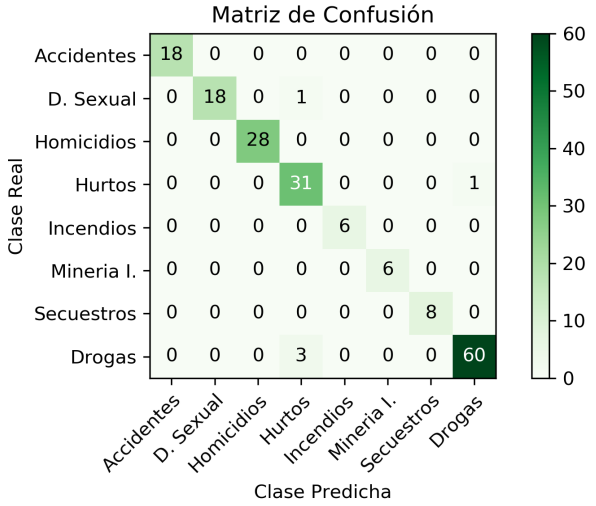


Figura 3 – Matriz de confusión del algoritmo de bosques aleatorios con el conjunto de datos de prueba.

Las curvas ROC y el área AUC fueron estimadas para cada tipo de delito, desde el punto de vista macro (métrica independiente para cada clase) donde se toma cada clase por igual y la micro donde se tiene en cuenta la contribución de todas las clases, es decir

tiene en cuenta el desequilibrio de las clases y en todos los casos se estimó un AUC de 1.0, lo que permite establecer que se obtuvo un buen desempeño (Figura 4).

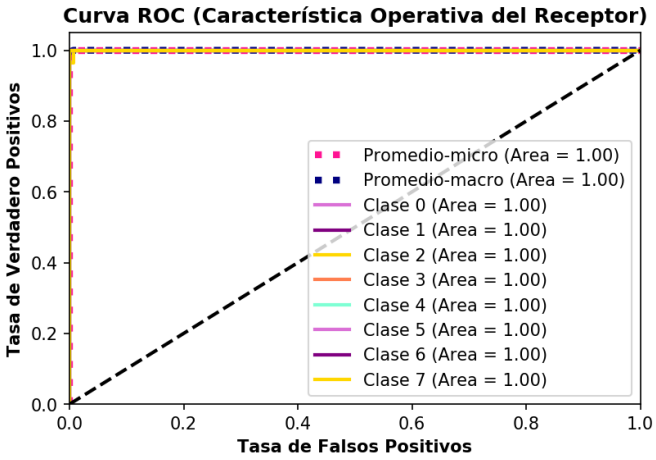


Figura 4 – Curva ROC del algoritmo bosques aleatorios con el conjunto de datos de entrenamiento.

Este trabajo fue diseñado para el idioma español, lo que presenta una ventaja dado que se cuenta con una cantidad limitada de trabajos desarrollados en este idioma, además se identificó que el enfoque plateado podría ser utilizado y adaptado para otros idiomas. También se obtuvo como producto una lista de palabras vacías significativamente mayor a la que comúnmente se encuentra en las librerías de programación y se encuentra adaptado para la temática criminal. Por otra parte, este trabajo posee una la limitación en la etapa de Lematización puesto que se optó por utilizar las reglas de la librería spaCy (Explosion AI, 2018) desarrollada en Python la cual todavía requiere incluir un amplio número de reglas, por lo que se identifica esta carencia de librerías en el idioma español. Otra de las dificultades que presenta el diseño es que puede ocurrir que en un reporte se identifique más de un tipo de delito, por lo cual debería adaptarse para asociaciones entre diferentes temáticas criminales.

5. Conclusiones

Los resultados obtenidos en este trabajo muestran un panorama alentador en el propósito de clasificar diferentes tipos de noticias criminales en formato de texto reportadas en diferentes medios digitales para el idioma español, a partir del uso de técnicas del procesamiento del lenguaje natural y el aprendizaje automático. Donde, estos resultados, metodología y arquitectura pueden ser utilizados para abordar, evaluar y cuantificar que delitos se están cometiendo en una ubicación particular a partir de las noticias.

Se construyeron ocho diccionarios de palabras relacionados con noticias criminales los cuales pueden ser implementados en sistemas de procesamiento de texto y de clasificación de documentos y de esta forma se puede reducir el preprocesamiento que se realiza a los documentos.

El algoritmo que obtuvo el mejor desempeño en todas las pruebas realizadas tanto en la etapa de entrenamiento como la de prueba fue el de bosques aleatorios (RF), siendo capaz de identificar ocho tipos de delitos donde se incluye la categoría de minería ilegal, con una exactitud del 97,22%, precisión del 98,36%, sensibilidad del 98,35%, puntaje F1 del 98,32%, MCC del 0,97% en la prueba realizada. También es importante notar que los algoritmos SVM y MLP obtuvieron resultados elevados y no deben ser descartados para estudios posteriores que implementen este tipo de metodología.

Se identificó que las etapas de recolección del conjunto de datos y el preprocesamiento son fundamentales en el desarrollo de los sistemas de procesamiento de texto y que existen algunas falencias en el desarrollo de las librerías para el idioma español puesto que sus características son limitadas por lo que se requiere mayores investigaciones y desarrollos en este campo.

Como trabajo futuro se plantea anexar un algoritmo de clasificación que permita distinguir inicialmente las noticias criminales de las no criminales para posteriormente implementar los resultados de este trabajo y construir un sistema más especializado. Además, es importante incluir nuevas categorías de delitos, aumentar las subdivisiones específicas de cada delito y adecuarlo para detectar más de un delito en un solo reporte. Asimismo, este trabajo puede ser adaptado para abordar problemáticas relacionadas con la detección de patrones y similitudes entre delitos. Igualmente, las agencias de aplicación de la ley y las organizaciones gubernamentales pueden beneficiarse de este tipo de desarrollos puesto que puede ser adaptado para aumentar la tasa de resolución de delitos y reducir los tiempos de procesamiento de los informes criminales.

Referencias

Alatrista-Salas, H., Morzán-Samamé, J., & Nunez-del-Prado, M. (2020). Crime alert! crime typification in news based on text mining. En *Lecture Notes in Networks and Systems* (Vol. 69, pp. 725–741). Springer. https://doi.org/10.1007/978-3-030-12388-8_50

Ananyan, S. (2004). Crime pattern analysis through text mining. *AMCIS 2004 Proceedings*. https://aisel.aisnet.org/amcis2004/236

Chau, A., Chau, M., Xu, J. J., & Chen, H. (2020). *Extracting Meaningful Entities from Police Narrative Reports Item Type Conference Paper Extracting Meaningful Entities from Police Narrative Reports*. http://hdl.handle.net/10150/105786

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: A general framework and some examples. *Computer*, *37*(4), 50–56. https://doi.org/10.1109/MC.2004.1297301

Estadística delictiva. (2019). https://www.policia.gov.co/grupo-información-criminalidad/estadistica-delictiva

Explosion AI. (2018). *spaCy · Industrial-strength Natural Language Processing in Python*. Official website. https://spacy.io/

Helbich, M., Hagenauer, J., Leitner, M., & Edwards, R. (2013). Exploration of unstructured narrative crime reports: An unsupervised neural network and point pattern analysis approach. *Cartography and Geographic Information Science*, *40*(4), 326–336. https://doi.org/10.1080/15230406.2013.779780

Iqbal, F., Fung, B. C. M., & Debbabi, M. (2012). Mining criminal networks from chat log. *Proceedings - 2012 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2012*, 332–337. https://doi.org/10.1109/WI-IAT.2012.68

Keyvanpour, M. R., Javideh, M., & Ebrahimi, M. R. (2011). Detecting and investigating crime by means of data mining: A general crime matching framework. *Procedia Computer Science*, *3*(3), 872–880. https://doi.org/10.1016/j.procs.2010.12.143

Kianmehr, K., & Alhajj, R. (2008). Effectiveness of support vector machine for crime hot-spots prediction. *Applied Artificial Intelligence*, *22*(5), 433–458. https://doi.org/10.1080/08839510802028405

Nokhbeh Zaeem, R., Manoharan, M., Yang, Y., & Barber, K. S. (2017). Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers and Security*, *65*, 50–63. https://doi.org/10.1016/j.cose.2016.11.002

Po, L., & Rollo, F. (2018). Building an Urban Theft Map by Analyzing Newspaper Crime Reports. *Proceedings - 13th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2018*, 13–18. https://doi.org/10.1109/SMAP.2018.8501866

Rao, D., & McMahan, B. (2019). *Natural language processing with PyTorch: build intelligent language applications using deep learning*.

Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. http://arxiv.org/abs/1811.12808

Vega, M. G. de la. (2014). *La noticia policial, un género periodístico diferente : un análisis del cómo y el porqué del lenguaje periodístico usado por Clarín y Crónica en el caso García Belsunce.* [Universidad de San Andrés]. https://repositoriosdigitales.mincyt.gob.ar/vufind/Record/RDUDESA_7fac05c3d43b0be65c1d3ce039459722

Zhang, N. I. S. M. W. T. (2010). *Fundamentals of Predictive Text Mining*. Springer, London. https://doi.org/https://doi.org/10.1007/978-1-84996-226-1

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 30/09/2020
Aceitação/Acceptance: 23/11/2020

Inteligencia de Negocios para el Análisis de la Accidentalidad Vial en la Ciudad de Popayán

Carlos H. Bolaños M.[1], Inés A. Damián F.[1], Héctor F. Muñoz M.[1], Andrés F. Negrete G.[1], Miguel A. Tunubalá M.[1], Christian G. Arias I.[1], María A. Varona T.[1]

cbolanos@unimayor.edu.co; idamian@unimayor.edu.co; hfmunoz@unimayor.edu.co; andresfng123@unimayor.edu.co; maicolanderst@unimayor.edu.co; charias98@unimayor.edu.co; mvarona@unimayor.edu.co

[1] Institución Universitaria Colegio Mayor del Cauca, Cra. 7#2-34, 190003, Popayán, Colombia.

Pages: 130-141

Resumen: En este artículo se presenta una solución basada en Inteligencia de Negocios para realizar el análisis e interpretación de los datos de accidentalidad vial, con el propósito de generar conocimiento que contribuya en la prevención de accidentes de tránsito en la ciudad de Popayán, los cuales se están convirtiendo en una de las mayores preocupaciones para las autoridades locales, con un promedio de más de 6 accidentes diarios según el Plan Vial y de Transporte Municipal. Haciendo uso de la Metodología propuesta por Kimball se realizó el análisis usando los datos obtenidos directamente del Departamento de Tránsito Municipal que permitieron integrar las diferentes variables, elaborar reportes y generar el conocimiento requerido para establecer posibles soluciones a la necesidad planteada.

Palabras-clave: Accidentalidad; Inteligencia de Negocios; Metodología Kimball; Prevención; Tránsito.

Business Intelligence for the Analysis of Road Accidents in the City of Popayán

Abstract: This paper presents a solution based on Business Intelligence to perform the analysis and interpretation of road accident data, in order to generate knowledge that contributes to the prevention of traffic accidents in the city of Popayán, which are being becoming one of the biggest concerns for local authorities, with an average of more than 6 accidents per day according to the Road and Municipal Transportation Plan. Using the Methodology proposed by Kimball, the analysis was performed using data obtained directly from the Municipal Traffic Department, which allowed integrating the different variables, preparing reports and generating the knowledge required to establish possible solutions to the need presented.

Keywords: Accidentality; Business intelligence; Kimball Methodology; Prevention; Transit.

1. Introducción

Los accidentes de tránsito en el año 2011 reportados por la Organización Mundial de la Salud son oficialmente declarados como un problema de salud pública, por ser una de las mayores causas de muerte a nivel mundial, según estimaciones hechas por esta entidad, a nivel mundial más de 1,2 millones de personas fallecen por esta causa y alrededor de 50 millones quedan lesionadas (Vargas, 2013), el factor humano (tanto conductor como peatón) es de gran incidencia en el desencadenamiento de los accidentes, debido a que en la gran mayoría de estos se registra como causal de un error humano, ya sea por distracción o imprudencia en el momento de transitar por las vías de la ciudad. Actualmente, Colombia presenta una alarmante cifra de muertes por accidentes de tránsito; según los datos entregados por el Instituto Nacional de Medicina Legal, a julio de 2018 se habían registrado 3.704 muertes por accidentes de tránsito, y ha sido considerada la segunda causa que más genera muertes en el País (Medicina Legal, 2018). En términos de tasas por 100.000 habitantes, Colombia para el 2013 cuenta con una tasa de 13,2 muertes, que si bien es un número inferior por tres puntos al promedio de la tasa registrada por la región de las Américas con una tasa de 16,1, es de resaltar que después del año 2002 es la segunda vez que Colombia cierra con una cifra superior a 13, presentando el mismo fenómeno de 2012. Lo anterior evidencia un retroceso en la lucha contra el flagelo de la inseguridad vial en la última década, dado que el mejor registro colombiano se presentó en el 2004 con un indicador de 12,1 muertes por cada 100.000 habitantes (Vargas, 2013). En cuanto respecta a la ciudad de Popayán, la accidentalidad genera un promedio de más de 6 accidentes diarios, cifras preocupantes a pesar de tener niveles menores que otras ciudades del País (Alcaldía de Popayán, 2019). El análisis de datos de accidentes de tránsito puede contribuir a entender las características en el comportamiento de los conductores, las condiciones de las vías, condiciones climáticas y detectar zonas de mayor ocurrencia de accidentes de tránsito. La información generada puede ayudar a las entidades de control a la toma de decisiones, permitiendo formular mejores políticas de control de la seguridad del tráfico(Chong et al., 2005). Un análisis de información cuya finalidad sea predecir los índices de accidentabilidad en la ciudad de Popayán puede constituir un apoyo importante para la comunidad y entidades responsables del tema de viabilidad de la ciudad de Popayán a la hora de tomar decisiones que permitan reducir al máximo estos índices para el bien de la comunidad.

2. Metodología

Para la implementación basada en inteligencia de negocios se empleó una investigación de tipo descriptiva bajo una metodología cuantitativa, en este caso la propuesta por Kimball, denominada ciclo de vida dimensional del negocio, con los siguientes principios básicos.

Centrarse en el negocio: Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.

Construir una infraestructura de información adecuada: Diseñar una base de información única, integrada, fácil de usar y de alto rendimiento donde se refleje la amplia gama de requerimientos de negocio identificados en la empresa.

Realizar entregas en incrementos significativos: Crear el almacén de datos (Data Warehouse) en incrementos entregables en plazos. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.

Ofrecer solución completa: Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

Para la construcción de una solución Data Warehouse / Bussiness Intelligence, se utilizó la metodología que propone Kimball realizando las siguientes tareas:

2.1. Planificación

Se determina el propósito del proyecto de Data Warehouse / Bussiness Intelligence, los objetivos específicos, el alcance, los riesgos y una aproximación a las necesidades de información del proyecto.

2.2. Análisis de requerimientos

Se conoce la organización, el negocio, las estrategias, las necesidades y se definen los requerimientos tanto inferidos como pedidos, y se priorizan los más críticos.

Se determinan los siguientes requerimientos en el análisis de los datos:

El lugar o sector con mayor número de accidentes, con número de muertes.

El número de accidentes que fueron causal de muerte por fecha, hora, lugar y diseño de la vía.

El promedio de accidentes en días festivos frente a el promedio de accidentes en días no festivos.

El número de implicados por clase de accidente y gravedad de accidente.

El número de vehículos involucrados por tipo, servicio, sector y tipo de accidente.

La cantidad mensual de implicados por tipo de víctima y ubicación.

2.3. Modelado dimensional

Se define el área a modelar, se especifica el nivel de granularidad, se definen las dimensiones, se identifican las tablas de hechos y medidas, se realiza el modelo gráfico de alto nivel, se identifican los atributos de dimensiones y tablas de hechos, se implementa el modelo dimensional detallado, se realiza la prueba del modelo, se revisa y valida el modelo propuesto para realizar los documentos finales.

En la figura 1 se representa el modelo dimensional el cual describe 3 principales hechos relevantes para el estudio:

1. Accidentalidad: Busca describir diferentes características en la ocurrencia de accidentes.
2. Víctimas involucradas: Se enfoca en la gravedad de las lesiones presentadas en las personas involucradas en un accidente de tránsito.
3. Vehículos involucrados: Describe las características de los vehículos.

Así mismo se representa el modelo dimensional que describe 9 dimensiones:

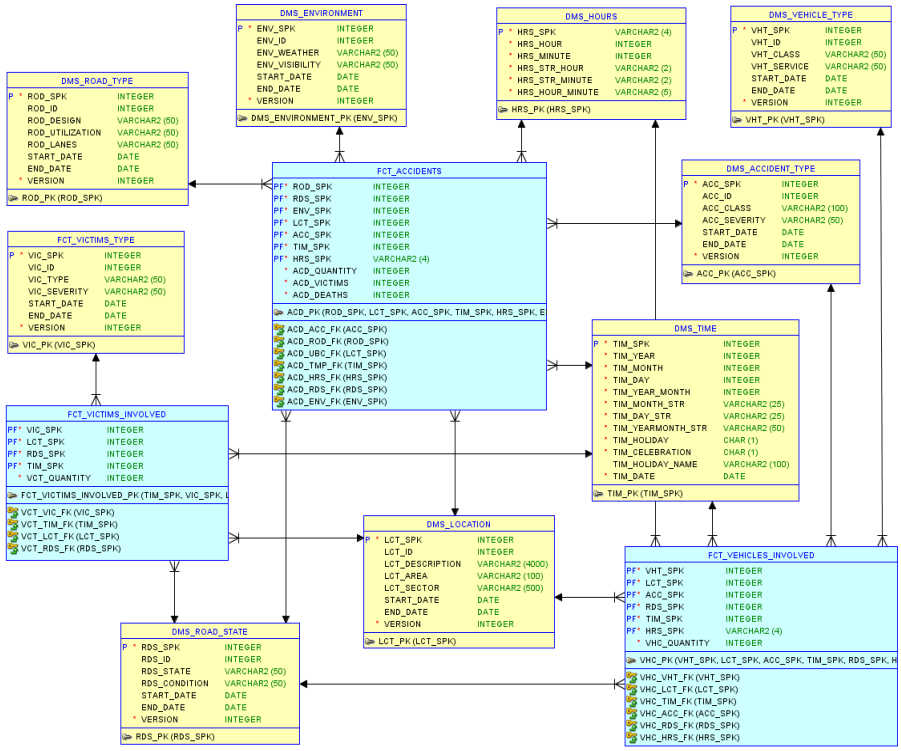


Figura 1 – Modelo dimensional

1. Tipo de accidente: Es una clasificación de los accidentes (choque, atropello, volcamiento, entre otros).
2. Ubicación: Describe la localización geográfica del accidente.
3. Tiempo: Esta dimensión nos permite tener una perspectiva temporal de los accidentes en días, mes, año y festividad
4. Horas: Describe el momento temporal de la ocurrencia del accidente en hora y minutos.

5. Tipo de vehículo: Describe la clase del vehículo (automóvil, motocicleta, etc.) así como también el tipo de servicio que presta (particular, público, oficial, o diplomático).
6. Tipo de víctima: Describe el rol de la persona en el momento del accidente (acompañante, peatón, conductor, pasajero, ciclista) así como la severidad de las lesiones presentadas (herido, muerte, sin gravedad).
7. Entorno: Describe las condiciones externas en el momento de la ocurrencia del accidente, condiciones meteorológicas y visibilidad.
8. Tipo de vía: Es una clasificación del diseño de la vía (glorieta, intersección, lote, paso a nivel, paso inferior, zona peatonal, puente, tramo lineal de vía).
9. Estado de la vía: Es una descripción de la vía al momento de ocurrir el accidente (bueno, con huecos, derrumbes, en reparación, fracturada, hundimiento, inundada, parchada) así como también la condición de la vía (húmeda, seca, lodo, material suelto, aceite en la vía).

2.4. Diseño Físico

Se definieron aspectos físicos del Data Warehouse, aspectos como cuán grande debe ser el sistema, la configuración, memoria, servidores y software instalado, entre otros.

2.5. Diseño del sistema de ETL

Se define el sistema de extracción, transformación y carga, buscando aumentar la calidad y consistencia de los datos, después se consolida la información y se carga en el Data Warehouse en un formato acorde para utilización de las herramientas definidas, para este proceso se trabajará con la herramienta Pentaho Data Integration (PDI, también llamado Kettle).

Cada proceso en Pentaho se crea con una herramienta gráfica donde se especifica qué hacer sin escribir código para indicar cómo hacerlo; debido a esto, se podría decir que PDI está orientado a los metadatos, siendo una herramienta de código abierto más popular disponible que nos ayuda a generar procesos de ETL, también soporta una amplia gama de formatos de entrada y salida, incluyendo archivos de texto, hojas de datos y motores de base de datos comerciales y gratuitos. Además, las capacidades de transformación de PDI permiten manipular datos con muy pocas limitaciones, en la Figura 2 se representa el proceso Job para ejecución ETL.

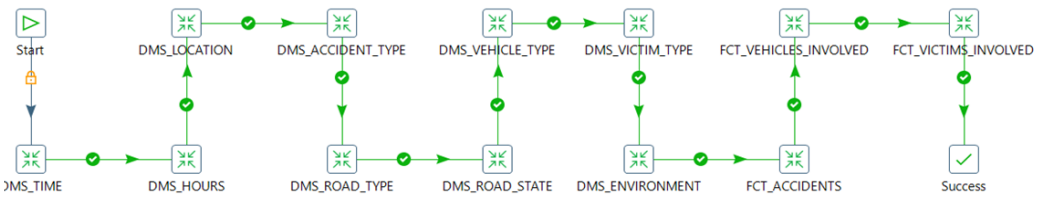


Figura 2 – Job para ejecución de ETL

2.6. Especificación y desarrollo de aplicaciones de BI

Se realizaron los tipos de informes, herramientas de análisis y aplicaciones analíticas que serán entregadas al usuario final, con que uso y consumo del Data Warehouse.

A continuación, en las siguientes figuras se evidencian los cubos resultantes:

En la Figura 3 se describe la ocurrencia de accidentes, cantidad de personas implicadas y la cantidad personas fallecidas, según sea el sector o barrio de la ciudad en un tiempo determinado. Esto nos permite identificar los sectores con mayor número ocurrencia de accidentes, implicados y muertes en la ciudad, en esta figura se puede observar que el barrio "El modelo" registra una mayor cantidad de accidentes con un total de 265, donde se involucraron 549 personas de las cuales 3 personas resultaron fallecidas en el año 2.015.

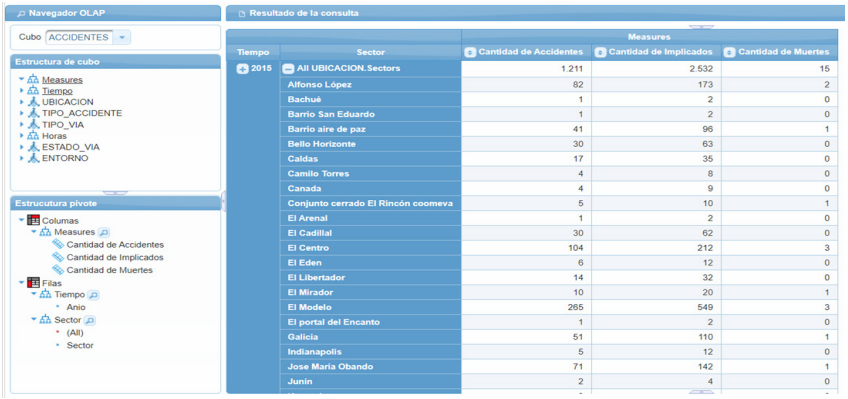


Figura 3 – Cubo accidentes

En la Figura 4 lo que se busca con el Cubo de vehículos es describir la cantidad de vehículos involucrados en un accidente según sean los tipos de vehículos y el tipo de servicio que prestan en un tiempo determinado. Así, podemos ver en la Figura 4 que los vehículos de tipo automóvil que prestan servicios particulares, son lo que más registran accidentes en la ciudad, según los datos registrados de los años 2.015, 2.016 y 2.017.

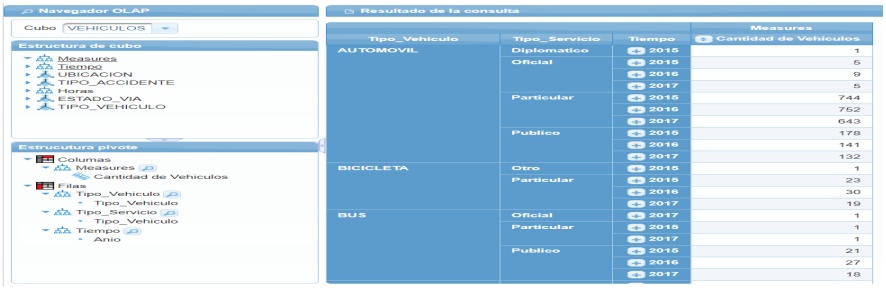


Figura 4 – Cubo vehículos

En la Figura 5 en el Cubo de implicados se describen la cantidad de personas implicadas en accidentes de tránsito, clasificados en el tipo de rol que tenía en el vehículo al momento del accidente y la gravedad de las lesiones presentadas. La Figura 5 muestra que el conductor del vehículo representa un alto número de registros de heridos y fallecimientos.



Tipo_Victima	Gravedad_Victima	Tiempo	Measures: Cantidad de Implicados
Acompanante	Herido	2015	85
		2016	91
		2017	60
	Muerte	2016	4
		2017	1
	Sin Gravedad	2015	1
		2016	2
Ciclista	Herido	2016	1
		2017	1
	Sin Gravedad	2015	1
Conductor	Herido	2015	259
		2016	221
		2017	172
	Muerte	2015	10
		2016	6
		2017	7
	Sin Gravedad	2015	2.060
		2016	2.072
		2017	1.690
Pasajero	Herido	2015	23

Figura 5 – Cubo de implicados

Una vez realizado el análisis de la información como se muestra en la Figura 6 en el Dashboard de accidentabilidad los resultados encontrados son clasificados por clase de accidente y gravedad en las lesiones de los involucrados. Esto permite que el usuario final comprenda la información que se está analizando y permita realizar una toma de decisiones.

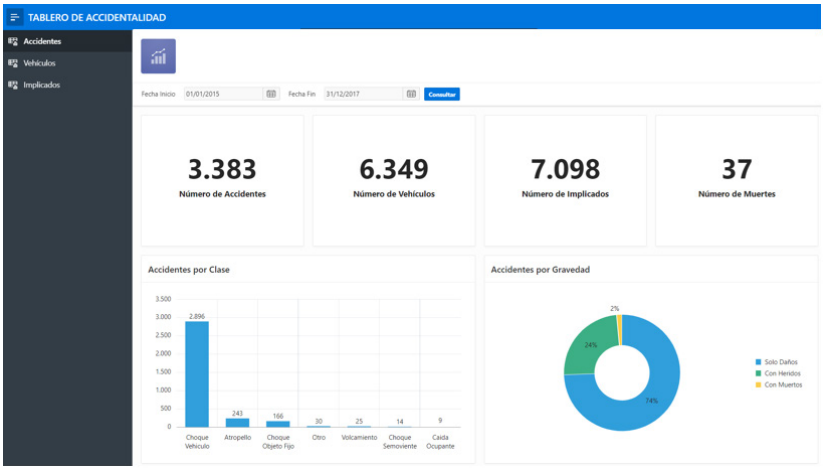


Figura 6 – Dashboard de Accidentabilidad

3. Resultados

Después del análisis de datos se encontró que los sectores con mayor nivel de accidentalidad es la zona centro de Popayán, puntualmente el barrio modelo y el barrio centro.

En consecuencia, se hace necesario fortalecer el personal de control de tránsito y la señalización vial en dichos sectores, como se evidencia en la Figura 7.

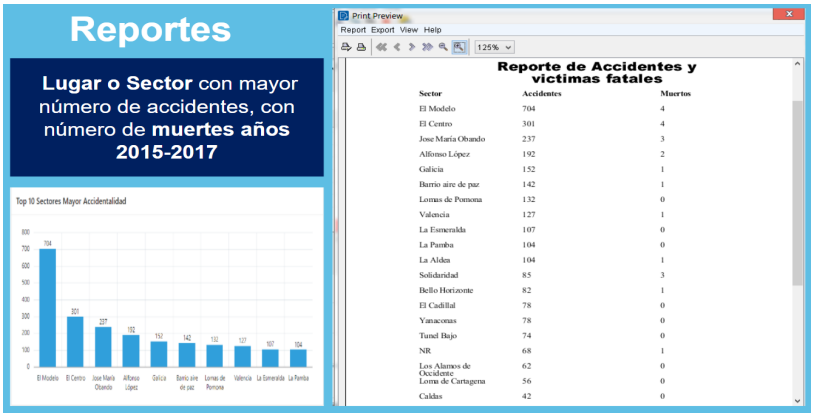


Figura 7 – Reporte de accidentes y víctimas fatales

En la Figura 8 se evidencia la incidencia de accidentalidad en días festivos es inferior a la de días no festivos, en proporción directa a la disminución de actividades en los días festivos.



Figura 8 – Reporte del promedio de accidentes días festivos y no festivos

En la Figura 9 resulta claro que los tramos de vía son los lugares donde mayor accidentalidad se presenta, muy seguramente ligado a la mayor velocidad de circulación y en segundo lugar se encuentran las intersecciones, donde la imprudencia es la principal causa de accidentalidad.

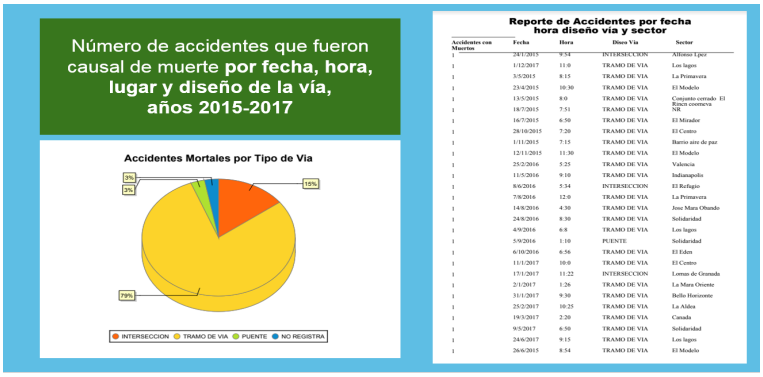


Figura 9 – Reporte de accidentes por fecha hora diseño vía y sector

En la Figura 10 se evidencia que los meses de mayor accidentalidad: octubre, abril y mayo, además se pueden evidenciar los meses de menor incidencia: febrero, julio y enero.

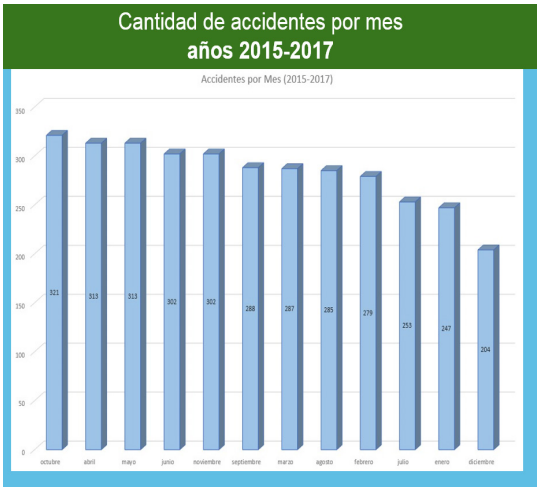


Figura 10 – Cantidad de accidentes por mes años 2015.2017

En la Figura 11 basados en datos entre el 01 de enero de 2015 a 31 de diciembre de 2017, el principal implicado en un accidente de tránsito es el conductor, seguido por el peatón y los acompañantes. En un porcentaje del 82% de los accidentes de tránsito no presentan gravedad, donde el 17% equivale a los heridos y el 1% a las muertes.

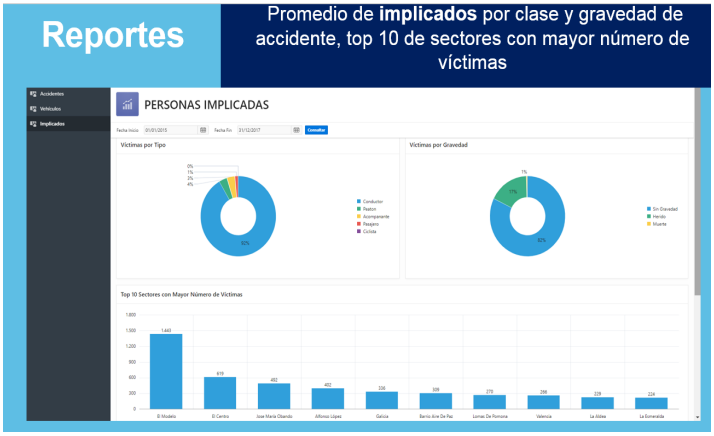


Figura 11 – Promedio de implicados clase y gravedad de accidente

En la siguiente figura los vehículos por tipo con mayores incidencias son los automóviles seguidos por las motocicletas y las camionetas y los vehículos clasificados por servicio el sector con mayor incidencia es el particular, seguido por el sector público. También se puede determinar la clasificación por vehículos por clase de accidente, el mayor porcentaje equivale choque de vehículo y en los vehículos por gravedad de accidente los daños representan al porcentaje más alto.



Figura 12 – Vehículos involucrados por tipo de servicio gravedad y clase de accidente

Finalmente, en la Figura 13 se representan los resultados de la accidentabilidad vial en la ciudad de Popayán entre los años 2015 a 2017, donde los factores con porcentajes más altos en número de accidentes de tránsito en la ciudad de Popayán se encontraron: el mes de octubre, el barrio modelo, el barrio centro, los días no festivos, el tramo en la vía y solo daños.

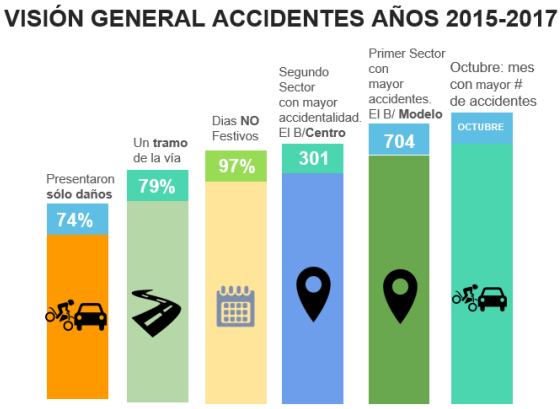


Figura 13 – Visión General de la accidentabilidad vial en Popayán entre los años 2015 a 2017

4. Conclusiones y trabajos futuros

Se puede realizar un proceso de adopción de nuevas tecnologías que permitan auditar y controlar el comportamiento de los incidentes viales de manera permanente y se hace obligatorio la importancia de contar con los entes territoriales para disminuir el impacto producido en los ciudadanos y en el presupuesto estatal.

La accidentalidad vial genera un profundo impacto en la ciudadanía en general y en los entes territoriales. Las personas se ven afectadas tanto en su integridad física como en pérdidas humanas y materiales. Los entes territoriales se han visto seriamente afectados ya que las demandas por falta de señalización y mantenimiento presentan un marcado incremento generando altísimos costos.

Resulta fundamental que las entidades territoriales asuman el compromiso de mantener un adecuado estado de la malla vial para no incurrir en los costos generados por multas, demandas y sanciones.

Después de realizar una solución basada en inteligencia de negocios se evidencia la importancia de conocer los factores más relevantes en los accidentes de tránsito y con estos resultados ayudar al Departamento de Tránsito Municipal en la toma de decisiones como alertas y prevención de los mismos.

Se podrán adoptar herramientas tecnológicas que permitan controlar la accidentalidad vial y realizar verificaciones permanentes para identificar comportamientos anómalos de manera temprana.

Para fortalecer el análisis de la información y obtener mejores resultados se recomienda acudir a otras fuentes de datos como el Registro Único Nacional de Tránsito RUNT y tener en cuenta el Sistema que integra el registro de infractores a nivel nacional y controla la no realización de trámites SIMIT.

Referencias

Alcaldía de Popayán. (2019). *Plan de Ordenamiento Territorial de Popayán*. Retrieved from: http://www.popayan.gov.co/sites/default/files/documentosAnexos/pot-popayan.pdf

Chong, M., Abraham, A., & Paprzycki, M. (2005). Traffic Accident Analysis Using Machine Learning Paradigms. *Informática*, 29(1), 89-98.

Medicina Legal. (2018). *Subdirección de Servicios Forenses Centro de Referencia Nacional sobre Violencia CRNV Boletín Estadístico Mensual.* Retrieved from: https://www.medicinalegal.gov.co/documents/20143/217010/Boletin+estadistico+mensual+-julio-2018.pdf/4266e5fc-8846-9a44-ac4f-637db93f790c

Vargas, D. (2013). *Comportamiento de muertes y lesiones por accidente de transporte*. Retrieved from: https://www.medicinalegal.gov.co/documents/20143/49517/Accidentes+de+transporte.pdf

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 19/09/2020
Aceitação/Acceptance: 13/11/2020

Niveles de Autoeficacia Matemática en Estudiantes Mujeres y Varones Aspirantes a la Carrera de Ingeniería

Claudia Mego-Sanchez[1], Ling Huaman-Sarmiento[1], Ivan Iraola-Real[1], Alonso Iraola-Arroyo[2]

iiraola@uch.edu.pe; ling.katterin@gmail.com; claudiasmg01@gmail.com; air_2021@outlook.com

[1] Universidad de Ciencias y Humanidades, Los Olivos, Lima 15314, Lima, Perú.

[2] Universidad Tecnológica del Perú, Ate, Lima 15487, Santa Clara, Perú.

Pages: 142-155

Resumen: En la carrera de ingeniería, la autoeficacia matemática es esencial, porque representa el nivel de confianza para resolver de forma óptima los problemas matemáticos. Esta autoeficacia ha sido estudiada identificando diferencias de género que desfavorecen a las mujeres; problemática por la cual, el objetivo del presente estudio fue identificar los niveles de autoeficacia matemática en mujeres y varones; así también, analizar su relación con la aplicación de estrategias de estudio en matemáticas. La muestra fue de 122 estudiantes aspirantes a las carreras de ingeniería de Lima – Perú, de los cuales 66 fueron mujeres (*54.1%*) y 56 varones (*45.9%*) entre 15 a 26 años de edad ($Media^{edad}$ *= 17.82; DE: 1.75*). Finalmente, en los resultados se identificó que las mujeres poseen menores niveles de autoeficacia matemática, que se relaciona con el uso de una menor cantidad de estrategias de estudio; pero poseen el mismo nivel de desempeño académico en matemáticas.

Palabras-clave: Autoeficacia matemática, diferencias de género, estrategias de estudio, carrera de ingeniería, desempeño académico.

Diagnostic study between the levels of mathematical self-efficacy among women and men aspiring to the engineering career

Abstract: In the engineering career, mathematical self-efficacy is essential, because it represents the level of confidence to optimally solve mathematical problems. This self-efficacy has been studied by identifying gender differences that disadvantage women; problematic by which, the objective of this study was to identify levels of mathematical self-efficacy in women and men; so too, analyze their relationship to the application of study strategies in mathematics. The sample was of 122 aspiring students to the engineering careers of Lima – Peru, of which 66 were women (*54.1%*) and 56 males (*45.9%*) between 15 and 26 years of age ($Mean^{age}$ *= 17.82; SD: 1.75*). Finally, the results identified that women have lower mathematical self-efficacy levels, which relates to the use of fewer study strategies; but they have the same mathematics academic performance level.

Keywords: Mathematical self-efficacy, gender differences, study strategies, engineering career, academic performance.

1. Introducción

La elección de la carrera profesional suele ser una decisión importante para la vida de las personas, debido a la existencia de una gran variedad de profesiones y oficios, teniendo en cuenta que esta elección está ligada a los gustos y preferencias de cada uno, y que a su vez estas van variando de acuerdo con las vivencias y experiencias que son parte de la autorrealización personal (Thenmozhi, 2018). Así, el caso peruano no es la excepción; sin embargo, entre todas las carreras universitarias existentes las que más se consideran complicadas son las carreras de ingeniería (Instituto Nacional de Estadística e Informática [INEI], 2015) porque demandan un amplio conocimiento y buen desempeño en las matemáticas, los que se van adquiriendo desde la escuela. Este nivel de conocimiento y desempeño en matemáticas es producto de la autoeficacia matemática; la cual es comprendida como el nivel de confianza y creencias positivas que los estudiantes poseen de sus aptitudes académicas para resolver problemas aritméticos (Bandura, 1997).

Por estas razones, resulta necesario profundizar en el estudio de la autoeficacia y sus orígenes (Bandura, 1977); para ello, urge conocer marco conceptual de la autoeficacia académica, la autoeficacia matemática (Bandura, 1997; Usher & Pajares, 2009) y sus diferencias de género en diversos estudios empíricos en población estudiantil (Peters, 2013). Así como la relación entre la autoeficacia matemática y la carrera de ingeniería, siendo un punto fundamental para el estudio de las ingenierías, puesto que permitiría mejorar su proceso de aplicación desde la formación profesional, así como implementar medidas en la praxis.

2. Marco conceptual

2.1. La autoeficacia y algunos estudios relacionados

El concepto autoeficacia surge desde la Teoría Cognitivo Social desarrollada por Albert Bandura (1977), en la que explica cómo las personas elaboran percepciones de sus propias capacidades en todo lo que hacen. La autoeficacia involucra un conjunto de juicios de los eficientes (o ineficientes) que las personas son al enfrentar situaciones inciertas según sus logros obtenidos (Bandura, 1997); que pueden variar según los diferentes ámbitos en los que se desarrollen (Bandura, 2006). Por ejemplo; a nivel personal y educativo, los estudios de Schunk y Meece (2006) demostraron que la autoeficacia se desarrollaba dependiendo del entorno familiar y educativo; siendo este último en el cual, los compañeros de estudios influían de forma significativa. De forma simultánea se estudiaron los beneficios de la autoeficacia en la reducción de la ansiedad (Bandura, 1997), en el incremento de la perseverancia (Pajares & Urdan, 2006) y del desempeño académico (McCoach, Gable & Madura, 2013). Pero, por otro lado, se estudiaron los efectos negativos de los débiles niveles de autoeficacia en el incremento del estrés (Pajares, 2002), la ansiedad (Roick & Ringeisen, 2017), y la procrastinación (Katz, Eilot

& Nevo, 2014); evidenciando la necesidad de profundizar los estudios de la autoeficacia en los entornos académicos.

2.2. La autoeficacia en el campo académico

El concepto autoeficacia involucra creencias que tienen las personas de sí mismos sobre la capacidad de desempeñar una labor, siendo así que los sujetos con más triunfos desplegarán más creencias positivas de autoeficacia (Bandura, 1997). Pero, por el contrario, las experiencias de fracaso contribuyen en las sensaciones de autoeficacia negativas generadas por la angustia, la ansiedad, la tristeza, el miedo, la ira, entre otros (Lima & Nunes, 2017). Y, como consecuencia del tipo de autoeficacia que se desarrolla, las personas con creencias positivas de sus propias capacidades van a emprender las tareas con más calma y tolerancia; pero, para aquellas con creencias negativas, el trabajo será más tedioso y complicado (Olivas & Barraza, 2016). En efecto, la autoeficacia influye en la comprensión de las exigencias propias, de modo que las personas con altos grados de eficacia tendrán mayor probabilidad de identificar situaciones problemáticas como retos, descartando que sea un riesgo (Armando, 2016). Y, en consecuencia, la autoeficacia, en el campo académico, predice resultados, permitiendo que los estudiantes se consideren más aptos para la resolución de diversas tareas como las de matemáticas (Lima & Nunes, 2017).

La autoeficacia en el campo académico es la confianza en las propias capacidades de poder afrontar con éxito las tareas educativas (Bandura, 1997), difiriendo en cada estudiante al desarrollar óptimas creencias de autoeficacia en una materia y en otra no (Bandura, 2006), como en las matemáticas. Así, la autoeficacia en el aspecto educativo genera mayor seguridad en los estudiantes reduciendo la ansiedad en las evaluaciones (Roick & Ringeisen, 2017) e incrementando el desempeño académico (Honicke & Broadbent, 2016; McCoach, et al., 2013).

2.3. La autoeficacia matemática y la carrera de ingeniería

Según Albert Bandura (1997) la autoeficacia desarrolla convicciones que permiten evaluar conflictos y lograr tareas en base a la confianza o desconfianza propia. Demostrando así ser esencial en diversos campos educativos (Bandura, 2006), para el desarrollo de la autoeficacia matemática (Usher & Pajares, 2009). Este tipo de autoeficacia se desarrolla debido a las experiencias exitosas (donde el estudiante percibe cierto grado de dominio), a la observación de sus compañeros con capacidades similares, a la persuasión positiva que reciben de las personas y a la propia percepción de seguridad adquirida en los estudios (Schunk, Meece & Pintrich, 2014).

La autoeficacia matemática surge de las experiencias de dominio, es decir, de vivencias exitosas en las matemáticas. Por ejemplo: en la resolución de problemas matemáticos el aprendiz que estudia situaciones de su contexto logrará mayor competitividad (Plaza, 2016), y si estudia una gama amplia de ejercicios matemáticos aprenderá estrategias y de destrezas alcanzando la excelencia (Tárraga, 2008). Y respecto a las experiencias vicarias, para Bandura (1997), la autoeficacia matemática también surge como producto de ver a otros enfrentar retos matemáticos, sintiendo así el deseo de afrontarlos, y al superarlos desarrollar autopercepciones positivas de cuán capaz se es. De esta manera,

la autoeficacia matemática es producto de la motivación que generan las experiencias vicarias, como consecuencia de observar a los compañeros de clase tener éxito en la resolución de problemas matemáticos (Usher & Pajares, 2009).

Se observa que, la autoeficacia matemática se desarrolla vinculando la educación a la realidad, principalmente en los estudiantes de ingeniería, para que así cuenten con la motivación de estudiar su entorno logrando la eficiencia (Plaza, 2016). Esto es posible aplicando la Educación Matemática Realista (EMR) porque el vincular la matemática a la realidad (Zolkower, Bressan & Gallego, 2006) favorece a que los estudiantes resuelvan cada vez mejor los ejercicios matemáticos (Yuanita, Zulnaidi & Zakaria, 2018) incrementando su confianza en sus logros de aprendizaje (Carmona, 2015) y su autoeficacia matemática (Bandura, 1997). Además, se debe considerar que, en el futuro, el estudiante será insertado al exigente mundo laboral, de modo que se verá envuelto en situaciones complejas, debiendo hacer uso de sus habilidades fomentando nuevos conocimientos y estrategias de análisis (Plaza, 2016).

2.4. Diferencia de género en la autoeficacia matemática e ingenierías

En el campo de la ingeniería, la autoeficacia matemática podría diferir entre mujeres y varones. Esto a consecuencia que, durante la historia de la humanidad, la mujer ha sido catalogada como alguien carente de las capacidades necesarias, siendo común que asuma labores dependientes (Rodríguez, 2014). Y el campo de la ingeniería no es la excepción, en el cual se cree que las mujeres tendrán menor desempeño que los varones con ser mujeres y en muchos casos por ser madres; situación a la que se suma la carencia de apoyo familiar y/o conyugal; que terminan por generar desmotivación para dedicarse a la ingeniería (Rodríguez, 2016).

Según Peters (2013) existe la tendencia a percibir que en las matemáticas las mujeres no son igual de eficaces que los varones. Hecho que se ha observado en el sector educativo que tiene como patrón definido que el género masculino sobresale en asignaturas relacionadas a los números, dejando a las féminas, como las mejores en demostrar sus sentimientos (Ruíz & Santana, 2018). Incluso, los estudiantes desde edades tempranas reconocen en sí mismos cuáles son las capacidades en las que mejor se desenvuelven, evidenciando que son los niños los que tienen mayor destreza en las matemáticas y las niñas en las artes y lingüística (Del Río, Strasser & Susperreguy, 2016). Al respecto, se tiene como ejemplo a los estudios de Joët, Usher y Bressoux (2011) quienes investigaron 395 estudiantes franceses de tercer grado de primaria, en quienes identificaron que las niñas superaban a los varones en el aprendizaje de la lengua francesa y los niños poseían mayor autoeficacia matemática producto de experiencias exitosas en esta materia.

Y en educación superior se han podido identificar hallazgos similares. Por ejemplo: los estudios de Michelle Peters (2013) quien investigó una muestra de 326 estudiantes universitarios identificando que los varones poseen mayores niveles de autoeficacia matemática a diferencia de las mujeres. Esta situación en educación superior genera preocupación debido a que los débiles niveles de autoeficacia matemática generan altos niveles de ansiedad ante los exámenes (Reiner, 2017; Usher & Pajares, 2009). Situación que preocupa más a las adolescentes aspirantes a las carreras profesionales, quienes, en Perú, manifiesta elevados niveles de ansiedad en los exámenes de admisión a las

universidades (Chávez, Chávez, Ruelas & Gómez, 2014) presentando deficiencias en su desempeño académico (Chávez, Morocho, Alvites, Vega & Solis, 2017).

Los estudios mencionados podrían explicar los motivos por los cuales, la carrera de ingeniería en su mayoría es estudiada por varones, debido a que durante mucho tiempo ha sido considerada netamente para varones (Rodríguez, 2016), hecho que influye negativamente en muchas mujeres quienes consideran que su género cuenta con muy poca aptitud para las matemáticas, lo cual puede reducir incluso la autoeficacia y su elección vocacional (Álvarez, 2017). Situación que debe ser atendida a fin de poder alcanzar en las mujeres la autoeficacia para participar de las carreras de la ciencia, tecnología y matemáticas (Tellhed, Backstrom & Bjorklund, 2016). Lo cual, se puede lograr teniendo en cuenta que en las carreras de ciencia, tecnología, ingeniería y matemáticas (Science, Techonology, Engineering and Mathemátics, STEM), las mujeres a pesar de tener menor nivel de autoeficacia matemática, poseen los mismos logros en matemáticas (Peters, 2013) y habilidades similares a pesar de que el ambiente de estudios no sea del todo equitativo (Riegle-Crumb & Moore, 2013).

Y de acuerdo a lo expuesto respecto a las diferencias entre mujeres y varones a nivel de autoeficacia matemática (Joët, et al., 2011) y las aparentes diferencias de desempeño en el campo de la ingeniería (Peters, 2013), la presente investigación propone como objetivo: identificar las diferencias entre los niveles de autoeficacia matemática en estudiantes mujeres y varones aspirantes a la carrera de ingeniería; y a la vez identificar cómo los niveles de autoeficacia matemática influyen en la aplicación de estrategias de estudio de las matemáticas aplicando los principios de la Educación Matemática Realista (EMR).

3. Metodología

En coherencia al objetivo, el estudio asume el enfoque cuantitativo de investigación no experimental de tipo descriptivo correlacional (Creswell, 2012) a fin de poder establecer la relación entre los niveles de autoeficacia en las matemáticas y las estrategias para poder estudiarlas.

3.1. Muestra

Identificada la población de 577 estudiantes aspirantes a las carreras de ingeniería de una universidad pública de prestigio de Lima – Perú, se aplicó un procedimiento de muestreo por conveniencia y se seleccionaron 122 participantes (quienes eran el 21% de la población y estaban próximos a realizar un examen de admisión a la universidad). Esta muestra estaba constituida por 66 mujeres (*54.1%*) y 56 varones (*45.9%*) entre 15 a 26 años de edad ($Media^{edad} = 17.82; DE: 1.75$).

3.2. Instrumentos

Escala de Autoeficacia Matemática (Reiner, 2017). La presente escala es la versión abreviada de la escala de Scott Reiner que posee por finalidad evaluar los niveles de confianza en las capacidades matemáticas que los estudiantes poseen. Esta escala posee 5 ítems que los organiza de forma unidimensional. Luego de su aplicación, la escala

fue sometida a un análisis de confiabilidad la escala obtuvo un coeficiente del Alfa de Cronbach de 0.85 demostrando confiabilidad (Field, 2009).

Escala de Autopercepción de Aplicación de Estrategias de la Matemática Realista (Iraola y Gonzáles, 2019). Esta escala evalúa la autopercepción que los estudiantes poseen de su propia aplicación de las estrategias de la matemática realista (EMR). Esta escala posee 16 ítems distribuidos en 4 ítems por las siguientes estrategias de aplicación de los principios de la matemática realista (principio de realidad, de reinvención guiada, de interacción e interconexión). En el análisis de confiabilidad la subescala de estrategias de aplicación del principio de realidad obtuvo un coeficiente del Alfa de Cronbach de 0.81, la estrategia del principio de reinvención guiada fue de 0.71, la estrategia del principio de interacción fue de 0.71, y la estrategia del principio de interconexión fue de 0.90, demostrando adecuados niveles de confiabilidad (Field, 2009).

3.3. Procedimientos

Identificada la problemática de los estudiantes, se procedió a realizar el estudio aplicando los instrumentos. Este proceso se llevó a cabo según los procedimientos de ética conservando el anonimato de los participantes (Goodwin, 2010). Y, debido a contar con una muestra que incluye estudiantes de 15 años (menores de edad en Perú) se procedió en conformidad a la Ley Nro, 27337 del Código Civil del Niño y del Adolescente (Congreso de la República del Perú, 2000) que según el Artículo 9 y 10, los considera libres de opinión y expresión. Luego de la aplicación de los instrumentos se elaboró una base de datos con el programa Statistical Package for the Social Sciences (SPSS, versión 23, IBM Knowledge Center, 2016) y se procedió con los análisis preliminares de confiabilidad de las escalas, que obtuvieron un coeficiente de consistencia interna de Alfa de Cronbach mayor o igual a 0.70 (Field, 2009). Finalmente, se procedió a realizar los análisis principales de diferencias de medias con el estadístico T de Student (Ahsnullah, Kibria & Shakil, 2014) y el análisis de correlación de variables (Creswell, 2012).

4. Resultados

Para el análisis de los resultados se procedió a realizar un estudio comparativo de acuerdo con el género para la variable autoeficacia matemática y las variables de la aplicación de los principios de la matemática realista.

4.1. Diferencia de Medias

Se procedió a identificar las diferencias de género según el análisis de diferencias de medias para muestras independientes con el estadístico T de Student (Ahsnullah, et al., 2014). Así, en la Tabla 1 se puede observar que solo la variable autoeficacia matemática reflejó una diferencia de media significativa, indicando que las mujeres presentan menores niveles de autoeficacia matemática que los varones. En cuanto a las variables desempeño académico en matemáticas y al uso de estrategias de aplicación de los principios de la matemática realista al momento de estudiar no se encontraron diferencias significativas. Estos datos se confirman en la Tabla 2.

Variables	Género	N	Media	Desv. Desviación	Desv. Error promedio
Autoeficacia Matemática	Varones	56	5.36	1.122	0.150
	Mujeres	66	4.91	0.857	0.105
Desempeño Académico (Promedio en matemáticas)	Varones	46	14.65	2.243	0.331
	Mujeres	60	14.33	3.139	0.405
Aplicación del Principio de Realidad	Varones	56	5.03	1.127	0.151
	Mujeres	66	5.09	1.085	0.133
Aplicación del Principio de Reinvención Guiada	Varones	56	5.44	0.773	0.103
	Mujeres	66	5.30	0.836	0.103
Aplicación del Principio de Interacción	Varones	56	4.88	1.291	0.172
	Mujeres	66	5.11	1.268	0.156
Aplicación del Principio de Interconexión	Varones	56	5.40	1.141	0.152
	Mujeres	66	5.13	1.173	0.144

Tabla 1 – Estadísticas de Grupo (T de Student)

En la Tabla 2 se observa que el valor de significatividad ($p<0.05$; $p=0.014^{varones}$, y $p=0.016^{mujeres}$) de la media de la variable autoeficacia matemática confirma menores niveles en las mujeres. Pero, en el desempeño académico en matemáticas los varones y las mujeres no evidenciaron un valor de significatividad que indique diferencias relevantes ($p>0.05$).

Variables		Prueba t para la igualdad de medias				
		t	*gl*	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar
Autoeficacia Matemática	Se asumen varianzas iguales	2.499	120	0.014	0.448	0.179
	No se asumen varianzas iguales	2.445	101.816	0.016	0.448	0.183
Desempeño Académico (Promedio en matemáticas)	Se asumen varianzas iguales	0.584	104	0.561	0.319	0.546
	No se asumen varianzas iguales	0.610	103.544	0.543	0.319	0.523
Aplicación del Principio de Realidad	Se asumen varianzas iguales	-0.320	120	0.750	-0.064	0.201
	No se asumen varianzas iguales	-0.319	115.212	0.751	-0.064	0.201
Aplicación del Principio de Reinvención Guiada	Se asumen varianzas iguales	0.916	120	0.362	0.134	0.147
	No se asumen varianzas iguales	0.922	119.079	0.358	0.134	0.146

Variables		Prueba t para la igualdad de medias				
		t	***gl***	**Sig. (bilateral)**	**Diferencia de medias**	**Diferencia de error estándar**
Aplicación del Principio de Interacción	Se asumen varianzas iguales	-0.956	120	0.341	-0.222	0.232
	No se asumen varianzas iguales	-0.955	116.090	0.342	-0.222	0.233
Aplicación del Principio de Interconexión	Se asumen varianzas iguales	1.309	120	0.193	0.276	0.210
	No se asumen varianzas iguales	1.312	117.746	0.192	0.276	0.210

Tabla 2 – Prueba de muestras independientes (T de Student)

4.2. Análisis de Correlaciones entre variables (muestra de mujeres)

Para el análisis de correlación se aislaron las muestras de varones y mujeres para identificar cómo las diferencias en el nivel de la autoeficacia matemática influían de alguna forma en el desempeño académico y en el uso de las estrategias de estudio aplicando los principios de la matemática realista. Y para la interpretación de los coeficientes de correlación se emplearon los criterios de correlación para ciencias sociales de Cohen (1992) mediante el cual se establecen rangos de correlación leves si el coeficiente *r* de Pearson es de 0.10 a 0.23, moderado si es de 0.24 a 0.36, y fuerte si es de 0.37 a más.

Asumiendo estos criterios, en la Tabla 3 se puede observar que en la muestra de mujeres la autoeficacia matemática se relaciona de forma positiva, moderada y significativa con la estrategia de estudio de la EMR de aplicación de la interacción ($r=0.31^{*}; p<0.05$); es decir, que para las mujeres los elevados niveles de confianza en sus capacidades matemáticas se relacionan con mayores posibilidades de interactuar con sus compañeros para estudiar matemáticas. Luego, la autoeficacia matemática se relaciona de forma positiva, fuerte y significativa con la estrategia de estudio de aplicar la interconexión de las matemáticas ($r=0.41^{**}; p<0.01$); lo que significa que, en las mujeres, los adecuados niveles de autoeficacia matemática se asocian a aplicar estrategias interconectando la aritmética, el álgebra, la trigonometría y la geometría. Adicionalmente, se observa que la autoeficacia matemática se relaciona de forma positiva, fuerte y significativa con el desempeño académico en las matemáticas ($r=0.38^{**}; p<0.01$); lo que indica que, si el nivel de confianza en las propias capacidades matemáticas es elevado, las calificaciones obtenidas en las matemáticas serán también elevadas (y viceversa).

Se puede apreciar que la autoeficacia matemática posee tres relaciones significativas y directamente proporcionales con el uso de estrategias de la EMR para estudiar matemáticas aplicando el principio de interacción y de interconexión matemática, así también con el desempeño académico en matemáticas.

Variables		1	2	3	4	5	6
1	*Autoeficacia Matemática*	1					
2	*Aplicación del Principio de Realidad*	-0.03	1				
3	*Aplicación del Principio de Reinvención Guiada*	-0.09	0.24*	1			
4	Aplicación del Principio de Interacción	0.31*	0.25*	0.33**	1		
5	*Aplicación del Principio de Interconexión*	0.41**	0.33**	0.26*	0.45***	1	
6	*Desempeño Académico (Promedio en matemáticas)*	0.38**	0.25	-0.21	0.34**	0.13	1

Tabla 3 – Correlaciones entre variables (Muestra de mujeres)

4.3. Análisis de Correlaciones entre variables (muestra de varones)

En la Tabla 4 se puede observar que en la muestra de varones la autoeficacia matemática se relaciona de forma positiva, fuerte y significativa con la estrategia de estudio de la EMR de aplicación del principio de realidad ($r=0.39^{*}; p<0.05$); es decir, que para los varones los elevados niveles de confianza en sus capacidades matemáticas se relacionan con mayores posibilidades de estudiar las matemáticas vinculándolas a la realidad. Luego, la autoeficacia matemática se relaciona de forma positiva, fuerte y significativa con la estrategia de estudio de aplicar la reinvención guiada ($r=0.59^{***}; p<0.001$); es decir, que para los varones los elevados niveles en su autoeficacia matemática se relacionan con mayores posibilidades de estudiar las matemáticas, reinventando estrategias de solución de problemas. También, se observó que la autoeficacia matemática se relacionó de forma positiva, moderada y significativa con la estrategia de estudio de aplicar la interacción ($r=0.28^{*}; p<0.05$); es decir que, para los varones, los elevados niveles de confianza en sus capacidades matemáticas se relacionan con mayores posibilidades de interactuar con otras personas para estudiar matemáticas. Finalmente, se identificó que la autoeficacia matemática evidenció una relación positiva, fuerte y significativa con la estrategia de estudio de aplicar la interconexión matemática ($r=0.60^{***}; p<0.001$); es decir que, para los varones, los elevados niveles de confianza en sus capacidades matemáticas influyen en sus preferencias por vincular las áreas matemáticas (aritmética, álgebra, geometría y la trigonometría) al momento de estudiar.

Variables		1	2	3	4	5	6
1	*Autoeficacia Matemática*	1					
2	*Aplicación del Principio de Realidad*	0.39**	1				
3	*Aplicación del Principio de Reinvención Guiada*	0.59***	0.11	1			
4	*Aplicación del Principio de Interacción*	0.28*	0.15	0.29*	1		

Variables		1	2	3	4	5	6
5	*Aplicación del Principio de Interconexión*	0.60***	0.46***	0.44**	0.52***	1	
6	*Desempeño Académico (Promedio en matemáticas)*	0.24	-0.11	0.23	-0.02	0.18	1

Tabla 4 – Correlaciones entre variables (Muestra de varones)

Se puede apreciar que, para los varones, la autoeficacia matemática posee cuatro relaciones significativas y directamente proporcionales con el uso de estrategias de la EMR para estudiar matemáticas de forma contextualizada, para reinventar estrategias de resolución de problemas matemáticos, para aplicar la interacción social y la interconexión matemática al momento de estudiar. Pero, a diferencia de las mujeres, en los varones no se establecieron relaciones entre la autoeficacia matemática y el desempeño académico en matemáticas.

5. Discusión

Desde la perspectiva de la Teoría Cognitivo Social de Albert Bandura (1977) la autoeficacia matemática es la confianza que las personas desarrollan en sus propias capacidades de resolución de problemas matemáticos (Bandura, 1997). Esta es una variable psicológica que en algunos estudios empíricos ha presentado diferencias significativas en cuanto al género, demostrando que este tipo de autoeficacia es de menor nivel en las mujeres; a diferencia de los varones quienes poseen mayores niveles de autoeficacia matemática (Joët, et al., 2011; Peters, 2013); siendo una condición que podría influir en la elección vocacional de muchas mujeres, quienes no evidencian mucho interés por estudiar las carreras de ingeniería (Rodríguez, 2016). Y es ante esta situación que la presente investigación pretendió identificar en una muestra de estudiantes aspirantes a las carreras de ingeniería, los niveles de autoeficacia matemática en estudiantes mujeres y varones aspirantes a la carrera de ingeniería; identificando también, cómo los niveles de autoeficacia matemática influyen en la aplicación de estrategias de estudio de la matemática.

Así, al realizar los estudios comparativos evidenciaron que los varones poseen mayores niveles de autoeficacia matemática, a diferencia de las mujeres aspirantes a las carreras de ingeniería. Este hallazgo confirma diversos estudios en muestras de educación básica de Joët (et al., 2011) y con muestras de estudiantes de educación secundaria de Reiner (2017). Pero, específicamente en muestras de estudiantes universitarios, Peters (2013) identificó diferencias significativas hallando menores niveles de autoeficacia matemática. En los análisis comparativos se identificó que no había diferencias significativas entre el nivel de desempeño matemático entre las mujeres y varones. Indicando que a pesar de que los varones aspirantes a las carreras de ingeniería poseen mayores niveles de confianza en sus capacidades matemáticas, sus calificaciones en estas asignaturas son las mismas que las mujeres. Hallazgo que puede discrepar con los estudios de *Riegle-Crumb y Moore (2013*) quienes identificaron en una muestra de doscientos estudiantes de un curso de ingeniería, que las mujeres presentaban menores niveles de autoeficacia

matemática, de interés por estudiar ingeniería, menor confianza en sus habilidades para la ingeniería, mayores niveles de ansiedad ante los exámenes y menor desempeño académico. O los estudios de Usher y Pajares (2009) quienes identificaron que la baja autoeficacia matemática genera ansiedad ante los exámenes y esto reduce el desempeño académico en las mujeres, los resultados presentes no lo confirman.

Finalmente, se identificó que los niveles de autoeficacia matemática en las mujeres se relacionaban con el uso de dos tipos de estrategias de la EMR para estudiar matemáticas; la primera de ellas consiste en aplicar el principio de interacción social para aprender matemática de forma cooperativa; y la segunda estrategia consiste en aplicar la interconexión matemática (vinculando las diversas áreas matemáticas: aritmética, álgebra, trigonometría y geometría). Y en el caso de los varones, su autoeficacia se relacionó con aplicar cuatro estrategias de la EMR: la primera que consistió en contextualizar la matemática a la realidad; la segunda implicaba reinventar los pasos para resolver problemas matemáticos; la tercera en socializar con sus compañeros para aprender matemáticas; y la cuarta estrategia consistió en aplicar la interconexión entre las matemáticas. Además, en las mujeres sus niveles de autoeficacia matemática sí se relacionan con obtener adecuadas calificaciones en matemáticas; mientras que en los varones no se evidenció esta relación. Estos hallazgos permiten confirmar los estudios de Peters (2013) en los que se evidencia quizás que los varones y las mujeres difieren en el nivel de la autoeficacia matemática y en el uso de las estrategias para estudiarlas; sin embargo, al final sus logros son los mismos.

6. Conclusiones

Según lo analizado se concluye que las mujeres poseen menores niveles de autoeficacia matemática a diferencia de los varones, pero ambos poseen los mismos niveles de desempeño académico en las matemáticas. Luego, se aprecia que, en las mujeres sus elevados niveles de autoeficacia matemática influyen en el uso de estrategias de estudio aplicando los principios de la EMR; por ejemplo, ellas relacionan las áreas matemáticas (álgebra, aritmética, geometría y trigonometría) y estudian interactuando con sus pares; lo que favorece su desempeño matemático. Finalmente, en los varones, la autoeficacia matemática se asocia con estrategias de estudio de la EMR de la contextualización de los problemas matemáticos a la realidad, de reinvención de estrategias de solución, de interacción con sus pares y relacionando las áreas matemáticas; pero a diferencia de las mujeres, su autoeficacia no se relacionó a su desempeño matemático. Esto permite reflexionar que, a nivel educativo y laboral, se debe dar a conocer que las mujeres y los varones poseen igual desempeño en las matemáticas; y así, desarrollar estrategias educativas que permitan elevar en ellas su confianza matemática.

Referencias

Ahsnullah, M., Kibria, G. & Shakil, M. (2014). *Normal and Student´s T distributions and their applications*.Miami: Atlantic Press.

Álvarez, M. (2017). *Autoeficacia según el género y su influencia en el ámbito científico-tecnológico* (Trabajo para optar al grado de Máster). Universidad de Salamanca, Facultad de Psicología, Salamanca.

Armando, G. (2016). *Creencias de autoeficacia y bournout en docentes: un enfoque preventivo*. (Tesis Doctoral en Psicología). Departamento de Psicología, Universidad Empresarial Siglo XXI, Córdoba.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 195-215.

Bandura, A. (1997). Self-efficacy: the exercice os control. New York, Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales en. F. Pajares & Urdan, T. (Eds.), Self-efficacy beliefs of adolescents, (pp. 307-337). Greenwich, Information Age Publishing.

Carmona, F. (2015). Educación matemática realista como agente articulador en la enseñanza de la teoría de conjuntos en el contexto social. Universidad de Antioquía, Medellín, Colombia, 2015.

Congreso de la República. (2000). *Código del Niño y del Adolescente. Ley N°. 27337*. Lima. Perú.

Creswell, J. (2012). *Educational Research: Planning, conducting, and evaluating quantitative an qualitative research* (4th Ed.). Upper Saddle River, NJ: Merrill.

Chávez, H., Chávez, J., Ruelas, S. & Gómez, M. (2014). Ansiedad ante los exámenes en los estudiantes del Centro Preuniversitario – UNMSM 2012-I, *Revista de Investigación en Psicología*, *17*(2), 187–201.

Chávez, H., Morocho, J., Alvites, C., Vega, J. & Solis, R. (2017). Relación entre los niveles de depresión y rendimiento académico en estudiantes preuniversitarios, del ciclo ordinario 2015-1 de una universidad nacional de Lima Metropolitana, *Revista de Investigación Psicológica*, *20*(1), 107–118.

Cohen, J. A power primer, *Psychological Bulletin*, *112*(1), 155–159.

Del Río, M., Strasser, K. y Susperreguy, M. (2016). ¿Son las habilidades matemáticas un asunte de género? Los estereotipos de género acerca de las matemáticas en niños y niñas de kínder, sus familias y educadoras. *Calidad en la Educación*, (45), 20-53.

Field, A. (2009). *Discovering statistics using SPSS* (3era Ed.). Lóndres: Sage Publications.

Goodwin, J. (2010). *Research in psychology:methods and design*. Córdova, Brujas.

Honicke, T. & Broadbent, J. (2016). The relation of academic self-efficacy to university student academic performance: A systematic review. *Educational Research Review, 17*, 63-84. http://dx.doi.org/10.1016/j.edurev.2015.11.002

IBM Knowledge Center. (2016). *Ibm.com*. Retrieved from: http://www.ibm.com/support/knowledgecenter/SSLVMB_23.0.0/spss/product_landing.dita

Instituto Nacional de Estadística e Informática (INEI). (2015). *Compendio Estadístico Perú*. Lima: Instituto Nacional de Estadística e Informática.

Iraola, I. & Gonzáles, E. (2019). *Escala de Autopercepción de la Aplicación de Estrategias de la Matemática Realista*. Lima, Perú. (Inédito).

Joët, G., Usher, E. & Bressoux, P. (2011). Sources of self-efficacy: an investigation of elementary school students in France. *Journal of Educational Psychology*, *103*(3), 649-663. doi: 0.1037/a0024048

Katz, I., Eilot, K. &Nevo, N. (2014).“I’ll do it later”: Type of motivation, self-efficacy and homework procrastination.*Motivation and Emotion*, 38, 111 - 119. doi: 10.1007/s11031-9366-1

Lima, P. & Nunes, A. (2017). *A motivação como importante fator desencadeador da aprendizagem da matemática: breve estudo à luz da Teoria Social Cognitiva*. Retrieved from: https://bit.ly/2PHFzbn

McCoach, D., Gable, R., & Madura, J. (2013). Instrument Development in the Affective Domain. New York, NY: Springer.

Olivas, A. & Barraza, A. (2016). *Expectativas de autoeficacia académica en alumnos de un bachillerato técnico de la ciudad de Durango* (1ra Ed.). México, Universidad Pedagógica de Durango.

Pajares, F. & Urdan, T. (2006). *Self-Efficacy beliefs of adolescents*. Florida, Frank Pajares and Tim Urdan.

Peters, M. (2013). Examining the relationships among classroom climate, selfefficacy, and achievement in undergraduate mathematics: A multi-level analysis. *International Journal of Science and Mathematics Education*, *11*(2), 459-480, 2013. doi:10.1007/s10763-012-9347-y

Plaza, L.F. (2016). Modelación matemática en ingeniería. *IE Revista de investigación educativa de la REDIECH*, *7*(13), 47-57.

Reiner, S. (2017). Students´mathematics self-efficacy, anxiety, and course level at a Community College, Ph.D. disertation, Dept. Education., Walden Univ., Washington.

Riegle-Crumb, C. & Moore, Ch. (2013). Examining gender inequality in a high school engineering course, American Journal of Engineering Education, *4*(1), 55–66.

Rodríguez, M. (2014). *Cómo orientar hacia la construcción del proyecto profesional: Autonomía individual, sistema de valores e identidad laboral de los jóvenes*. Bilbao, España: Desclée de Brouwer.

Rodríguez, O. (2016). *Why women do not want to be engineers? Case: Female students of Industrial Engineering in UPTC,* (Master Thesis) Universidad Politécnica de Cartagena, España.

Roik, J. & Ringeisen, T. (2017). Self-efficacy, test anxiety, and academic success: A longitudinal validation. *International Journal of Educational Research, 83*, 84-93. https://doi.org/10.1016/j.ijer.2016.12.006

Ruíz, J. M., & Santana, L. E. (2018). Elección de carrera y género. *Revista Electrónica de Investigación y Docencia (REID)*, (19), 7-20. doi.org/10.17561/reid.v0i19.3470

Schunk, D., y Meece, J. (2006). Self-efficacy development in adolescence. En F. Pajares y R. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 71-96). s. l.: Information Age Publishing.

Schunk, D., Meece, J. & Pintrich, R. (2014). *Motivation in education: Theory, research, and applications*. New Jersey: Pearson.

Tárraga, R. (2008). *¡Resuélvelo! Eficacia de un entrenamiento en estrategias cognitivas y metacognitivas de solución de problemas matemáticos en estudiantes con dificultades de aprendizaje*. (Tesis Doctoral). Departamento de Psicología Evolutiva y de la Educación, Universidad de Valencia, Valencia.

Tellhed, U., Backstrom, M., & Bjorklund, M. (2016). Will I Fit in and Do Well? The Importance of Social Belongingness and Self-Efficacy for Explaining Gender Differences in Interest in STEM and HEED Majors. *Sex Roles*; (77), 86–96. doi 10.1007/s11199-016-0694-y

Thenmozhi, C. (2018). Vocational Guidance and its Strategies. *International Journal of Education*, 7(1), 20-23.

Usher, E. & Pajares, F. (2009). *Sources of self-efficacy in mathematics: Avalidation study," Contemporary Educational Psychology*, 34, 89–101.

Yuanita, P., Zulnaidi, H. & Zakaria, E. (2018). The effectiveness of Realistic Mathematics Education approach: The role of mathematical representation as mediator between mathematical belief and problem solving, *PLoS ONE*, *13*(9), 1-20.

Zolkower, B., Bressan, B. & Gallego, F. (2006). La corriente realista de didáctica de la matemática. Experiencias de un grupo de docentes capacitadores. Yupana, 3, 11–33.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 14/09/2020
Aceitação/Acceptance: 24/11/2020

Evaluación de la precisión de los modelos de clasificación binaria para la identificación de noticias verdaderas o falsas en Costa Rica

Eilyn Salazar Miranda[1], Gustavo Tenorio Arce[2], Luis Naranjo Zeledón[3]

esalazarm@ucenfotec.ac.cr; gtenorioa@ucenfotec.ac.cr; lnaranjo@ucenfotec.ac.cr

[1] Eilyn Salazar-Miranda, San Pablo, 40901, Heredia, Costa Rica.

[2] Gustavo Tenorio-Arce, Cumming, GA 30040, Estados Unidos.

[3] Luis Naranjo-Zeledón, San José, 10101, Costa Rica.

Pages: 156-170

Resumen: Este artículo describe el proceso investigativo para la creación de un primer conjunto de datos beta con noticias difundidas en Costa Rica y la evaluación del nivel de precisión de modelos de minería de datos para la detección de noticias como verdaderas o potencialmente falsas. Para tal efecto se realiza un análisis lingüístico del contenido de las noticias utilizando la herramienta LIWC para la extracción de características cuantitativas y se ofrecen seis variaciones para modelos como Regresión Logística, Árboles de Decisión, Bosques Aleatorios, Máquinas de Soporte de Vectores y Redes Neuronales, haciendo una comparación de las variables lingüísticas y analizar los resultados obtenidos en los modelos los cuales resultan satisfactorios para la escala de evaluación propuesta.

Palabras-clave: Noticias falsas; conjunto de datos beta; fake news detection; minería de datos; news dataset, español.

Evaluation of the precision of the binary classification models for the identification of true or false news in Costa Rica

Abstract: This article describes the investigative process for creating a first beta dataset with news broadcast in Costa Rica and evaluating the level of accuracy of data mining models for news detection as true or potentially false. For this purpose, a linguistic analysis of the news content is carried out using the LIWC tool to extract quantitative characteristics and six variations are offered for models such as Logistic Regression, Decision Trees, Random Forests, Vector Support Machines and Neural Networks, making a comparison of the linguistic variables and analyzing the results obtained in the models which are satisfactory for the proposed evaluation scale.

Keywords: Fake news; beta dataset; fake news detection; data mining; news dataset, spanish.

1. Introducción

En los últimos años, la difusión de noticias falsas en redes sociales y sitios web de noticias, ha cobrado relevancia y despertado interés de diversos investigadores a nivel mundial como (Edell, 2018), (Shu, Sliva, Wang, Tang, & Liu, 2017) y (Granskogen, 2018) quienes concuerdan en señalar que las noticias falsas representan un problema al distorsionar la opinión, emociones y acciones de los diversos lectores expuestos a su contenido y pueden traer repercusiones de gran impacto a nivel social y económico.

Costa Rica no se ha visto exenta de este fenómeno y con el fin de combatir este problema han surgido proyectos de tipo periodístico para la identificación de noticias falsas, como Doble Check, financiado por la Vicerrectoría de Docencia de la Universidad de Costa Rica (doblecheck.cr, 2018), #NoComaCuento promovido por el periódico La Nación (Grupo Nación, s.f.), así como otras iniciativas promovidas por el Gobierno de la República de Costa Rica y las entidades competentes quienes han brindado atención a este tema anunciando acciones al respecto (presidencia.go.cr, 2018). Sin embargo, la dificultad de este tipo de iniciativas surge cuando el número de reportes de contenido falso aumenta y el recurso humano disponible es insuficiente para realizar la validación de los hechos que permiten identificar el contenido potencialmente falso del verdadero.

Por esta razón y considerando que la clasificación de noticias como verdaderas o potencialmente falsas usando modelos de minería de datos es un tema poco explorado en el ámbito costarricense; el presente artículo busca describir el proceso llevado a cabo para la creación de un conjunto de datos beta, así como la aplicación de técnicas de aprendizaje automático a este conjunto de datos, y presentar los resultados obtenidos.

En la sección 2 se exploran las redes sociales más utilizadas por los costarricenses para el consumo de noticias y cuáles de estas redes fueron de mayor interés para la investigación. En la sección 3 se describen los elementos que investigadores han explorado en este tema (Contenido de la noticia, Contexto social, e Información temporal) para la generación de las variables cualitativas y cuantitativas a incluir en el conjunto de datos beta.

La sección 4 incluye la metodología de la investigación, describiendo los pasos seguidos para la extracción del contenido de sitios web y redes sociales, el proceso de limpieza de los datos e identificación de variables usando LIWC para crear finalmente un conjunto de variables cualitativas y cuantitativas que podían ser usadas por los modelos de minería, y el proceso seguido para la aplicación de los modelos de aprendizaje automático que nos permitirían comparar las precisiones obtenidas.

Se debe destacar que, al momento de iniciar la investigación en el año 2018, tras realizar las consultas bibliográficas respectivas, no se encontraron conjuntos de datos en español, razón por la cual parte de los objetivos de esta investigación incluye la creación de un conjunto de datos beta con noticias costarricenses. En años posteriores otros investigadores han abordado la creación de nuevos conjuntos de datos con noticias de países como México (Posadas-Durána, Gómez-Adornob, Sidorovc, & Escobara, 2019), (Acosta, 2019)y España (Abonizio, Morais, Tavares, & Junior, 2020).

2. Contexto del uso de redes sociales en Costa Rica

Entender cuáles redes sociales se utilizan mayoritariamente en Costa Rica para leer y compartir noticias fue necesario para determinar las que serían de interés para el análisis. De acuerdo con el Informe 2018 elaborado por Latinobarómetro (Corporación Latinobarómetro, 2018) WhatsApp, Facebook y YouTube son las redes sociales más utilizadas en Latinoamérica y Costa Rica es el líder en donde el uso de WhatsApp alcanza un 83%, Facebook un 77% y YouTube 59%.

Estos mismos datos son sustentados por el informe de resultados del estudio de opinión sociopolítica "Audiencias Noticiosas De Medios Digitales: Redes Sociales, WhatsApp Y Noticias Falsas" de la Universidad de Costa Rica (Brenes, Pérez, Siles, Camacho, & Solís, 2019), el cual analiza elementos como las características de la población que hace uso de estas redes sociales, medios periodísticos usados en Costa Rica y la percepción de la población en torno a la difusión de noticias falsas, y elementos que son analizados por los costarricense para determinar la veracidad de una noticia.

Por medio de una colaboración con la iniciativa Doble Check (Chinchilla-Ugalde, 2019), también se nos señala que Facebook y WhatsApp son las redes sociales en donde reciben más reportes de contenido potencialmente falso. De estas Facebook la consideramos como una red social de interés para llevar a cabo la investigación por la posibilidad de extraer datos, mientras que WhatsApp tiene una mayor complejidad para su proceso de recolección por lo que no se tomó en cuenta. Otras redes sociales como Twitter no son tan utilizadas por los costarricenses para compartir contenido según Doble Check. Tampoco se encontró evidencia de conjuntos de datos con noticias extraídas de redes sociales o medios de noticias en Costa Rica u otra información asociada a la noticia como información del post o personas involucradas, entre otros elementos.

Con el objetivo de comprender cuáles son las características de una noticia y los atributos que pueden ser significativos al clasificar contenido falso, a continuación, se presenta la caracterización de noticias falsas en redes sociales.

3. Caracterización de las noticias falsas en redes sociales

A diferencia de una noticia publicada en prensa, las noticias difundidas a través de redes sociales pueden ser analizadas tomando en cuenta no solo el contenido del evento descrito

Elementos	Descripción
Contenido de la noticia	El contenido de la noticia incluye los aspectos lingüísticos y visuales de la noticia. Tomando en consideración los elementos representativos de una noticia identificados por (Shu, Sliva, Wang, Tang, & Liu, 2017), se consideró la extracción de los siguientes elementos: Fuente, Título, Texto del cuerpo e Imagen / Video.
Contexto social	Dentro del contexto social, (Shu, Mahudeswaran, Wang, Lee, & Liu, 2019) identifican los aspectos relacionados con el post y las relaciones de los usuarios a través de la red.
Información espacio temporal	Involucra toda la información asociada con el lugar y tiempo en el que los diferentes usuarios involucrados interactuaron con el contenido compartido

Tabla 1 – Elementos en el contenido de una noticia

por la noticia, sino que es posible recopilar elementos como la información espacio temporal de la noticia y su contexto social. Diversos autores como (Ruchansky, Seo, & Liu, 2017), (Zhou & Zafarani, 2018), (Shu, Mahudeswaran, Wang, Lee, & Liu, 2019) se han dado a la tarea de describir estos elementos e identificar patrones que pueden ser de utilidad al momento de identificar noticias falsas empleando cada uno de estos factores. Se presenta un detalle sobre la caracterización de cada uno de estos elementos propuestos por (Shu, Mahudeswaran, Wang, Lee, & Liu, 2019), en la tabla 1.

4. Metodología de la investigación

En nuestro caso debido a que al momento de realizar la investigación no se encontraron conjuntos de datos con noticias costarricenses, el primer reto consistió en identificar cuáles eran los requisitos necesarios para crear dicho conjunto de datos, desde entender cuáles eran las fuentes de datos de las cuales podíamos extraer la información, y los mecanismos para la extracción, hasta que criterios íbamos a emplear para etiquetar el contenido como potencialmente falso o verdadero.

El primer paso del proceso de clasificación de noticias consistió en entender las diferentes definiciones y tipos existentes asociadas a este término, esto basado en la revisión bibliográfica de estudios similares en el área como (Shu, Mahudeswaran, Wang, Lee, & Liu, 2019) y (Pérez & Gardey, 2016). Como resultado de este proceso se identifican las definiciones que se detallan en la Tabla 2.

Tipo de Noticia	Referencia	Definición
Noticias satíricas	(Lubeck, 2009)	La sátira es un género literario, de acuerdo con (Lubeck, 2009), emplea el humor al hacer comentarios sobre personas o actividades y sus vicios, defectos o errores percibidos. “En el periodismo, la sátira comúnmente se burla de las noticias o utiliza la parodia representada como una noticia convencional. Si bien las noticias satíricas se definen por su naturaleza cómica, el uso del humor inexpresivo para crear lo que se denomina “noticias falsas”, su objetivo subyacente es hacer declaraciones sobre personas reales, eventos y tendencias, a menudo con la intención de influir en el cambio”. (p.2).
Rumores	(Pérez & Gardey, 2016) (Shu, Sliva, Wang, Tang, & Liu, Fake News Detection on Social Media: A Data Mining Perspective, 2017)	Según (Pérez & Gardey, 2016) se define este concepto como información cuya veracidad está en duda o no puede corroborarse. En relación con este tema, (Shu, Sliva, Wang, Tang, & Liu, 2017) la tarea más relacionada con la detección de noticias falsas es la clasificación de veracidad del rumor. La clasificación de veracidad de los rumores se basa en gran medida en las otras subtareas, lo que requiere que las posturas u opiniones se puedan extraer de las publicaciones relevantes.

Tipo de Noticia	Referencia	Definición
Teorías de conspiración	(Merriam-Webster, Conspiracy theory, s.f.)	El (Merriam-Webster, Conspiracy theory, s.f.) define la teoría de conspiración como "Una teoría que explica un evento o conjunto de circunstancias como resultado de un complot secreto por conspiradores usualmente poderosos".
Desinformación (disinformation)	(Merriam-Webster, "Disinformation." Merriam-Webster.com Dictionary, s.f.)	El (Merriam-Webster, "Disinformation." Merriam-Webster.com Dictionary, s.f.) define la desinformación como "Información falsa difundida deliberadamente y a menudo encubierta (como un rumor) con el fin de influenciar la opinión pública y obstaculizar la verdad".
Falta de información (misinformation):	(Sullivan, 2018)	La información falsa que se difunde, independientemente de si hay intención de engañar. De acuerdo con (Sullivan, 2018) "Desinformación (disinformation)" implica una intención definida de engañar, mientras que "Falta de información (misinformation)" no implica ninguna implicación sobre la intención.
Clickbait	(Merriam-Webster, Clickbait, s.f.) (Shewan, 2017)	El (Merriam-Webster, Clickbait, s.f.) define el Clickbait como "Algo (como un titular) diseñado para que los lectores quieran hacer clic en un hipervínculo, especialmente cuando el enlace lleva a contenido de dudoso valor o interés". (Shewan, 2017) Identifica las siguientes características: Un titular llamativo y convincente, Imágenes divertidas o memorables / video, Tono humorístico, o hace un llamamiento fuerte a una emoción específica e Intención de fomentar el intercambio social.

Tabla 2 – Tipos de noticias falsas

El segundo paso llevado a cabo fue comprender qué atributos del contenido textual de la noticia son relevantes para la identificación del contenido potencialmente falso. Tras realizar una revisión bibliográfica fue posible identificar qué elementos eran mencionados por múltiples autores. La tabla 3 resume algunos de estos atributos para la clasificación de contenido considerados en la investigación:

Elemento	Referencias
Signos de puntuación y cantidad de signos.	Los caracteres de puntuación como punto, coma, guiones, signos de interrogación y de exclamación también son características lingüísticas que pueden ser efectivas para categorizar el texto engañoso del texto verdadero. (Anoop, Manjary, Deepak, & Lajish, 2018), han identificado que noticias de tipo satírico tienen mayor cantidad de signos de puntuación.

Elemento	Referencias
Cantidad de adverbios. Cantidad de verbos, métricas de legibilidad.	(Perez-Rosas, Kleinberg, Lefevre, & Mihalcea, 2017) mencionan que los autores de noticias falsas usan más adverbios, verbos y signos de puntuación que los autores de noticias legítimas. La comprensibilidad o legibilidad del texto es otra característica más para la comprensión falsa o engañosa. En general, esto incluye características como número de caracteres, palabras complejas, palabras largas, número de sílabas, tipos de palabras, número de párrafos y varias otras métricas de legibilidad.
Correlación semántica	Los artículos reales comprenden los pares correlacionados de palabras de contenido y oraciones que se correlacionan entre sí y se dice que esta correlación es semánticamente coherente (Badaskar, Agarwal, & Arora, 2008).
Repetición del titular en el contenido de la noticia	(Pennebaker, Boyd, Jordan, & Blackburn, 2015) indica que los titulares resultan relevantes para detectar la sátira. Mientras que los artículos de noticias legítimos reportan material nuevo en la primera línea, los artículos satíricos tienden a repetir material del título. Una diferencia sintáctica notable entre los artículos satíricos y legítimos fue la longitud de la oración y la complejidad: los artículos satíricos tienden a agrupar un mayor número de cláusulas en una oración.
Orientación semántica positiva o negativa y tiempo verbal	La presencia de polaridad en la sátira se ha observado en métodos anteriores; tal y como se indica por orientación semántica positiva o negativa, emotividad y palabras emocionales (Perez-Rosas, Kleinberg, Lefevre, & Mihalcea, 2017). El contenido legítimo incluye más palabras funcionales (por ejemplo, pronombres como él, ella), negaciones y expresiones de relatividad. Por otra parte, el lenguaje utilizado al informar contenido falso usa más palabras sociales y positivas, expresa más certeza y se enfoca en acciones presentes y futuras.
Autor	En muchas ocasiones, las noticias falsas son generadas y compartidas por sitios desconocidos y carecen de información del autor
Cantidad de citas o referencias	Otro de los aspectos que resaltan diversos autores, es el hecho que una noticia verdadera, por lo general está acompañada de referencias de las fuentes que confirman la información presentada, o URLs, links o detalle de la fuente, que permiten la validación de los hechos que expresan.
Uso de infinitivos en el titular.	De acuerdo con (Chinchilla-Ugalde, 2019) en noticias de Costa Rica, se ha detectado una relación entre el titular de la noticia haciendo uso de infinitivos y su posterior clasificación como noticia falsa.

Tabla 3 – Atributos para clasificación de contenido

4.1. Dependencias para la creación del conjunto de datos

Debido a la ausencia de conjuntos de datos con noticias en el contexto costarricense fue necesario la utilización de diferentes instrumentos para la recolección de estos yla consecuente creación de un conjunto de datos beta. Se aplicaron encuestas dirigidas a profesionales en periodismo y filosofía, para entender los hábitos de consumo de noticias de los costarricenses y cómo perciben las personas la información que se les hace llegar. En Costa Rica se no cuenta con sitios de verificación de hechos realizados por expertos (como FactCheck por ejemplo) y los procesos de verificación de medios como Doble Check tampoco hacen una diferenciación de todo el contenido. Por lo tanto, se agrupa cualquier noticia satírica, rumores, clickbaits o falta de información dentro de una misma categoría como potencialmente noticia falsa, por la complejidad y amplitud que tiene este término.

Identificación de las fuentes de datos

Para la selección de las fuentes de datos se consideraron elementos como la información recopilada a través de las encuestas a expertos, así como el uso de APIs de dominio público como: "Facebook for Developers" (Facebook , 2018), Twitter for Developers (Twitter , 2018) y Webhose.io (Webhose.io, 2018) que permitieran extraer contenido de las noticias generadas en Costa Rica.

Estrategias de selección de noticias "verdaderas" y "potencialmente falsas"

Autores como (Zhou & Zafarani, 2018) resaltan la importancia de la figura del editor o emisor de la noticia como un elemento que puede ser empleado al momento de identificar una noticia falsa. Partiendo de esto, se consideran como noticias verdaderas aquellas que provienen de sitios periodísticos que gozan de prestigio en el medio periodístico local y que se rigen por las normativas del Colegio de Periodistas de Costa Rica y cumplen su función informativa de manera ética y profesional.

A través de la colaboración con Doble Check se nos facilitó un listado de 113 perfiles públicos de Costa Rica en la red social Facebook, relacionados previamente con la difusión de contenido falso y que han sido objeto de investigación por esta iniciativa lo cual facilitó la extracción de posible contenido falso mediante la utilización de herramientas adicionales que se explicarán más adelante.

4.2. Creación del conjunto de datos beta

Para la creación del conjunto de datos se siguieron dos estrategias diferentes, la primera para la extracción de datos de medios periodísticos reconocidos y la segunda para la extracción de datos de perfiles públicos de redes sociales con noticias potencialmente falsas, y se estableció un proceso de 6 pasos descrito en la tabla 4.

Pasos	Descripción del proceso
Extracción de datos de noticias del medio costarricense	La recolección de los datos iniciales se hizo utilizando el sitio web Webhose.io en donde se hace un filtro por país y por rango de tiempo de máximo de 30 días. Igualmente se seleccionan los tipos disponibles de sitios web para extraer contenido (blogs, discusión, sitios de noticias, reseñas) en lenguaje español. Para hacer el proceso más ágil se hizo un script para conectarse al API de este sitio web y obtener mayor cantidad de resultados y enviarlos a una hoja de cálculo.
Extracción de noticias en redes sociales	Teniendo el listado de perfiles de Facebook conocidos por esparcir contenido falso en Costa Rica se utilizó una herramienta de WebScraping para obtener los datos de las publicaciones. Esto se hace de esta manera por la imposibilidad de utilizar el API público de Facebook actualmente. A esta herramienta se le provee una lista de direcciones web de perfiles de Facebook a los cuales luego se pueden obtener ciertos atributos de la publicación considerados importantes para el análisis. El atributo más importante es el de la URL de la noticia que se está compartiendo por medio del perfil ya que provee las fuentes de la información potencialmente falsa. Seguido a este proceso se hace uso nuevamente de un script para conectar a estos sitios por medio de un API y obtener los atributos de la noticia.

Pasos	Descripción del proceso
Detección de características psicolingüísticas	Tomando en cuenta los diferentes atributos que se logran extraer de las noticias en el paso anterior, se seleccionan los campos "Título de la noticia" y "Texto de la noticia" para su procesamiento en la herramienta de análisis de texto. De acuerdo con (Brenes, Pérez, Siles, Camacho & Solís, 2019) elementos como el contenido de la noticia, así como el titular son dos de los aspectos a los que los costarricenses le prestan mayor atención. Otros atributos como el URL, la URL de la fuente, medio periodístico y tipo de sitio fueron conservados en el conjunto de datos, pero no tomados para el análisis a fin de evitar sesgo al usar atributos como el medio periodístico. Para la creación del conjunto de datos beta se hace uso del software "Linguistic Inquiry and Word Count" o LIWC, el cual es un programa de computadora que analiza textos y es capaz de calcular cómo las personas usan diferentes categorías de palabras a través de una gran diversidad de textos. La herramienta ofrece soporte en diferentes idiomas, siendo el idioma inglés el que agrupa mayor cantidad de palabras y el más actualizado. Se contactó a los creadores del diccionario LIWC2007 en español a fin de verificar si existía una versión más reciente y con características y variables compuestas similares a la versión en inglés LIWC2015. Al no existir uno más reciente, la recomendación fue hacer la traducción de los textos al idioma inglés.
Aplicación de tareas de pre-procesamiento de datos.	Sobre el conjunto de datos disponible se aplicaron una serie de tareas de pre-procesamiento, que se describen a continuación: 1. Renombrar las variables del conjunto de datos beta. 2. Adición de variables "Título Inglés" y "Texto Inglés", dichas variables corresponden a las traducciones a este idioma de las variables "Título" y "Texto" respectivamente, para su posterior análisis con el diccionario en inglés LIWC2015. 3. Estandarización del texto a minúsculas. 4. Eliminación del ruido de las noticias. En el conjunto de datos beta, se detectó que algunas noticias incluían al inicio del texto, expresiones que incluían la ciudad, así como la fecha de la noticia. 5. Lematización del contenido. Procesar el conjunto de datos en LIWC con las palabras lematizadas ofrece mejores resultados en términos de cantidad de palabras y clasificaciones realizadas.
Clasificación de los medios informativos como verdaderos o potencialmente falso.	Se procedió a emplear el conjunto de datos de salida, como resultado del pre procesamiento de los datos descritos anteriormente y se añadió una columna adicional al conjunto de datos llamada "Clasificación", la cual puede tomar los valores 1 y 0, en donde 1 representan las noticias que se consideran "verdaderas" o fueron difundidas por sitios reconocidos por los costarricenses y que tienen una trayectoria reconocida; mientras que 0 representa aquellos sitios "potencialmente falsos", que no son tan conocidos, o que representan la opinión de determinados sectores o gremios y han sido vinculados con la difusión de noticias falsas.
Análisis de texto y creación de conjunto de datos beta	Como parte de este proceso de identificar las variables psicolingüísticas del texto, se analizó la columna "textoPreprocesado" tanto para la creación del conjunto de datos beta en español como en inglés. Como resultado de este proceso, se obtuvieron en total 80 variables para el conjunto de datos beta en inglés y 69 variables en el conjunto de datos beta en español. Los conjuntos de datos finales tanto en español como en inglés, así como un conjunto de datos sin ningún tipo de procesamiento, han sido puestos a disposición del público y pueden ser descargados del siguiente enlace: https://zenodo.org/record/3519855#.XbjyS9V7nIV

Tabla 4 – Proceso para la creación del conjunto de datos beta

4.3 Aplicación de los modelos de minería

El proceso de aplicación de modelos de minería de datos está dividido en cuatro etapas que son Comprensión, Preparación, Modelado y Evaluación de los datos. Para efectos de

reproducibilidad del trabajo se facilitan los scripts y archivos de datos utilizados y está accesibles en el siguiente enlace: https://github.com/gtenorioa/fakenewscr.git

Comprensión de los datos

Se analiza el conjunto de datos creado en el idioma español y empleando los mismos textos de la noticia se procedió a crear un conjunto de datos en inglés utilizando la herramienta Translate de Google Sheets (Google, 2019), a fin de realizar la comparación de resultados. Asimismo, se quiere determinar qué características contribuyen de mejor manera en el problema de clasificación que se pretende resolver.

Ambos conjuntos de datos usan las mismas observaciones, varían únicamente el idioma en que se encuentra el texto, así como las características extraídas en cada idioma. Para la generación de las variables cuantitativas se usó la herramienta LIWC con los diccionarios LIWC2007 en español y LIWC2015 en inglés, cómo se mencionó en el paso 3 del proceso para la creación del conjunto de datos beta, descrito en la tabla 4. La diferencia en la cantidad de variables cuantitativas generadas se debió justamente la diferencia en dichos diccionarios.

En total se cuenta con 318 observaciones de noticias "potencialmente falsas" las cuales tienen un valor de 0 y 286 observaciones "verdaderas" de sitios oficiales con un valor de 1. La preparación de los datos, modelado y evaluación se hizo en el lenguaje de programación R.

El conjunto de datos beta en español cuenta con 69 variables: 5 variables cualitativas, más 64 variables cuantitativas continuas, agrupadas en 10 categorías (Variables resumen, dimensión lingüística, procesos sociales, procesos psicológicos, procesos cognitivos, procesos de percepción, relatividad, preocupaciones personales, lenguaje informal y signos de puntuación).

El conjunto de datos beta en inglés cuenta con 80 variables: 5 variables cualitativas, más 75 variables cuantitativas continuas, agrupadas en 12 categorías (Variables resumen, Dimensión Lingüística, Procesos Psicológicos, Procesos Sociales, Procesos cognitivos, Procesos de percepción, Dirección, Orientación en el tiempo, Relatividad, Preocupaciones Personales, Lenguaje informal, Signos de Puntuación).

Preparación de los datos

Tareas iniciales de preparación incluyeron la creación de un conjunto de entrenamiento y uno de prueba con 70% y 30% de los datos respectivamente. Como parte del proceso de selección de variables a utilizar en los modelos se agruparon las variables basados en su nivel de prioridad: 1=Alta, 2=Media y 3 Baja, como se muestra en la Tabla 5.

Categorías LIWC2015 - inglés	**Categorías LIWC2007 – español**	**Prioridad**
Variables resumen	Variables resumen	1
Dimensión Lingüística	Dimensión Lingüística	1
Signos de Puntuación	Signos de Puntuación	1
Procesos Psicológicos	Procesos Psicológicos	2

Categorías LIWC2015 - inglés	Categorías LIWC2007 – español	Prioridad
Preocupaciones Personales	Preocupaciones Personales	2
Lenguaje informal	Lenguaje informal	2
Procesos Sociales	Procesos Sociales	3
Procesos cognitivos	Procesos cognitivos	3
Procesos de percepción	Procesos de percepción	3
Relatividad	Relatividad	3

Tabla 5 – Categorías de priorización de variables.

Modelado de los datos

Se llevaron a cabo diferentes experimentos con las variables priorizadas mostradas en la tabla 6, a fin de crear modelos que aportaran el mayor nivel de significancia estadística al ser comparado con otras variables. Debido a que los conjuntos de datos en inglés y español no necesariamente contienen las mismas variables (por la diferencia de diccionarios), se trabajó únicamente con variables que estuvieran en ambos conjuntos de datos. En total se crearon 6 modelos distintos por cada idioma empleando combinaciones de variables basados en su prioridad y los factores para la identificación de noticias falsas como preposiciones, signos de exclamación y palabras que expresan lenguaje informal, entre otras.

Modelo	Variables utilizadas
Modelo 1	Cantidad de palabras, cantidad de palabras por oración, palabras mayores de seis letras, pronombres, artículos, adverbios, verbos, signos de pregunta, signos de exclamación, apóstrofes, verbos auxiliares, palabras funcionales y palabras que denotan lenguaje informal.
Modelo 2	Adverbios, palabras mayores de seis letras, palabras que denotan emociones negativas.
Modelo 3	Signos de pregunta, palabras que denotan lenguaje informal y (Ramírez-Esparza, Pennebaker, & García, 2007) preposiciones.
Modelo 4	Todas las variables
Modelo 5	Cantidad de palabras, palabras que denotan confianza o influencia en el estado social, palabras que denotan pensamiento analítico, palabras que denotan tono emocional positivo o negativo, cantidad de palabras por oración, palabras mayores a seis letras, palabras del diccionario utilizadas y palabras funcionales.
Modelo 6	Cantidad de palabras, palabras que denotan confianza o influencia en el estado social, palabras que denotan pensamiento analítico, palabras que denotan tono emocional positivo o negativo, cantidad de palabras por oración, palabras mayores a seis letras, palabras del diccionario utilizadas, palabras funcionales, pronombre, pronombres personales, pronombres impropios, artículos, preposiciones, adverbios, conjunciones, palabras que denotan emociones negativas, verbos, adjetivos, comparaciones, signos de interrogación y números, cuantificadores, todos los signos de puntuación, punto, dos puntos, punto y coma, signos de exclamación, guiones, apóstrofes y paréntesis.

Tabla 6 – Modelos de Datos propuestos.

Se debe señalar que todas estas variables son variables cuantitativas y fueron generadas con la herramienta LIWC basados en los diccionarios utilizados. Para más detalle del significado de las variables del diccionario en español consultar (Ramírez-Esparza, Pennebaker, & García, 2007) y para el diccionario en inglés consultar (Pennebaker, Boyd, Jordan, & Blackburn, 2015).

Para cada uno de estos modelos descritos en la tabla 6 se aplicaron 5 técnicas de minería de datos: Regresión Logística, Árboles de decisión, Bosques Aleatorios, Máquina de Soporte de Vectores y Redes Neuronales y se usó un proceso de validación cruzada (cross folding) de tres iteraciones y 10 pliegues de re-muestreo (resampling), usando la función trainControl() del lenguaje de programación R para el entrenamiento de los modelos. Esta función es parte de la librería Caret (Kunh, 2019).

Evaluación de los resultados

Tras evaluar el valor de la precisión obtenidos en cada uno de los modelos en ambos conjuntos de datos, en general los modelos en español presentan mejores resultados como se puede apreciar en la tabla 7, teniendo que el algoritmo de árboles de decisión obtuvo una precisión de 63,44% al clasificar una noticia como verdadera o potencialmente falsa. Según vemos en la tabla 8, para los modelos en inglés el algoritmo de regresión logística fue el que mostró mejores resultados con una precisión de 62.77%.

Algoritmo	Precisión	M1	M2	M3	M44	M5	M6
Regresión logística	Precisión	54,84%	54,84%	54,84%	62,37%	61,22%	54,84%
Máquinas de SP	Precisión	49,54%	49,54%	54,39%	61,11%	57,52%	56,88%
Árboles de decisión	Precisión	55,43%	55,43%	51,40%	63,44%	59,14%	57,48%
Bosques Aleatorios	Precisión	55,14%	55,14%	55,65%	57,14%	62,75%	55,65%
Redes Neuronales	Precisión	60.22%	60.22%	50.28%	60.22%	59.67%	57.46%

Tabla 7 – Resumen resultados de los modelos del conjunto de datos de español.

Algoritmo	Precisión	M1	M2	M3	M44	M5	M6
Regresión logística	Precisión	57,80%	60,58%	55,91%	62,77%	54,47%	54,74%
Máquinas de SP	Precisión	52,94%	52,10%	50,00%	58,33%	53,64%	54,87%
Árboles de decisión	Precisión	54,01%	56,63%	53,46%	52,59%	53,79%	53,79%
Bosques Aleatorios	Precisión	60,00%	57,41%	55,77%	56,36%	57,27%	55,66%
Redes Neuronales	Precisión	59.12%	55.25%	54.14%	61.33%	59.12%	60.22%

Tabla 8 – Resumen resultados de los modelos del conjunto de datos de inglés.

Otro de los aspectos a destacar es el hecho de que el modelo 4, el que contiene todas las variables, fue el que mostró mejores resultados en ambos conjuntos de datos, en cuanto a precisión se refiere. Finalmente, basado en el criterio de evaluación descrito en la Tabla 9, actualmente los resultados obtenidos se consideran satisfactorios, ya que el modelo que obtuvo la precisión más alta es de 63.77%.

La rúbrica de evaluación fue propuesta por los autores de ésta investigación, teniendo en cuenta los resultados de la precisión obtenida por estudios similares en la detección de noticias falsas, empleando conjuntos de datos de noticias con conjuntos de datos de menos de 5000 observaciones, por ejemplo (Horne & Adalı, 2017) y (Posadas-Durána, Gómez-Adornob, Sidorovc, & Escobara, 2019). Se debe destacar que, en ambos trabajos, se hace uso de algoritmos de aprendizaje automático como Máquinas de soporte vectorial, Regresión logística, bosques aleatorios entre otros y se obtienen precisiones que varían entre el 51 y 77% de precisión. Considerando que las precisiones obtenidas empleando el conjunto de datos beta con noticias costarricenses se mantiene en dichos rangos, se considera el experimento como satisfactorio.

Evaluación	Criterio
Sobresaliente	100%= Precisión >= 70% al predecir si una noticia es potencialmente falsa
Satisfactorios	69% >= Precisión > 51% al predecir si una noticia es potencialmente falsa
Deficiente	51% = Precisión >= 0% al predecir si una noticia es falsa

Tabla 9 – Criterios de evaluación de la precisión definidos

5. Conclusiones de la investigación

Tras la realización de la investigación, se obtienen las siguientes conclusiones:

Los modelos del conjunto de datos beta en español mostraron mejores resultados, en comparación con el conjunto de datos beta en inglés, de lo que podría concluirse que, al realizar la traducción de las noticias, se perdió en cierta medida precisión.

Al analizar el texto de las noticias se debe tener en cuenta las características del lenguaje a analizar ya que elementos como el tamaño de las palabras complejas varía de un idioma a otro.

El proceso de recolección de datos en redes sociales que se asocian a noticias de Costa Rica es un proceso que puede requerir más de una fase de extracción de datos y preprocesamiento para extraer el contenido de la noticia.

La creación de un conjunto de datos y la selección de las variables a utilizar es un proceso complejo que requiere de conocimiento del área de estudio o el acompañamiento de expertos a fin de garantizar la calidad del conjunto de datos. (En el caso del presente trabajo se hizo uso de la herramienta LIWC para la identificación de variables lingüísticas y otros procesos psicológicos y sociales que fueron de importancia para la creación del conjunto de datos).

El proceso de modelado y evaluación de modelos de aprendizaje automático conlleva múltiples iteraciones y experimentos, sobre todo cuando se trabaja con múltiples variables, en busca de mejores resultados.

A pesar de que para la realización del trabajo se emplearon conjuntos de datos beta fue posible obtener resultados satisfactorios con una precisión de 63,44% para el conjunto de datos en español y 62,77% para el conjunto de datos en inglés dejando en evidencia que a través del análisis del texto si es posible realizar el proceso de clasificación binaria de noticias.

Se espera que este proyecto sirva como referencia a trabajos futuros que permitan reducir los efectos negativos causados por las noticias falsas, favoreciendo de esta manera a los diferentes usuarios que acceden a la información.

Referencias

Abonizio, H. Q., Morais, J. I., Tavares, G. M., & Junior, S. B. (2020, 05). Language-Independent Fake News Detection: English, Portuguese, and Spanish Mutual Features. Retrieved from https://www.mdpi.com/1999-5903/12/5/87/pdf

Acosta, F. A. (2019). Construcción de un dataset de noticias para el entrenamiento de clasificadores automatizados. Retrieved from http://oa.upm.es/53685/

Anoop, Manjary, Deepak, & Lajish. (2018, 12). Leveraging Heterogeneous Data for Fake News. Retrieved from https://www.researchgate.net: https://bit.ly/2CnkT5A

Badaskar, S., Agarwal, S., & Arora, S. (2008). Identifying real or fake articles: Towards better language modeling. In Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II. Springer International Publishing.

Brenes, C., Pérez, R., Siles, I., Camacho, S., & Solís, C. (2019, 07 31). Informe de resultados del estudio de opinión sociopolítica: "Audiencias Noticiosas de Medios Digitales: Redes Sociales, Whatsapp y Noticias Falsas". Retrieved from http://cicom.ucr.ac.cr: https://bit.ly/2PNs1uR

Chinchilla-Ugalde, D. (2019). Entrevista sobre detección de noticias falsas en Costa Rica Doble Check . (E. Salazar-Miranda, & G. Tenorio-Arce, Interviewers)

Corporación Latinobarómetro. (2018). http://www.latinobarometro.org/latdocs/INFORME_2018_LATINOBAROMETRO.pdf. p. 77. Retrieved from Informe 2018: http://www.latinobarometro.org/

doblecheck.cr. (2018). https://doblecheck.cr. Retrieved from doblecheck: https://doblecheck.cr/quienes-somos/

Edell, A. (2018, 01 11). I trained fake news detection AI with >95% accuracy, and almost went crazy. Retrieved from https://towardsdatascience.com: https://bit.ly/3kElErY

Facebook. (2018). Facebook for developers. Retrieved from https://developers.facebook.com: https://developers.facebook.com/tools

Google. (2019). Google Translate. Retrieved from Support.google.com: https://support.google.com/docs/answer/3093331?hl=en

Granskogen, T. (2018, 01). Automatic Detection of Fake News in Social Media using Contextual Information. Retrieved from https://ntnuopen.ntnu.no: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2559124

Grupo Nación. (2019). Retrieved from www.nacion.com: https://www.nacion.com/gruponacion/

Horne, B. D., & Adalı, S. (2017, 2017). This JustIn:Fake News Packs a Lot in Title,Uses Simple, Repetitive Content in Text Body,More Similar to Satire than Real News. 6. Retrieved from https://arxiv.org: https://arxiv.org/pdf/1703.09398.pdf

Kunh, M. (2019). TrainControl. Retrieved from https://www.rdocumentation.org: https://bit.ly/3gRbLFb

Lubeck, M. A. (2009). Satire of News. Retrieved from Encyclopedia of Journalism: http://sk.sagepub.com/reference/download/journalism/n342.pdf

Merriam-Webster. (n.d.). "Disinformation." Merriam-Webster.com Dictionary. Retrieved from Merriam-Webster.com dictionary: https://www.merriam-webster.com/dictionary/disinformation

Merriam-Webster. (n.d.). Clickbait. Retrieved from Merriam-Webster.com dictionary: https://www.merriam-webster.com/dictionary/clickbait

Merriam-Webster. (n.d.). Conspiracy theory. Retrieved from https://www.merriam-webster.com: https://www.merriam-webster.com/dictionary/conspiracy%20theory

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The Development and Psychometric Properties of LIWC2015. Retrieved from Universidad de Texas: https://bit.ly/3fFCALe

Pérez, J., & Gardey, A. (2016). Definición de rumor. Retrieved from https://definicion.de/rumor/

Perez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. Retrieved from https://aclweb.org: https://aclweb.org/anthology/C18-1287

Posadas-Durána, J.-P., Gómez-Adornob, H., Sidorovc, G., & Escobara, J. J. (2019, 05). Detection of fake news in a new corpus for the Spanish language. Journal of intelligent and Fuzzy systems, 1,8. doi:10.3233/JIFS-179034

presidencia.go.cr. (2018, 12 17). Anuncian acciones contra noticias falsas en línea. Retrieved from https://presidencia.go.cr: https://bit.ly/3fVBqeQ

Ramírez-Esparza, N., Pennebaker, J. W., & García, F. A. (2007, 06 01). Psicología del Uso de las Palabras: Un Programa de Computadora que Analiza Textos en Español [The Psychology of Word Use: A Computer Program that Analyzes Texts in Spanish]. Revista Mexicana de Psicología, 24. Retrieved from https://bit.ly/2DS47Ml

Ruchansky, N., Seo, S., & Liu, Y. (20017). CSI: A Hybrid Deep Model For Fake News Detection. 26th ACM International Conference on Information and Knowledge Management (CIKM) 2017. doi:10.1145/3132847.3132877

Shewan, D. (2017, 09 22). You Can't Not Click: Weighing the Pros & Cons of Clickbait. Retrieved from https://www.wordstream.com/: https://www.wordstream.com/blog/ws/2014/07/15/clickbait

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2019). Fake News Net :A Data Repository with News Content, Social Context and Spatio temporal Information for Studying Fake News on Social Media. Arizona, USA: College of Information Sciences and Technology, Penn State University, University Park.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter. Retrieved from https://arxiv.org/abs/1708.01967

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017, 06). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved from https://dl.acm.org/citation.cfm?id=3137600: https://bit.ly/2DV7LVL

Sullivan, M. (2018). Why Dictionary.com chose "misinformation," not "disinformation" as word of the year. Retrieved from https://www.fastcompany.com: https://bit.ly/3kBCMPd

Twitter. (2018). Developer. Retrieved from https://developer.twitter.com: https://developer.twitter.com/en

Webhose.io. (2018). Output Reference. Retrieved from webhose.io: https://docs.webhose.io/docs/output-reference-5

Zhou, X., & Zafarani, R. (2018). Fake News: A Survey of Research, Detection Methods,and Opportunities. ACM Comput. Surv.1. ACM.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 22/08/2020
Aceitação/Acceptance: 15/10/2020

Propuesta Arquitectónica de un Sistema de Recomendación Híbrido Adaptativo

Daniela Varela[1], Jose Aguilar[1,2], Julián Monsalve-Pulido[1], Edwin Montoya[1]

danielavarelatabares@gmail.com; aguilar@ula.ve; jmsistemas@gmail.com; emontoya@eafit.edu.co

[1] Universidad EAFIT, Carrera 49 # 7 sur 50, 050021, Medellín, Antioquia, Colombia

[2] Universidad de Los Andes, Facultad de Ingeniería, Núcleo La Hechicera, Final Av. Alberto Carnevali, Mérida 5101, Mérida, Venezuela.

Pages: 171-184

Resumen: Existe una gran variedad de sistemas de recomendación principalmente basados en contenido y filtros colaborativos. Cada uno presenta ventajas y desventajas, por lo que combinarlos en un sistema híbrido es un enfoque prometedor, y para hacerlo, existen múltiples propuestas en la literatura. La mayoría de los enfoques híbridos son estáticos, por lo que es de interés actual explorar su capacidad para adecuarse a las características, tanto del dominio como de los datos disponibles. Este trabajo propone una arquitectura de recomendación híbrida adaptativa, que responde al comportamiento dinámico del entorno por medio del uso de métricas (meta-características), a partir de las cuales se determina la configuración híbrida para la recomendación. En los experimentos en un caso de estudio se muestran sus capacidades adaptativas, ejemplificando su funcionamiento, y evidenciando su flexibilidad para ser implementada de varias maneras y en múltiples contextos.

Palabras-clave: Sistemas de recomendación; Algoritmo hibrido; Filtro colaborativo; Basado en Contenido.

An Adaptive Hybrid Recommendation System architecture

Abstract: There is a wide variety of recommendation systems, mainly based on content and collaborative filters. Each one has advantages and disadvantages, so combining them in a hybrid system is a promising approach, and to do that, there are multiple proposals in the literature. Most hybrid approaches are static, so it is of current interest to explore their ability to follow the characteristics, both domain and available data. This work proposes an adaptive hybrid recommendation architecture, which follows the dynamic behavior of the environment through the use of metrics (meta-characteristics), from which the hybrid configuration for the recommendation is determined. In the experiments in a case study, its adaptive capacities are shown, exemplifying its operation, and showing its flexibility to be implemented in various ways and in multiple contexts.

Keywords: Recommendation systems; Hybrid algorithm; Collaborative filter; Content Based.

1. Introducción

Los sistemas de recomendación se han vuelto muy populares por sus capacidades de predecir qué elemento le puede gustar a un usuario en función de sus intereses. El término en la actualidad tiene una connotación amplia, ya que describe a cualquier sistema que produce recomendaciones individualizadas. Dichos sistemas representan un campo de gran interés porque estamos inmersos en un entorno en el que la cantidad de información sigue creciendo exponencialmente, superando ampliamente la capacidad de cualquier individuo para gestionarla.

Los sistemas de recomendación se dividen en categorías según la información que explotan y cómo la aprovechan. Los sistemas colaborativos (CF, por sus siglas en inglés) utilizan el comportamiento de los usuarios agrupados según sus afinidades, los basados en contenido (CB, por sus siglas en inglés) analizan la información disponible (por ejemplo, características de elementos), y los basados en conocimiento explotan el conocimiento del dominio de uso del sistema de recomendación. Cada enfoque de recomendación tiene sus fortalezas y debilidades, sus casos de uso óptimos, o situaciones en las que no pueden ser implementados.

En ese sentido, se han propuesto sistemas de recomendación híbridos que combinan múltiples recomendadores para aumentar la calidad de las recomendaciones y superar los inconvenientes individuales. Siguiendo la taxonomía de Burke (Burke, 2007), los métodos de hibridación pueden clasificarse en tres grupos básicos: monolíticos, paralelizados y canalizados. Los híbridos monolíticos incorporan cada sistema individual en un nuevo sistema de recomendación. Los híbridos paralelos usan varios recomendadores, uno al lado del otro, y emplean un mecanismo de hibridación específico para agregar sus resultados, que generalmente es un proceso de ponderación. Los híbridos canalizados realizan un proceso por etapas secuenciales (en cascada), donde la última produce las recomendaciones al usuario.

Aunque los sistemas de recomendación híbridos han demostrado sus méritos (Bourkoukou et al., 2018), el procedimiento de construir y ajustarlos sigue siendo un proceso complejo, tedioso y lento. Las investigaciones previas sobre sistemas de recomendación híbrido se han centrado principalmente en enfoques de hibridación estática (Aksel et al., 2010). La estrategia fija tiene problemas en dominios dinámicos, y no tienen en cuenta los problemas de optimización que se puedan presentar. Por ejemplo, en la hibridación ponderada, el problema es cómo optimizar el peso de cada sistema de recomendación para adecuarse a las entradas de sistema que son cambiantes en el tiempo (Suriati et al., 2017).

En este trabajo se propone una arquitectura de recomendación hibrida ponderada adaptativa, tal que la predicción es una combinación lineal de las diferentes predicciones. De esta manera, se integran los algoritmos de recomendación existentes, a través de un proceso de combinación dinámica que busca adaptarse al contexto. Las ponderaciones que se plantean como elemento de hibridación final, pueden ser pesos o la proporción que aporta cada tipo de algoritmo en la recomendación final. Es decir, el esquema híbrido propuesto se clasifica dentro de los ponderados (hibridación paralela), y para la ponderación establece dos esquemas: cuántas recomendaciones aporta cada enfoque

de recomendación en el ranking final, o cuál es la relevancia en peso de cada uno para el cálculo de la predicción final, que matemáticamente se expresa como una suma ponderada simple.

En el trabajo se presentada detalladamente la arquitectura, y muestra su capacidad adaptativa en un caso de estudio. Para la experimentación, el sistema hibrido usa un algoritmo de recomendación basada en contenido y dos de filtrado colaborativo. Los resultados experimentales muestran sus capacidades adaptativas en el contexto de estudio, lo que incentiva próximos trabajos que extiendan las capacidades adaptativas de la arquitectura. El artículo se organiza de la siguiente manera. En la sección inicial se presenta la arquitectura propuesta, seguido del caso de estudio, y se culmina con las conclusiones.

2. Arquitectura Propuesta

2.1. Descripción de la Arquitectura

La arquitectura de recomendación hibrida ponderada adaptativa se presenta en la Figura 1. Sus componentes se agrupan en externos e internos, los cuales se describen a continuación.

2.2. Componentes Externos al proceso de Hibridación

A continuación, se describen los componentes del sistema que no actúan directamente en el proceso de hibridación, que pueden ser consumidos como servicios externos ya existentes, pero que generan sus datos de entrada:

- **Sistema de perfilamiento de usuario:** es el responsable de la extracción de las características de los usuarios (ejemplo: el estilo de aprendizaje en el caso de estudiantes), para tener más información sobre sus intereses, preferencias, necesidades, etc.
- **Sub-Base de Datos**: en este paso se utiliza el perfil de los usuarios (es decir, toda la información que se tenga del usuario), para pre-filtrar una pequeña porción de los datos, que eventualmente se pueden adaptar al usuario. Estos ítems serán conocidos como ítems candidatos para el usuario en cuestión.
- **Base de conocimiento:** contiene la información de todos los metadatos de los ítems (artículos, patentes, objetos de aprendizaje, videos, etc.) que se pueden recomendar, así como la retroalimentación proveniente de las interacciones de los usuarios y demás datos que alimenten el sistema.
- ***Mecanismo de Caracterización de la Retroalimentación*:** establece la retroalimentación para el sistema de recomendación hibrido por la recomendación realizada. La retroalimentación puede determinarse de manera explícita e implícita. En el caso explícito, se usan las estrategias de calificación directa por parte de los usuarios. Pero además de estas, se usan retroalimentaciones implícitas, que vienen de observaciones del entorno, como resultados en los exámenes, referencias que hacen con el material recomendado, visitas a los ítems, entre otros.

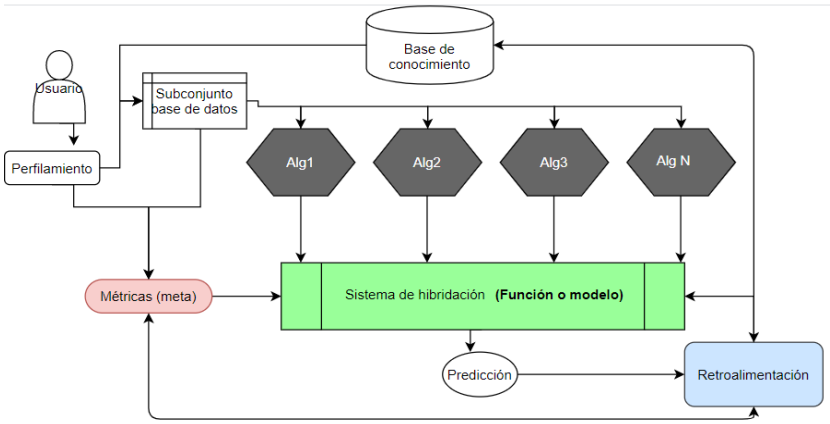


Figura 1 – Arquitectura de recomendación hibrida ponderada adaptativa

En general, no hay estándares para convertir una retroalimentación implícita en calificaciones explícitas, y en los trabajos previos se ha hecho según la aplicación particular y los objetivos de la implementación. En nuestro enfoque se consideran ambos tipos de retroalimentación, de la manera como se muestra en la Tabla 1. Los valores de cada acción fueron definidos de acuerdo con la importancia que tiene cada una de ellas dentro del contexto de un sistema y proceso de recomendación. Ahora bien, dichos valores se muestran para ejemplificar el proceso de retroalimentación combinado en este trabajo, pero pudiesen modificarse según el interés que se le quisiese dar en otro contexto de estudio.

Fuente de retroalimentación	**Tipo**	**Valor de Retroalimentación**
Click (A)	Implicita	2.0
Guardar (B)	Implicita	3.0
Tiempo abierto (C)	Implicita	3.0 (si se abre más de un tiempo t)
Recomendar a otros (D)	Implicita	3.5
Rating	Explicita	Valor dado por el usuario

Tabla 1 – Tipos de Retroalimentaciones

La Tabla 1 define el valor de retroalimentación según el tipo, y convierte una retroalimentación implícita en explicita. Si en una misma sesión ocurren varios eventos implícitos (por ejemplo, guardar (B) y recomendar a otros (D)), entonces se puede usar para determinar cómo retroalimentación el promedio de ellas, el máximo, entre otros posibles enfoques de agregación. En este trabajo se utiliza el promedio simple entre ellas.

2.3. Componentes Internos al proceso de Hibridación

Son los componentes que intervienen en el proceso de hibridación:

- **Algoritmos de Recomendación:** son los diferentes sistemas de recomendación que se están usando en el proceso para recomendar, los cuales pueden ser de diferentes tipos.
- **Métricas:** son indicadores numéricos usados para describir a un usuario y sus ítems candidatos en un momento específico en el tiempo. Son variables dinámicas, cuyos valores altos o bajos reflejan el desempeño satisfactorio o deficiente de los algoritmos de recomendación dentro del sistema. En ese sentido, las métricas determinan la ponderación (relevancia) que se le dará a cada algoritmo de predicción, derivado de cómo se están comportando en un caso particular. Ellas definirán el valor del peso o proporción de recomendaciones de cada algoritmo, guiando de esta manera el proceso adaptativo del sistema de recomendación híbrido propuesto. Las métricas son computadas en tiempo de ejecución usando como insumo la salida del módulo "Mecanismo de Caracterización de la Retroalimentación". Este módulo realiza dos tareas en específico:
  - **Definir las métricas:** establece las expresiones matemáticas que calculan indicadores sobre la recomendación realizada, y están basados en las características de funcionamiento/rendimiento de los algoritmos de predicción usados, capturadas por el módulo "Mecanismo de Caracterización de la Retroalimentación".
  - **Convertir esas métricas en una ponderación a darle a cada algoritmo:** la ponderación puede ser un peso de relevancia del algoritmo, o proporción de participación del algoritmo en la recomendación a hacer.
- ***Sistema de hibridación*:** es el motor del sistema de recomendación, y recibe la salida del módulo de métricas definidas como ponderaciones y de los algoritmos de recomendación, para determinar la predicción (recomendación) final. El sistema de hibridación puede realizar una combinación de recomendaciones de cada uno de los algoritmos, de forma lineal o no lineal. En este trabajo se utiliza una combinación lineal.

3. Casos de Estudio

El objetivo de esta sección es describir el proceso seguido por el sistema de recomendación híbrido, y analizar su capacidad adaptativa. El paso inicial es el *perfilamiento de los usuar*ios, que permite la selección de los *ítems candidatos*. Ahora bien, en este caso de estudio, al ser un conjunto de datos relativamente pequeño, no se presentan inconvenientes computacionales para utilizarlos completo. Se asume que los ítems candidatos son los mismos para todos los usuarios.

Los datos utilizados son registros de 12 meses de Desktop, una plataforma de comunicaciones internas desarrollada por CI&T, que contiene alrededor de 73.000 interacciones de los usuarios con más de 3.000 artículos. Los datos ofrecen información de los artículos, como el título, el texto y el ID; y de las interacciones, como el tipo

de evento, el momento en que se realizó (*timestamp*), el ID del usuario y del ítem involucrado[1].

Con la intención de reflejar el comportamiento del sistema en el tiempo, la evaluación se hizo dividiendo los datos ordenados cronológicamente, utilizando los datos de tiempos menores a un tiempo de corte *t* para el entrenamiento. Para la prueba, el cálculo del error se hizo sobre el valor real de la calificación dada por el usuario en los artículos después del tiempo de corte.

Los tipos de interacciones definidos en el conjunto de datos original son "*view*", "*like*", "*comment*", "*follow*" y "*bookmark*". Estos se convirtieron, respectivamente, en los valores de "click", "rating", "recomendar", "tiempo abierto" y "guarda", para darles los valores según la Tabla 1 .

3.1. Algoritmos de Recomendación

Se seleccionaron tres algoritmos (uno CB y dos CF), para ejemplificar la implementación del sistema de hibridación propuesto. La recomendación basada en contenido sugiere ítems al usuario en función de la similitud entre su perfil y el contenido de estos, teniendo en cuenta que el perfil se define por los ítems con los que el usuario ha interactuado en el pasado o con atributos predefinidos. Para caracterizar los ítems, se utilizan descriptores como palabras claves, tags o vectores numéricos que permiten realizar una comparación cuantitativa entre ellos (B.Thorat, M. Goudar, & Barve, 2015). Cuando se recomienda bajo esta aproximación, no se presenta variabilidad en los resultados, siempre van a tener una similitud más alta los ítems cuyo contenido sea del agrado del usuario, dejando a un lado ítems novedosos e interesantes (Cacheda, Carneiro, Fernández, & Formoso, 2011).

Los sistemas que utilizan filtrado colaborativo se basan en la recopilación de interacciones de muchos usuarios, y en explotar las similitudes o diferencias entre estas para recomendar. Existen dos grandes categorías dentro del CF: basado en modelos y basado en memoria. Los primeros construyen inicialmente un modelo que represente el comportamiento de las interacciones de los usuarios, y los basados en memoria utilizan el concepto de "vecindario" entre usuarios o elementos para hacer recomendaciones (Cacheda et al., 2011).

En este trabajo, se implementa un CB donde el perfil del usuario está determinado únicamente por sus calificaciones pasadas, un algoritmo de CF basado en la similitud entre usuarios (basado en memoria), y un modelo de descomposición en valores singulares (SVD, por sus siglas en inglés) que se basa en la factorización de la matriz en factores latentes.

3.2. Determinación de las métricas

Se proponen tres métricas en función de las características de los motores de recomendación empleados, para correlacionar sus rendimientos. Es importar acotar, que la definición de las métricas dependerá de los algoritmos a utilizar, y del dominio y las características de los datos. Primero, se define cada métrica, explicando su propósito

[1] https://www.kaggle.com/gspmoreira/articles-sharing-reading-from-cit-deskdrop

y cómo se espera que sus valores reflejen el rendimiento de los algoritmos relacionados. Luego, se propone una ecuación para su cálculo y se muestran los resultados, para valorar su comportamiento.

Para analizar las métricas se dividen los datos en dos muestras con diferente número de interacciones, siguiendo un orden cronológico en el tiempo. Ellas simulan dos estados diferentes del sistema, tal que las interacciones del segundo grupo pueden ser interpretadas como futuras respecto a las del primer grupo (ver Tabla 2).

	# Interacciones	# usuarios	# items	Dispersion (sparsity)
Grupo 1	1,000	234	129	0.9777
Grupo 2	40,000	1525	1773	0.9923

Tabla 2 – Análisis de métricas

3.2.1. Métrica 1 – Actividad del usuario

Describe la cantidad de interacciones del usuario, representando en qué tan llena está su fila específica dentro de la matriz. Se define como el número de interacciones del usuario dividido por el máximo número de interacciones realizadas por un usuario (ver ecuación 2). Toma valores entre 0 y 1.

$$M1(u) = \frac{n_u}{max_{u \in U}} \tag{1}$$

La actividad del usuario se relaciona con el desempeño de los tres algoritmos, de la manera como se describe a continuación, tal que las siglas CB describen al algoritmo basado en contenido, MF al de factorización de matriz, y CF al de filtro colaborativo, respectivamente:

- **CB:** un usuario con mayor M1 tiene más información disponible para configurar su perfil y definir sus atributos, llevando a una recomendación más acertada. No se puede calcular el perfil del usuario si este no ha tenido interacciones, es decir, cuando M1 es cero.
- **MF:** su rendimiento depende del nivel de dispersión de la matriz, como ha sido establecido en (Da Silva, De Moura, & Caloba, 2018), donde comparan los resultados para diferentes niveles de dispersión, sin arrojar una conclusión general porque se comporta de manera desigual en diferentes escenarios. En este trabajo se considera que es mejor tener una matriz menos dispersa, puesto que se tiene más información para inferir características de los usuarios. Además, aun cuando no se tengan interacciones (fila vacía), el algoritmo SVD arroja un valor por defecto.
- **CF:** la similitud entre usuarios es determinada por los ratings. Entre más interacciones tenga un usuario, tendrán más información para extraer sus vecinos. Si un usuario no ha interactuado con ningún ítem, no se puede establecer su vecindario, y el algoritmo no funciona.

En la Tabla 3 se observa el resultado de aplicar el cálculo de M1 a cuatro usuarios con diferentes números de interacciones en los dos grupos de estudio. Si bien M1 da una idea del nivel de actividad de un usuario respecto a los demás, y varía dinámicamente con la cantidad de datos, falla en cuanto a reflejar el nivel de dispersión de la matriz, porque su valor depende en todo momento del usuario más activo. Para el caso reflejado en la Tabla 3, teniendo en cuenta que el total de ítems en cada grupo es 129 y 1773, respectivamente, y que la media de las interacciones de los usuarios es 2 y 13, la gran mayoría de ellos ha interactuado con muy pocos ítems, y los más activos son casos atípicos. Se debe mejorar la métrica explorando otras formas de determinar el nivel de dispersión de los datos para un usuario particular.

Grupo 1			Grupo 2		
Id usuario	**# calificaciones**	**M1**	**Id usuario**	**# calificaciones**	**M1**
-696495	36	1	-696495	504	0.83
-804958	14	0.39	804958	210	0.38
950040	3	0.083	-983140	61	0.1
218054	1	0.028	229373	2	0.003

Tabla 3 – Resultados

3.2.2. Métrica 2 – Actividad de los ítems

Describe el número de interacciones sobre los ítems candidatos. Representa qué tantos usuarios han calificado esos ítems. Ella se define por el número de ítems calificados por más del X% de los usuarios, respecto al total de ítems candidatos (ver ecuación 3). Toma valores entre 0 y 1.

$$M2(u) = \frac{\#\ items\ con\ más\ de\ Y\ iteraciones}{\#\ total\ de\ items} \quad (2)$$

Y: X% del total de usuarios.

Para las pruebas realizadas, se tomó un valor X = 5. En este caso, M2 es la proporción de ítems candidatos que ha sido calificado por más del 5% de los usuarios. La actividad de los ítems tiene relación con el desempeño de los algoritmos MF y CF, así:

- **MF:** el análisis para este algoritmo es equivalente al que se hizo para la primera métrica. El nivel de dispersión afecta el rendimiento del modelo en término generales, sin definir exactamente en qué medida.
- **CF:** como ya se mencionó, este algoritmo se afecta fuertemente por la dispersión, si se tiene una mayor interacción sobre los ítems es probable que se amplíe el conjunto de ítems que recibirán una predicción adecuada. Su rendimiento se ve afectado por la cantidad de información disponible para reconocer posibles vecinos, puesto que la similitud entre ellos se basa en sus patrones de calificación.

Como no se tienen ítems candidatos para cada usuario, M2 es igual para todos los usuarios en su respectivo tiempo. El 5% de los usuarios corresponde a 11 (de 234) y

76 (de 1525), respectivamente, para los grupos 1 y 2. Como se observa en la Figura 2, pocos ítems superan ese umbral establecido, los valores encontrados para M2 son muy bajos (0.1 y 0.006), por lo que se podría utilizar un umbral menor para obtener valores mayores. Esta forma de calcular la actividad de los ítems tiene la ventaja de que varía con el tiempo de acuerdo con los cambios en el número de usuarios e ítems, además de poderse ajustar cambiando el umbral.

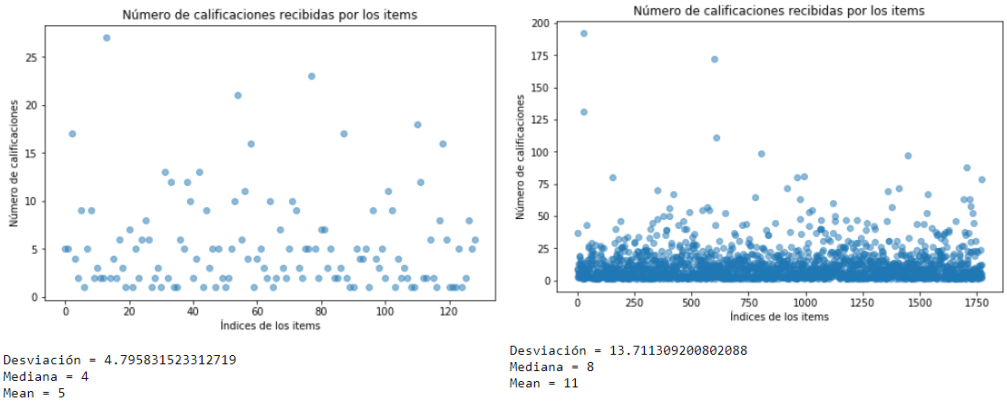


Figura 2 – Número de clasificaciones para cada ítem

3.2.3. Métrica 3 – Descriptores de los ítems

Describe qué tan adecuados son los descriptores para representar los ítems, es decir, qué tan bien caracterizan a sus respectivos ítems. Se calcula para los ítems candidatos, utilizando la media de los promedio de los valores TD-IDF para las 10 palabras claves de cada ítem candidato. Se extraen las 10 primeras palabras claves de cada ítem, con su respectivo valor TD-IDF, y se saca el promedio para caracterizar cada ítem. Finalmente, se promedia también esta caracterización individual para sacar un valor general. La actividad de los ítems tiene relación con el desempeño del algoritmo CB, así:

- **CB:** el perfil de cada ítem, y, por consiguiente, del usuario, está definido en base a los descriptores. Si estos tienen la capacidad de describir bien el contenido de los ítems, el algoritmo CB funciona mejor; mientras que en caso contrario los perfiles de usuarios e ítems no reflejarán sus características reales. El valor de M3 nunca es cero.

En este caso se asume igual para los usuarios del grupo 1 y el grupo 2. Los valores promedio de M3 por ítem fueron 0.221 y 0.218, respectivamente. Como se evidencia con los resultados de M3, el valor promedio de los descriptores tiene muy poca variabilidad, y será muy similar independiente del subconjunto de ítems que se tome, lo que lleva a concluir que esta métrica no aporta información determinante calculada de esta forma, con los descriptores que se tienen. Futuros trabajos deberán explorar otras formas de caracterizar los descriptores.

3.3. Determinación de las Ponderaciones

Después del proceso de definición y cálculo de las métricas, se continúa con su utilización para establecer la relevancia de cada algoritmo dentro de la recomendación final. Se usa un sistema difuso para dar un valor final de importancia a cada algoritmo. Para ello, se definen un conjunto de reglas difusas que establecen la relación entre los algoritmos de predicción y las métricas calculadas. El modelo difuso es definido por reglas difusas del tipo:

<Si M1 y M2 y M3 entonces CF y MF y CB>

Donde todas las variables difusas son definidas por tres conjuntos difusos: alto, medio y bajo. Se plantean unas funciones de membresía sencillas, como la mostrada en la Figura 3, para todas las variables difusas.

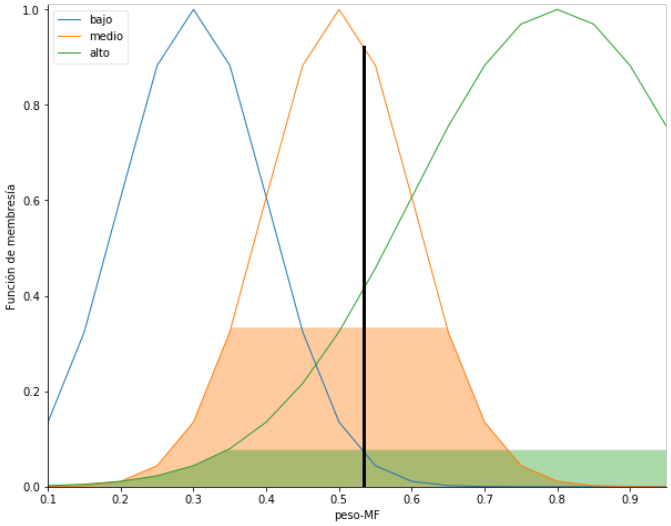


Figura 3 – Conjuntos difusos de las variables difusas

En general, las reglas reflejan diferentes tipos de usuarios y momentos dentro del ambiente de recomendación. Cada fila de la Tabla 4 describe una posible regla. Al introducir los valores puntuales de las métricas, el sistema de reglas difuso genera un valor de peso o proporción de participación de cada algoritmo en la hibridación final. Para convertir los valores finales en tres valores que sumen 1 (100%), se normalizan dividiendo cada peso sobre la suma de los tres pesos (ver ecuación 4).

$$w_{i-normalizado} = \frac{w_i}{\sum_j^3 wj} \tag{3}$$

Según lo mostrado en la Tabla 4, el sistema hibrido permite seleccionar los algoritmos de recomendación que en algún momento se deben activar, según los valores de las métricas.

Las métricas establecen los contextos funcionales del sistema de recomendación híbrido, lo que permite establecer como salida del proceso de inferencia difuso la combinación adecuada de los algoritmos de predicción a usar.

M1	M2	M3	CF	MF	CB
ALTO	ALTO	*	ALTO	ALTO	**
ALTO	*	ALTO	**	**	ALTO
BAJO	ALTO	*	BAJO	MEDIO	**
BAJO	*	ALTO	**	**	MEDIO
NULO	ALTO	*	NULO	MEDIO	**
NULO	*	ALTO	**	**	NULO
ALTO	BAJO	*	MEDIO	MEDIO	**
ALTO	*	BAJO	**	**	MEDIO
BAJO	BAJO	*	BAJO	BAJO	**
BAJO	*	BAJO	**	**	BAJO
NULO	BAJO	*	NULO	BAJO	**
NULO	*	BAJO	**	**	NULO

Explicación de símbolos utilizados: * Esa variable puede tomar cualquier valor, no es relevante para esa regla. ** Esa regla no afecta, o no es relevante, para ese algoritmo de recomendación.

Tabla 4 –Reglas del sistema difuso

4. Experimentación

Para dar un ejemplo del proceso de inferencia difuso para dar una recomendación final, se selecciona el usuario con ID: -696495. Ese usuario es el más activo del primer grupo, con 36 interacciones de 129 ítems posibles, por lo que es válido suponer que solo se activan las reglas de la segunda y séptima filas de la Tabla 4: un usuario activo, ítems con pocas interacciones y bien descritos. Así, el proceso de inferencia difuso determina los porcentajes 45%, 27.5% y 27.5% para CB, MF y CF respectivamente (según ecuación 4). Posteriormente, el usuario interactúa solo con 5 ítems (los supondremos como los valores correctos). A continuación, el algoritmo hibrido utiliza los porcentajes para hacer una suma ponderada de las predicciones de afinidad de los tres algoritmos respecto a los ítems.

Al comparar las recomendaciones hechas por los algoritmos de predicción y el hibrido, con los que realmente interactúa el usuario (los valores correctos), se obtienen los resultados de la Tabla 5, que muestra el error cuadrático medio (RMSE) entre lo recomendado y lo correcto. Si bien el RMSE es bajo, y es menor en la hibridación que en los dos primeros algoritmos implementados individualmente, no lo es en el caso del CF. Por este comportamiento, se procede a realizar la hibridación desde el segundo enfoque (proporción de participación de cada algoritmo en el ranking final).

Usuario	CB	MF	CF	Híbrido
-696495	0.2395	0.2690	0.068	0.177

Tabla 5 – Resultados del error cuadrático medio

Para este experimento, se considera el top 20, y se determina cuántos de los 5 ítems que el usuario ya se sabe que prefirió en el futuro son predichos dentro de esa recomendación. Los porcentajes representan cuántos elementos del top 20 serán recomendados por cada algoritmo. Para este caso, 9 son de CB, y 6 para MF y CF. En los resultados se obtiene que, dentro de los 5 ítems seleccionados por el usuario, 2 de ellos aparecen entre los recomendados por el ranking híbrido. Cuando se realiza el mismo experimento con los algoritmos individuales se obtienen 2, 1 y 1 ítems, respectivamente. Finalmente, para dar una idea general del desempeño de la hibridación respecto al de los algoritmos individuales, se realizó un experimento de aciertos, en el que se calcula cuántos de los ítems con los que el usuario ya interactuó son predichos dentro de sus top 20. Se hizo para cada caso: CB, MF, CF e Híbrido, tomando diferentes valores aleatorios de pesos para el caso del hibrido (test1 a test5). Los usuarios han interactuado, respectivamente, con 36, 14 y 3 ítems de abajo hacia arriba. Los resultados se muestran en la Tabla 6.

	Número de ítems predichos dentro del top 20								
USUARIO ID	CB	MF	CF	TEST0	TEST1	TEST2	TEST3	TEST4	TEST5
-696495	6	4	12	10	6	5	14	13	5
-804958	7	7	3	9	8	8	7	7	9
950040	1	1	1	2	1	1	2	2	1

Tabla 6 – Resultados general de desempeño de hibridación

Este ensayo realiza el proceso de hibridación por medio del enfoque de suma ponderada. Los resultados muestran que, en general, los máximos aciertos se presentan en casos de hibridación y varían de acuerdo con los usuarios, hipótesis que se defiende en este trabajo. Así, se logra verificar que la combinación debe ser dinámica según el usuario y el contexto.

5. Conclusiones

En este trabajo se propuso una arquitectura de recomendación híbrida adaptativa, además se elaboró un prototipo de esta, usando algoritmos de recomendación basados en contenido y en filtrado colaborativo. Para el caso de estudio se utilizó como esquema de predicción/ponderación una combinación lineal de las ponderaciones, y se exploraron dos maneras de hacer esas recomendaciones: como pesos o proporción de ítems a recomendar. Los resultados y ejemplos presentados reflejan el principio de funcionamiento de cada parte del proceso, y los beneficios de utilizar un sistema de

recomendación hibrido que se adecúe a los datos presentes en el momento de ejecución, que configure dinámicamente la hibridación para cada usuario, en función de las ventajas y desventajas de los diversos enfoques de recomendación determinada en tiempo real. Además, con el sistema difuso propuesto y las métricas definidas, se evidencia la capacidad de solucionar problemas existentes en los algoritmos de recomendación individuales, como el arranque en frío.

Para próximos trabajos se seguirán desarrollando los otros aspectos adaptativos de la arquitectura. Por ejemplo, que el motor de hibridación considere modelos no lineales adaptativos para la combinación de los algoritmos. También, otros trabajos futuros tendrán que analizar diferentes formas de extracción de las características implícitas y explicitas a considerar en la retroalimentación, y su incorporación en las métricas, para tomar en cuenta la diversidad de aspectos que están presentes en un proceso de recomendación. Igualmente, en el prototipo elaborado se usa un sistema difuso para inferir la ponderación, al cual se le pueden dar capacidades adaptativas bajo un enfoque supervisado, tal que trate de minimizar un error, a partir de lo recomendado y la recomendación ideal.

Por otro lado, el esquema hibrido debe ser incorporado en sistemas de recomendación autonómicos como el propuesto en (Monsalve-Pulido, Aguilar, Montoya, & Salazar, 2020), tal que la determinación de sus métricas pueda ser hecha de manera más autónoma y dinámica, y debe ser probado en sistemas de recomendación basados en conocimiento, como el propuesto en (Aguilar, Valdiviezo-Díaz, Riofrio, 2017). Finalmente, la propuesta debe ser integrada a otros proyectos de análisis de datos (Aguilar, Cordero, Buendía, 2018), para explotar los datos del contexto en la adaptación del algoritmo.

Agradecimientos

Este trabajo ha sido apoyado por el proyecto 64366: “Contenidos de aprendizaje inteligente a través del uso de herramientas de Big Data, Analítica Avanzada e IA”- Ministerio de Ciencia – Gobernación de Antioquia - República de Colombia.

Referencias

Aguilar J., Valdiviezo-Díaz P., Riofrio R. (2017) A general framework for intelligent recommender systems, Applied Computing and Informatics, 13(2), 147-160. doi: 10.1016/j.aci.2016.08.002

Aguilar J., Cordero J., Buendía O. (2018) Specification of the Autonomic Cycles of Learning Analytic Tasks for a Smart Classroom, Journal of Educational Computing Research, 56(6), 866-891. doi: 10.1177/0735633117727698

Aksel, F., & Birtürk, A. (2010). An adaptive hybrid recommender system that learns domain dynamics. In *Proceeding European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (p. 49), Barcelona, Spain.

Bourkoukou, O., & El Bachari, E. (2018). Toward a hybrid recommender system for e-learning personnalization based on data mining techniques. *International Journal on Informatics Visualization*, 2(4), 271-278.

Burke, R. (2007). Hybrid web recommender systems. In P. Brusilovsky, W. Nejdl, A. Kobsa (Eds.), *The adaptive web: Methods and Strategies of Web Personalization* (pp. 377-408). Berlin, Germany: Springer.

Cacheda, F., Carneiro, V., Fernández, D., & Formoso, V. (2011). Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web*, 5(1), 1-33.

Da Silva, J. F. G., de Moura Junior, N. N., & Caloba, L. P. (2018). Effects of data sparsity on recommender systems based on collaborative filtering. In *Proceeding* International Joint Conference on Neural Networks (pp. 1-8). Rio, Brazil, IEEE.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering education*, 78(7), 674-681.

Monsalve-Pulido, J., Aguilar, J., Montoya, E., & Salazar, C. (2020). Autonomous Recommender System Architecture for Virtual Learning Environments. *Applied Computing and Informatics*. doi: 10.1016/j.aci.2020.03.001

Suriati, S., Dwiastuti, M., & Tulus, T. (2017). Weighted hybrid technique for recommender system. *Journal of Physics Conference Series*, 930(1).

Thorat, P. B., Goudar, R. M., & Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, 110(4), 31-36.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 28/08/2020
Aceitação/Acceptance: 16/10/2020

Clasificación de Imágenes de Cámaras de Vigilancia Volcánica usando Visión Artificial

César D. Martínez[1], Roberto C. Uribe[1], Carlos Ospina[2], Rubiel Vargas-Cañas[1]

cesarmartinez@unicauca.edu.co; robertocarlos@unicauca.edu.co; cospina@sgc.gov.co; rubiel@unicauca.edu.co

[1] Universidad del Cauca, Calle 5 # 4-70, 190003, Popayán, Colombia.

[2] Servicio Geológico Colombiano, Calle 5B N.º 2-14, 190002, Popayán, Colombia.

Pages: 185-198

Resumen: Los Observatorios Vulcanológicos en el mundo, y en Colombia, cuentan con cámaras de vigilancia ubicadas en sectores aledaños a volcanes donde se pueden presentar manifestaciones importantes de actividad volcánica en superficie. En Colombia estas cámaras capturan una imagen por minuto, así el volumen de información puede ser difícil de manejar en tiempo de crisis volcánica. Así, los sistemas automáticos pueden ayudar a mejorar la efectividad haciendo un diagnóstico de la actividad volcánica. En este trabajo se muestra la implementación de técnicas de aprendizaje de máquina y visión por computador tales como: extracción de características mediante *Bagoffeature* y algoritmos de clasificación, para etiquetar automáticamente imágenes de diferentes volcanes: Nevado del Ruiz, Nevado del Huila y Puracé. Las imágenes se dividieron en nueve categorías probando varios modelos de clasificación, alcanzando una exactitud superior al 88%. Finalmente se implementó una interfaz de usuario que permitió evaluar el rendimiento del sistema por los operarios.

Palabras-clave: Actividad volcánica; Extracción de característica; Procesamiento de imágenes; Monitoreo visual; Algoritmos de clasificación.

Classification of Volcanic Surveillance images using artificial vision

Abstract: Volcanological Observatories around the world, and in Colombia, have surveillance cameras located near to the volcanoes, where they may present significant manifestations of surface volcanic activity. These cameras acquire around one image per minute; therefore, the volume of information may be difficult to manage, especially in times of volcanic activity crisis. To that extent, automatic systems may help to improve difficulty by making a diagnosis of volcanic activity. In this study, machine learning and computer vision techniques were implemented such as: extraction of characteristics using Bagoffeature and classification algorithms to automatically classify images of different volcanoes: Nevado del Ruiz, Nevado del Huila and Puracé. The images were divided into nine labels and two models were trained, as a result, they achieved a final accuracy of over 88%. Finally, a user interface was deployed that allows evaluating the system performance by the operators.

Keywords: Volcanic activity; Feature extraction; Image processing; Visual monitoring; Classification algorithms.

1. Introducción

América Latina y Colombia en específico, es una región con un gran número de volcanes activos junto con grandes poblaciones expuestas a sus peligros. En la historia de la región, numerosas erupciones han demostrado las dificultades que genera vivir cerca de centros eruptivos activos, impactando en el desarrollo y organización de las comunidades (Ferrés, Escobar & Bowman, 2011). Ejemplos de ello hay varios, la erupción volcánica ocurrida en el volcán Nevado del Ruiz el 13 de noviembre de 1985 fue la tercera con mayor cantidad de vidas perdidas (cerca de 25000) a nivel mundial, desde 1.800 y la primera en América latina, con más de 17 municipios afectados, entre los que cuenta la tragedia de Armero, ciudad totalmente arrasada, dejando más de 200.000 personas afectadas directa e indirectamente (Hermelin, 2005). También, se recuerda la trágica erupción del volcán de fuego en Guatemala el 3 de junio del 2018 dejando un saldo de 114 personas fallecidas, 197 desaparecidas, más de 3.200 albergadas y aproximadamente un estimado de 1.7 millones de personas afectadas directa e indirectamente (Romano, 2019).

En Colombia el Servicio Geológico Colombiano (SGC), tiene como funciones la evaluación y monitoreo de la actividad volcánica de los volcanes activos del país, el estudio de las amenazas y de los riesgos naturales; para tal efecto se crearon Observatorios Vulcanológicos y Sismológicos (OVS) ubicados en Manizales (OVSM) y posteriormente, los de Pasto (OVSP) y Popayán (OVSPo) ("Servicio Geológico Colombiano", 2020). Como técnica de vigilancia volcánica, las apreciaciones visuales de imágenes logradas por medio de cámaras de vigilancia son muy importantes para el diagnóstico efectivo y oportuno del estado de actividad de un volcán (Samaniego, 2003). Por su parte, los avances en el procesamiento digital de imágenes (PDI) y Aprendizaje Automático han sido de gran ayuda en diversos campos, que para el caso de los OVS's han aplicado técnicas de aprendizaje de máquina para la clasificación de actividades sísmicas (Asim, Moustafa, Niaz, Elawadi, Iqbal & Álvarez, 2017) una de las principales tareas del monitoreo volcánico, pero no se ha trabajado en técnicas de visión artificial para el monitoreo visual a través del aprendizaje de máquina.

Aunque en la revisión de la literatura se encontró solo un trabajo que hace uso del procesamiento de imágenes para monitorear volcanes, si se encontraron documentos relacionados con el monitoreo volcánico y las técnicas de aprendizaje automático, por ejemplo, Lakiy, (Cadena & Muñoz, 2018) desarrollaron un sistema automático-supervisado para la aplicación del procesamiento primario en tiempo real para la detección de sismos tipo LP, VT o TRE y su clasificación, usando coeficientes de predicción lineal, energía espectral y amplitudes en el dominio del tiempo y redes neuronales artificiales (RNA), su validación alcanzó un rendimiento 92%; Tag EQ (León, Ordoñez & Bucheli, 2019) presentan un sistema de etiquetado semiautomático de señales sísmicas, a partir de un catálogo de eventos sísmicos, utilizando clasificadores como las máquinas de soporte vectorial y bosques de decisión las cuales presentan una precisión por encima del 80%. Por otro lado (Witsil & Johnson, 2020) utilizan algoritmos de detección para alimentar una red neuronal artificial (RNA) entrenada para clasificar la actividad

observable en cinco etiquetas. Como resultados obtenidos de la clasificación alcanzan una precisión general del 93%. De manera general la clasificación automática agiliza procesos que son de gran importancia para el tipo de actividad que se está analizando y que permiten un procesamiento de miles de datos en cortos lapsos de tiempos. La combinación de técnicas de visión por computador es escasa en temas relacionados con vigilancia volcánica convirtiéndose en un trabajo pionero en Colombia.

Con el objetivo de avanzar en la clasificación automática de imágenes de cámaras de vigilancia volcánica, en este trabajo se implementó un sistema basado en PDI y técnicas de aprendizaje de máquina, para ello, se desarrolló un sistema que posibilitan la extracción de características de las imágenes de los fenómenos volcánicos y/o atmosféricos para así entrenar clasificadores que realizan un etiquetado automático de los datos obtenidos mediante las cámaras.

El presente artículo está organizado de la siguiente manera: la sección 2 detalla las etapas de la metodología planteada para el desarrollo de la clasificación, en la sección 3 se evidencian los resultados obtenidos para cada etapa de la metodología, la sección 4 se presenta una discusión de los resultados obtenidos y finalmente la sección 5 muestra las conclusiones.

2. Metodología

El desarrollo de esta propuesta siguió las seis etapas de la metodología CRISP-DM (Osorio, 2019), la cual es una metodología estándar que ha sido ampliamente probada y demuestra su utilidad en problemas de manejo de altos volúmenes de datos, los algoritmos utilizados se implementaron en la plataforma de programación Matlab usando el Toolbox *Statistics and Machine Learning*. A continuación, se detalla cada una de las etapas de la metodología.

2.1. Adquisición de imágenes

Para este trabajo se partió de las imágenes de tres de los volcanes vigilados por el SGC, las cámaras se nombran de acuerdo al lugar donde se encuentran ubicadas: (i) volcán nevado del Ruiz, se trabajó con las cámaras web ubicadas en los sitios conocidos como: Piraña Azufrado, Gualí y Cisne; (ii) volcán Puracé, se usaron las cámaras web ubicadas en los sitios conocidos como: Minas y Lavas Rojas; y (iii) volcán nevado del Huila, donde se usaron las cámaras web ubicadas en los sitios conocidos como: Tafxnu y Maravillas. Las imágenes fueron recibidas y almacenadas de forma organizada en una unidad de red que permite la búsqueda y visualización de imágenes de los últimos nueve años, los OVS's hacen uso de recursos compartidos que pueden ser accedidos desde cualquier OVS's en caso de ser necesario.

2.2. Construcción del conjunto de datos

La construcción del conjunto de datos fue realizado a través de VolCAM (Figura 1) (Ospina & Narváez, 2020) el cual es un aplicativo de escritorio desarrollado en el SGC, que tiene por objetivos apoyar las labores del monitoreo volcánico, a partir de la organización y análisis de material multimedia asociado a fenómenos volcánicos

y actualmente funciona para cámaras de vigilancia del SGC. El programa permite la observación del catálogo de imágenes desplegado por cada cámara, con el fin de realizar la visualización de las imágenes y el proceso de etiquetamiento manual para realizar una generalización, categorización y agrupar a aquellas que pertenecen a escenas de una misma familia.

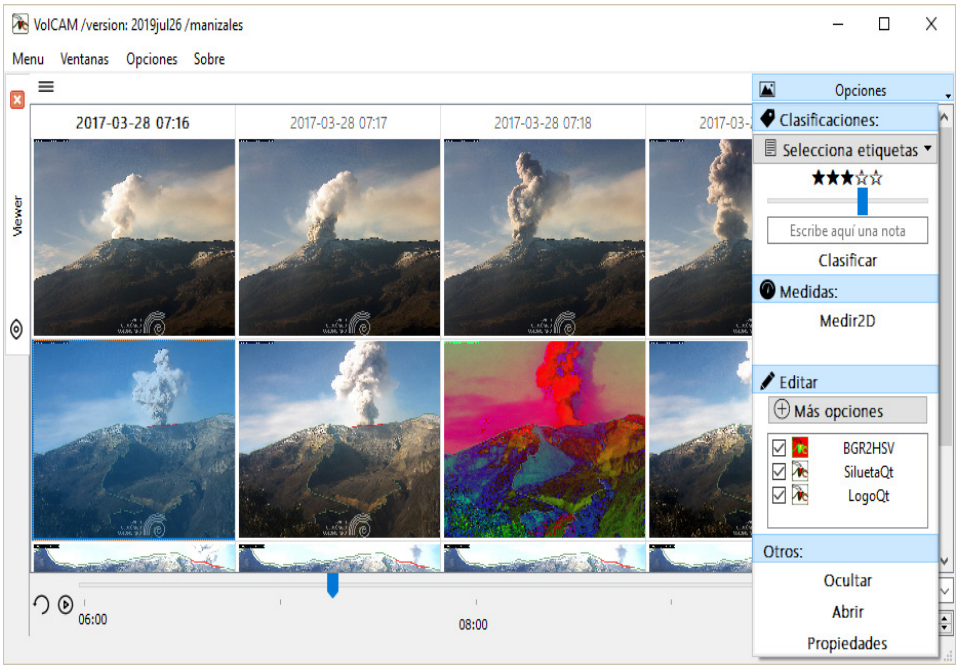


Figura 1 – Despliegue de catálogo de imágenes usando VolCAM

2.3. Preparación del conjunto de datos

Etiquetado

El etiquetado fue realizado como consenso de tres diferentes observadores, quienes dividieron los 'tipos de visibilidad' en tres: visibilidad alta, media y baja. Las etiquetas se asignan de acuerdo con qué tanto es posible apreciar la silueta o contorno del edificio volcánico, incluyendo glaciar y tomando en cuenta la cantidad de cielo despejado entre cámara y volcán. Así, la Visibilidad alta, cuando es posible apreciar con claridad la silueta del edificio volcánico y gran parte del cielo se encuentra despejado entre la cámara y el volcán. La Visibilidad media se da cuando es posible apreciar parte de la silueta del edificio volcánico, además de que, parte del cielo se encuentra despejado entre la cámara y el volcán. Por su parte la Visibilidad baja se asigna a las imágenes que es imposible apreciar con claridad la silueta del edificio volcánico y gran parte del cielo se encuentra obstruido entre la cámara y el volcán.

Adicionalmente, se crearon seis etiquetas más: (i) Columna de gas (Figura 2a)) asociada a imágenes donde se aprecia la salida de gases desde el volcán, que se aprecian

principalmente de color blanco y que se interpreta como la liberación “sin mayor esfuerzo” de gases volcánicos; (ii) Erupción (Figura 2 b)), se refiere a una emisión violenta de material volcánico a la superficie que en las imágenes se puede distinguir por la presencia de una columna eruptiva con presencia de colores marrones o grisáceos propios de la ceniza y otros materiales volcánicos; (iii) Escarcha, se asigna a imágenes en las que se puede observar cristales de hielo sobre el lente de la cámara; (iv) Lluvia, se establece para imágenes en las que se es posible apreciar gotas de agua muy probablemente debidas a lluvias locales; (v) Incandescencias (resplandor nocturno), se refiere a imágenes adquiridas en horas de la noche en las que es posible apreciar un resplandor asociado a la salida de gases calientes o la ocurrencia de erupciones; y por último, (vi) Noche, se asigna a imágenes capturadas en horas de la noche.



a. b.

Figura 2 – Ejemplos de dos etiquetas. a) Columna de gas, b) Erupción

La recolección de imágenes pertenecientes a algunas etiquetas contempló el estudio de un intervalo de tiempo inferior a otras etiquetas, en especial erupciones e incandescencias, que, debido a la escasez, demandaron un lapso más largo del periodo de estudio (2015-2019).

Redimensionamiento de las imágenes

Para disminuir el costo computacional, las imágenes originales se redimensionan a un tamaño de 200 pixeles de alto, 200 de ancho y 3 de profundidad o espacio color [200, 200, 3]. Este parámetro se aplicó directamente a todo el conjunto de datos.

Partición del conjunto de datos

El conjunto de datos en general, se fragmentó en dos conjuntos: conjunto de entrenamiento con 70% de las imágenes y conjunto de prueba con 30% de las imágenes.

Extracción de características

Se creó un vocabulario visual de 500 palabras usando una función extractora de características personalizadas, especificada como función extractora de características visuales propuesta por (Csurka, 2004) que codifica las características visuales de una

imagen utilizando el algoritmo de vecino más cercano y construye un histograma de características para cada imagen. La función luego incrementa los intervalos del histograma en función de la proximidad del descriptor a un centro de clúster particular. La longitud del histograma corresponde al número de palabras visuales que construyó el paquete de características. El histograma se convierte en un vector de características para cada imagen.

2.4. Modelado

Definición de Arquitecturas

Se establecieron dos arquitecturas para la clasificación: una lineal donde todas las imágenes están a una misma fase, un mismo nivel, y la segunda, ensamble de clasificadores en cascada, donde diferentes clasificadores pueden discriminar grupos de imágenes a través de diferentes fases, esto debido a los requerimientos del SGC lo cual también permite realizar la comparación y encontrar el de mejor desempeño.

Las etiquetas asignadas a las imágenes se ordenan de forma abreviada de la siguiente manera: Noche (Noc), Incandescencia (Inc), Visibilidad Alta (V.A), Visibilidad Media (V.M), Visibilidad Baja (V.B), Columna de Gas (C.G), Erupción (Er), Lluvia (Llu), y Escarcha (Esc).

Arquitectura lineal

La primera arquitectura se organizó en un conjunto origen, que contiene todos los subconjuntos con las diferentes etiquetas (Figura 3) la cual dio inicio al diseño de otras arquitecturas.

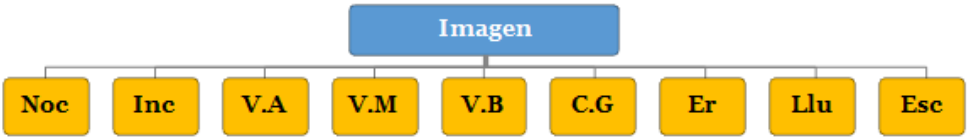


Figura 3 – Arquitectura de clasificación lineal

Ensamble de clasificadores en cascada

El ensamble de clasificadores contempla cinco fases (denotados en los cambios de color en la Figura 4). Este modelo busca como primera medida diferenciar imágenes diurnas de imágenes nocturnas (Fase 1); posteriormente, las imágenes son enviadas a una siguiente fase de acuerdo con su naturaleza. En el caso de noche busca asignar la etiqueta incandescencia o normal, para el día entre los tipos de visibilidad (Fase 2) y según su caso encontrar su contenido, columna de gas, erupción, Normal (Fases 3 y 4) y en el caso de baja visibilidad, se agregan lluvia y escarcha (Fase 5).

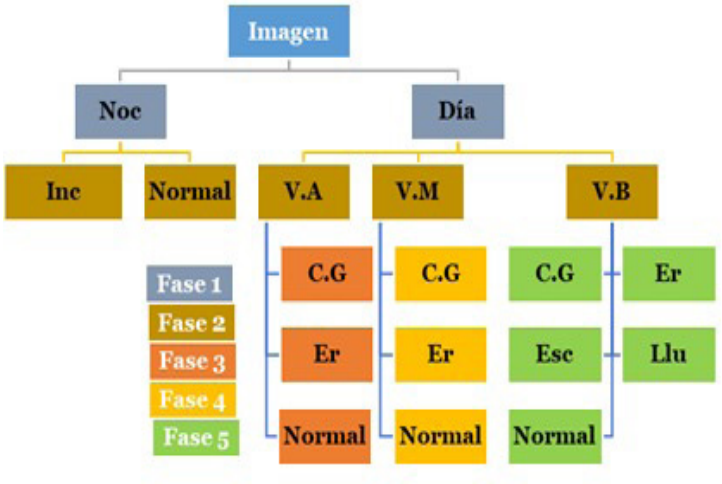


Figura 4 – Ensamble de clasificadores en cascada

2.5. Evaluación

Para el proceso de evaluación de desempeño, se usaron: la matriz de confusión que sirve para mostrar de forma explícita cuándo una clase es confundida con otra (Alegre, Pajares & De la escalera, 2016) y los indicadores derivados de ella como: su exactitud, que muestra la proporción entre las predicciones correctas que ha hecho el modelo y el total de predicciones con el conjunto de prueba. Los indicadores de rendimiento usados en el proceso de evaluación fueron: Exactitud, Precisión, Sensibilidad, Especificidad, Puntuación F1 y Coeficiente de correlación de Matthews CCM (Liu, Cheng, Yan & Chen, 2015).

3. Resultados

3.1. Construcción del conjunto de datos

Dentro de la construcción del conjunto de datos se obtuvieron un total de 18.849 imágenes, las cuales se organizaron en nueve etiquetas posibles (Tabla 1).

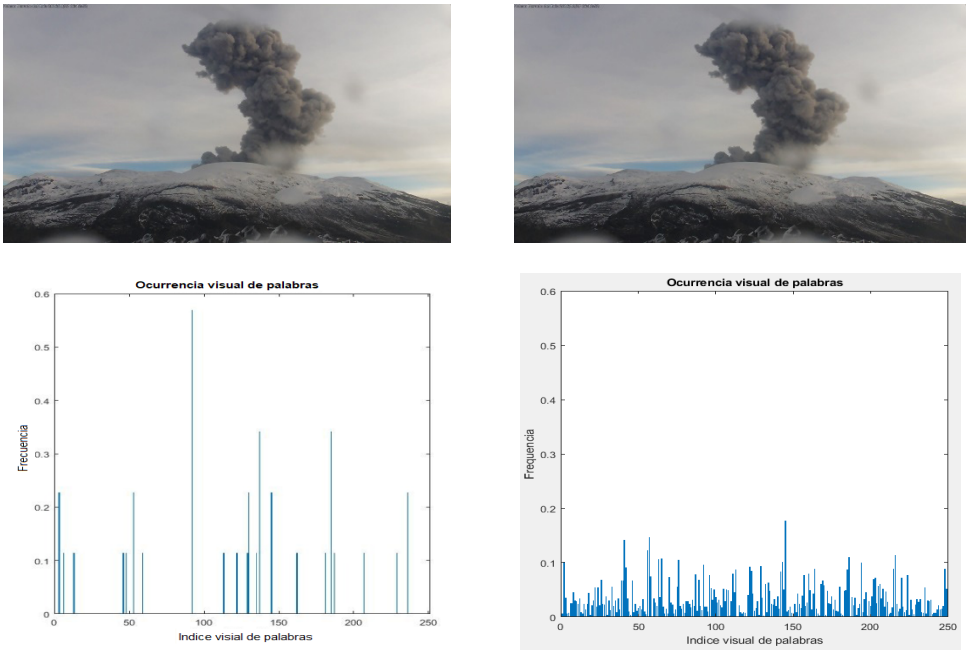
Tipo de imagen	Etiquetas	No. imágenes
Columna de gas	CG	2501
Erupción	ER	1183
Escarcha	ESC	2140
Incandescencia	INC	1808
Lluvia	LLU	2068
Noche	NOC	3422

Tipo de imagen	Etiquetas	No. imágenes
Visibilidad Alta	VA	2192
Visibilidad Baja	VB	1618
Visibilidad Media	VM	1917
Total:		18849

Tabla 1 – Conjunto de datos organizado por etiquetas

3.2. Preparación del conjunto de datos

Para probar cada arquitectura del sistema, se dividió, aleatoriamente las imágenes, en dos conjuntos: conjunto de entrenamiento (70%) y conjunto evaluación (30%) y se eligió el tipo de paquete de características de 500 palabras visuales (Figura 5).



a. b.

Figura 5 – a) Vocabulario visual 250 palabras en una imagen de erupción, b) Vocabulario visual 500 palabras en escala de 250 en una imagen de erupción

3.3. Modelado

Arquitectura lineal

Los clasificadores con mejor desempeño en la arquitectura lineal fueron las máquinas de soporte vectorial 'SVM' (Farhat, 1992) y el agrupamiento usando los k vecinos más

próximos 'KNN' (Hechenbichler & Schliep, 2004) (Tabla 2). Con desempeño por clase muy parecido para cada una de las etiquetas.

Clasificador	Exactitud (%)
SVM	89.8
KNN (k=3)	88.7
KNN (k=5)	86.6

Tabla 2 – Resultado del desempeño de los tres mejores clasificadores

Ensamble de clasificadores en cascada

En esta arquitectura, se realizó un proceso similar a la arquitectura lineal con la misma cantidad de clasificadores, resultando en un mejor desempeño. Esta arquitectura consiste en un clasificador para cada fase, encargado de realizar la discriminación y asignación de etiquetas. La Tabla 3 muestra el tipo de clasificador, las etiquetas, el rendimiento de cada fase por separado y el promedio de ella.

Fase	Arquitectura 2		
	Clasificador	Etiquetas	Exactitud
1	SVM	DÍA	100
		NOCHE	
2	SVM	INC	99.5
		NOC_NOR	
	KNN	VA	89.1
		VB	
		VM	
3	SVM	VA_CG	89.6
		VM_ER	
		VM_NOR	
4	SVM	VM_CG	91.2
		VM_ER	
		VM_NOR	
5	SVM	ESC	90.8
		LLU	
		VB_CG	
		VB_ER	
		VB_NOR	
Promedio			93.2

Tabla 3 – Visualización del desempeño del clasificador según su fase

3.4. Evaluación

A continuación se presentan los resultados alcanzados en la discriminación de las diferentes imágenes, para cada una de las arquitecturas planteadas. Matriz de confusión e indicadores de rendimiento calculados a partir de ella, para la arquitectura lineal, e indicadores de rendimiento para el ensamble de clasificadores en cascada.

Matriz de confusión

En primer lugar, para la arquitectura lineal, la matriz de confusión (Figura 6) muestra que las etiquetas con menos errores, mayor potencia discriminadora, son: las de noche, incandescencia, lluvia y escarcha, con seis, diez, veintidós y veinticinco errores en clasificación, respectivamente. Se observa a su vez que la etiqueta donde más errores se encuentra es la de erupción, 213 errores totales, la cual clasifica erróneamente 140 imágenes como columna de gas, de igual manera en columna de gas, el clasificador se equivoca en 93 imágenes clasificándolas como erupción, evidenciando la similitud que manejan estas dos etiquetas. Se observa un comportamiento similar entre los tipos de visibilidad, ejemplo de ello es la visibilidad media que confunde 118 veces con la visibilidad baja y a su vez la visibilidad baja es confundida en 80 imágenes como visibilidad media.

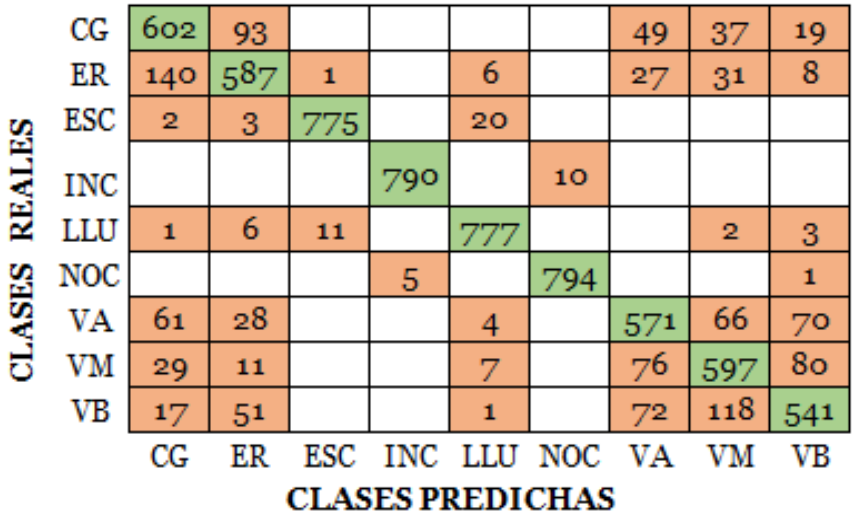


Figura 6 – Matriz de confusión para la arquitectura lineal

Indicadores de rendimiento

La Tabla 4 muestra las métricas de rendimiento para la arquitectura lineal. Incandescencia es la etiqueta que posee los valores más altos de todos los indicadores. Se observa que las etiquetas más exactas y precisas son las de escarcha, lluvia, noche e incandescencia, las otras etiquetas bajan un poco el valor respecto a ellas. En cuanto a la exactitud

todas las etiquetas superan satisfactoriamente y con creces el valor mínimo establecido inicialmente en 75%.

Indicador\ Clase	CG	ER	ESC	INC	LLU	NOC	VA	VB	VM	Desviación
Sensibilidad	0.753	0.734	0.969	0.988	0.971	0.993	0.714	0.746	0.676	0.137
Precisión	0.707	0.754	0.985	0.994	0.953	0.988	0.718	0.702	0.749	0.135
Especificidad	0.961	0.970	0.998	0.999	0.994	0.998	0.965	0.960	0.972	0.017
Puntuación 1	0.729	0.744	0.977	0.991	0.962	0.990	0.716	0.723	0.711	0.135
C. Matthews	0.694	0.712	0.974	0.989	0.958	0.989	0.681	0.688	0.678	0.152

Exactitud promedio = 0.898

Tabla 4 – Indicadores de rendimiento arquitectura lineal

Respecto a la arquitectura en cascada, la Tabla 5 evidencia las métricas de la fase 5 Los indicadores de exactitud se muestran en la Figura 7.

Indicador\Clase	ESC	LLU	VB CG	VB ER	VB NOR	Desviación
Sensibilidad	0.896	0.927	0.958	0.802	0.750	0.088
Precisión	0.966	0.881	0.829	0.794	0.878	0.065
Especificidad	0.992	0.969	0.951	0.948	0.974	0.018
Puntuación 1	0.930	0.904	0.889	0.798	0.809	0.059
C. Matthews	0.914	0.879	0.862	0.747	0.769	0.072

Exactitud promedio = 0.908

Tabla 5 – Indicadores de la fase 5, visibilidad baja

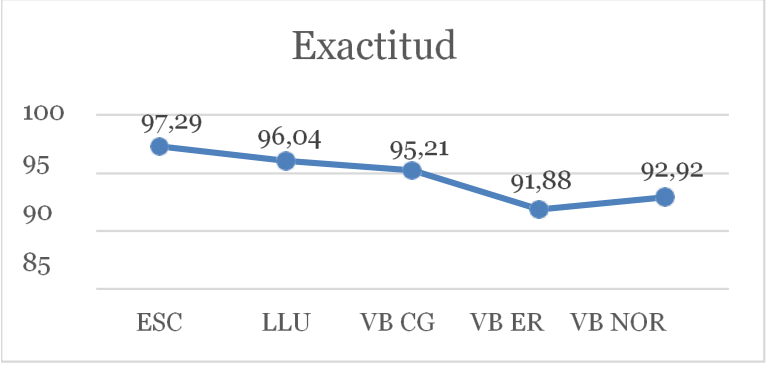


Figura 7 – Exactitud de visibilidad baja en la fase cinco del ensamble de clasificadores

4. Discusión

En este trabajo se implementaron dos arquitecturas de clasificación y se obtuvieron índices de desempeño de 89.8% y 93.2%, exactitud promedio, para las arquitecturas lineal y ensamble de clasificadores en cascada, respectivamente. De acuerdo con los trabajos relacionados en la literatura: Lakiy (Cadena & Muñoz, 2018) alcanzó resultados de 92% usando señales y RNA; Tag EQ (León, Ordoñez & Bucheli, 2019), utilizando los clasificadores SVM y bosques de decisión presentaron resultados con una precisión por encima del 80%. Por su parte, (Witsil & Johnson, 2020) quienes usaron imágenes, clasificaron únicamente tres etiquetas asociadas a la actividad volcánica: resplandor nocturno, cobertura de nubes y emisiones no perceptibles durante condiciones sin nubes, obtuvieron una precisión general del 93%, lo cual es superado por el modelo de ensamble de clasificadores, aquí presentado, que clasifica nueve clases con una precisión general del 93.2%.

En la arquitectura lineal el clasificador tiene la tarea de identificar entre las nueve etiquetas disponibles para una sola fase, el SVM presento una exactitud del 89.8%. Esta arquitectura se considera como primer punto de partida, línea de base, del trabajo, puesto que, en un inicio, la idea más inmediata que se planteó fue realizar la clasificación de imágenes a partir de un único entrenamiento y esperar resultados, pero se evidenció un menor desempeño respecto el ensamble en cascada. El ensamble en cascada presentó un mayor rendimiento en el sistema de clasificación de monitoreo volcánico, la arquitectura manejó una combinación de dos clasificadores SVM, KNN dependiendo de la fase en que se encontraban, la fase uno de mayor rendimiento arrojó una clasificación del 100% para distinción de noche y día con un clasificador SVM. En la fase dos las imágenes nocturnas, donde para los indicadores los valores son prácticamente iguales entre noche normal e incandescencia, su porcentaje de acierto del 99.5% es casi perfecto para ambas etiquetas acompañado de un clasificador SVM. A comparación de (Witsil, 2019) el resplandor nocturno (incandescencia en este trabajo) es la forma más común de actividad volcánica de su conjunto de datos etiquetado y está clasificado por la ANN con precisión 96% y recuperación 96%.

En el ensamble cascada la fase dos también trata sobre los tipos de visibilidad, donde se presentan los casos de nubosidad los porcentajes de acierto son del 89.1% con un clasificador tipo KNN. Para (Witsil. 2019) la identificación de la nubosidad es menos precisa (62% de recuperación y 70% de precisión) y se confunde con mayor frecuencia con períodos de inactividad. En el caso de las fases 3, 4 y 5 el ensamble en cascada evidencia que los clasificadores con rendimientos más altos fueron tipo SVM con porcentaje de aciertos de exactitud del 89.6%, 91.2% y 90.8% respectivamente y dentro de ello por cada fase las subcarpetas con mayor exactitud fueron visibilidad alta-normal, visibilidad media-erupción y visibilidad baja-escarcha respectivamente.

5. Conclusiones

Se implementó un sistema de clasificación de imágenes de cámaras de vigilancia volcánica basado en procesamiento digital de imágenes y aprendizaje de máquina, la extracción de características se realizó a través del paquete Bagoffeatures de MATLAB. Adicionalmente se compararon diferentes clasificadores y se seleccionó el de mayor

rendimiento, que para el caso de la arquitectura lineal fue un SVM con una exactitud de 89.8%, y para el ensamble en cascada una combinación de SVM y KNN con una exactitud promedio de 93.2%. En ambos casos se superó el 75% esperado al inicio del trabajo.

Los resultados encontrados en este trabajo demuestran que las técnicas implementadas pueden ser aplicadas como apoyo a las labores de clasificación automática y/o asistida de imágenes logrados de cámaras de vigilancia volcánica, para los OVS's de Colombia. Para mejorar este trabajo, se recomienda incrementar el conjunto de datos en la etapa de entrenamiento y refinar la metodología de etiquetado de los datos. Adicionalmente, con una mayor cantidad de datos correctamente etiquetados, se puede experimentar la clasificación haciendo uso de técnicas de aprendizaje profundo.

De acuerdo a los resultados obtenidos se recomienda explorar con algoritmos como árboles de decisión, discriminantes lineales y cuadráticas, o redes neuronales que permitan que los clasificadores a utilizar en los procesos puedan presentar mejoras, además tener una realimentación y estén en constante reentrenamiento de forma automática.

Agradecimientos

Este trabajo fue posible gracias al OVSPo quienes proporcionan la base de datos de las cámaras web, al programa de estado joven del ministerio de trabajo de la práctica laboral ordinaria número 220355-9 resolución 284 que financiaron, parcialmente, este proyecto, y a la Universidad del Cauca.

Referencias

Alegre E., Gonzales P. & Escalera A. (2016.) Conceptos y Métodos en Visión por Computador. España.

Asim K, Moustafa S, Niaz I, Elawadi E, Iqbal T & Martínez Á. (2019). "Seismicity analysis and machine learning models for short-term low magnitude seismic activity predictions in Cyprus," Soil Dynamics and. Earthquake. Engineering, 130. https://doi.org/10.1016/j.soildyn.2019.105932.

Cadena O. & Meneses J. (2018). "LAKIY, un sistema práctico, automático-supervisado, basado en RNA para el procesamiento primario de sismicidad volcánica en tiempo real". Retrieved from: https://cutt.ly/rd2TQuJ.

Csurka G., Dance C., Fan L., Willamowski J. & Bray. (2004). Visual Categorization with Bags of Keypoints. Work Stat Learn Comput Vision, ECCV. 1.

Ferrés D., Escobar R., & Bowman L. (2015). Reflexiones de los volcanólogos sobre su papel profesional en la gestión del riesgo volcánico en el contexto de América Latina. Revista Geológica de América Central, 52(1), 9-31.

Farhat N. (1992). "Photonit neural networks and learning mathines the role of electron-trapping materials," *IEEE Expert*. 7(5), 63–72.

García C. (2019) “Metodologías de desarrollo de proyectos de minería de datos - Una visión centrada en CRISP-DM,” Retrieved from: https://cutt.ly/Nd2Ym04.

Hechenbichler K. & Schliep K. (*2004).* “Weighted k-Nearest Neighbors Techniques and Ordinal Classification” https://doi.org/10.5282/ubm/epub.1769.

Hermelin, M., (2005). Desastres naturales en Colombia 1979-2004. Medellín, Colombia. Sello Editorial EAFIT.

Liu Y, Cheng J, Yan C, Wu X and Chen F. (2015). “Research on the Matthews correlation coefficients metrics of personalized recommendation algorithm evaluation” *International Journal of Hybrid Information Technology*, 8(1), 163–172. https://doi.org/ 10.14257/ijhit.2015.8.1.14.

Ospina C. & Narváez A. (in press). VolCAM, a desktop application for volcanic monitoring using surveillance cameras. Management Review, (Prevision Screen –August/2020)

Revista Ibérica de Sistemas e Tecnologias de Informação; Lousada Iss. E27, (Mar 2020): 359-370.

Romano, L.E. (2019). 14 Observaciones que surgen del reciente desastre en el volcán de Fuego, 2018, Guatemala. Revista de estudios Latinoamericanos sobre Reducción del Riesgo de Desastres REDER, 3(1), 109-112.

Samaniego P., Pennec J., Eissen J., Hall M., Molina. I., Robin C., Mothes P., Yepes H., Ramóm P., Monzier M. & Egred J. (2003). Los peligros volcánicos asociados con el Tungurahua. Retrieved from: https://cutt.ly/nd2U7mn.

Servicio Geológico Colombiano. (2020). Observatorios Vulcanológicos. Colombia. Retrieved from: https://cutt.ly/Yd2I0Ex

Witsil A. & Johnson J., (2020) Volcano video data characterized and classified using computer vision and machine learning algorithms. Geoscience Frontiers. doi:10.1016/j.gsf.2020.01.016.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 11/08/2020
Aceitação/Acceptance: 03/10/2020

Aplicación de métodos lineales de reducción de dimensión y redes neuronales convolucionales para la predicción de estructura secundaria de proteínas

Diego Fernando Peña[1], Néstor Díaz[1], Víctor Andrés Bucheli[2]

dfepena@unicauca.edu.co; nediaz@unicauca.edu.co; victor.bucheli@correounivalle.edu.co

[1]Universidad del Cauca, Calle 5 # 4-70, 190003, Popayán, Colombia.

[2]Universidad del Valle, Calle 13 # 100-00, 76001, Cali, Colombia.

Pages: 199-213

Resumen: Deducir la estructura secundaria de una proteína a partir de su cadena de aminoácidos es una importante etapa de clasificación intermedia que es utilizada para obtener estructuras nativas con un mayor grado de precisión. Si bien actualmente los modelos de aprendizaje profundo han demostrado tener los mejores resultados de predicción, estos denotan un elevado costo computacional debido al gran número de parámetros que los definen. En este sentido, surge la necesidad de buscar métodos que permitan la implementación de modelos más simples que puedan superar o igualar de forma eficiente el desempeño de arquitecturas robustas. En este artículo se realiza la predicción de la estructura secundaria a través de redes neuronales convolucionales donde se demuestra que la compresión del espacio de alta dimensión original a un conjunto de características más compactas puede mejorar el proceso de aprendizaje de una red neuronal convolucional con un mayor número de parámetros.

Palabras-clave: Análisis de Componentes Principales; Análisis Discriminante Lineal; Estructura Secundaria; Redes Neuronales Convolucionales; Reducción de Dimensión.

Application of Linear dimensionality reduction techniques and convolutional neural networks for protein secondary structure prediction

Abstract: The deduction of protein secondary structure only with its amino acid chain depicts an important intermediate stage used to obtain native protein structures with higher accuracy. In spite of current models based on deep learning have presented the best results, they are defined by an important number of parameters which is reflected in a considerable computational cost. In this sense, there is a need for searching methods for the implementation of simpler models which can reach or improve the performance of robust architectures. This paper uses one-dimensional convolutional neural networks to accomplish protein secondary

structure prediction where is possible to demote that embed the high dimensional data into more compact features can improve or match the learning process in a convolutional neural network with a greater number of inputs and parameters.

Keywords: Convolutional Neural Networks; Dimensionality reduction; Linear Discriminant Analysis; Principal component analysis; Secondary structure prediction.

1. Introducción

Las proteínas son constituyentes primordiales de todos los organismos vivos debido a su capacidad para ejecutar diversas funciones biológicas fundamentales (Khoury et al., 2014). La gran variedad de estas biomoléculas y sus funciones está directamente relacionada con las diferentes variaciones en las secuencias de aminoácidos que se pueden presentar, además de diversos factores estructurales que han sido divididos en cuatro niveles organizacionales (estructura primaria, secundaria, terciaria y cuaternaria) (Jiang et al., 2017). No obstante, identificar la estructura terciaria (estructura nativa) brinda información importante acerca de su función, lo que podría ser de gran ayuda para el avance de estudios científicos que puedan aportar un mejor entendimiento de procesos biológicos, químicos y patológicos o en la implementación de nuevas herramientas y aplicaciones médicas como vacunas (Dunitz & Scheraga, 2004; Khoury et al., 2014). A pesar de que existen métodos experimentales que permiten determinar las estructuras terciarias de las proteínas como, por ejemplo, la cristalografía de rayos X y la resonancia magnética nuclear de proteínas (RMN), la aplicación de estas técnicas supone una inversión considerable de tiempo y dinero. Esta situación ha representado un gran problema debido a que se ha creado una brecha entre el número de secuencias de proteínas descubiertas y el número de estructuras 3D definidas (Khoury et al., 2014). Dado el gran avance en los sistemas informáticos uno de los principales enfoques para intentar resolver dicho problema se ha basado en la aplicación de algoritmos de aprendizaje de máquina que puedan generar estructuras nativas a partir de secuencias de aminoácidos.

La predicción de la estructura secundaria puede definirse como un puente entre la estructura primaria y la estructura terciaria, y denota la conformación local en el espacio de la cadena de aminoácidos (Jiang et al., 2017). Tradicionalmente la estructura secundaria se representa con tres conformaciones estándar, sin embargo, en el presente trabajo se tendrá en cuenta la expansión realizada en (Kabsch & Sander, 1983) donde se proponen ocho clases objetivo: tres tipos de hélices (G para hélice, H para hélices-α, I para hélices-π), dos tipos de Lámina (E para lamina-β extendida y B para puente-β) y tres tipos de giro (T giro con enlace de hidrogeno, S para laso de alta curvatura y L para giro irregular).

Si bien para intentar resolver este problema se ha recurrido a enfoques clásicos de aprendizaje de máquina (Jiang et al., 2017), en la actualidad son los modelos de aprendizaje profundo los que han presentado los mejores resultados especialmente para el problema donde se tienen en cuenta ocho clases de estructura secundaria. Dentro de las arquitecturas más recientes se destaca la aplicación de redes neuronales convolucionales (Angioloni, 2019; Fang et al., 2018; S. Wang et al., 2016), y la integración de estas mismas

con otro tipo de redes neuronales como las recurrentes (Guo et al., 2018, 2019; Kumar et al., 2020; Z. Li & Yu, 2016; Zhang et al., 2018). Si bien los anteriores modelos han obtenido algunos de los mejores resultados, estos enfoques denotan arquitecturas con un número elevado de parámetros, sin contar que algunos de ellos (Fang et al., 2018; Guo et al., 2019; Z. Li & Yu, 2016), incrementan el número de características a procesar. Esto representa un problema reflejado en el costo computacional y en el número de parámetros a optimizar.

En este trabajo se retoma el modelo de clasificación propuesto en (Angioloni, 2019), el cual está basado en una red neuronal convolucional (RNC) y representa un enfoque alternativo más sencillo de predicción de estructura secundaria que brinda resultados cercanos a modelos más complejos como el descrito en (S. Wang et al., 2016). El principal aporte presentado en este documento es demostrar que el uso de métodos de reducción de dimensión (RD) lineales como análisis de componentes principales (PCA) y análisis discriminante lineal (LDA) (Jenike & Albert, 2009), pueden igualar o mejorar significativamente el proceso de aprendizaje de la RNC entrenada con un número mayor de características. Adicionalmente, se realizó un ventaneo alternativo al propuesto en (Angioloni, 2019), así como un estudio más profundo de algunos hiperparámetros a utilizar y la adición de la etiqueta perteneciente al tipo de aminoácido como característica de entrada a la RNC.

El resto de este trabajo está dividido en 5 secciones principales. En la Sección 2 se describe el conjunto de datos utilizado, así como su procesamiento. En la siguiente sección, se habla sobre la aplicación de métodos de RD como funciones de proyección. Posteriormente en la sección 4 se describe en detalle el modelo de predicción utilizado y sus diferentes parámetros. La sección 5 establece la descripción e interpretación de los resultados obtenidos. Finalmente, en la sección 6 se realizan unos comentarios finales a manera de conclusiones y trabajo futuro.

2. Procesamiento de datos

Para el desarrollo de este trabajo se usaron cuatro conjuntos de datos públicos: **1)** CB6133 obtenido del repositorio PISCES (G. Wang & Dunbrack, 2003), **2)** CB513 (Cuff & Barton, 1999), **3)** CASP10 , y **4)** CASP11 (Moult et al., 2014). El CB6133 cuenta con secuencias no homólogas que son utilizadas únicamente para procesos de entrenamiento y validación puesto que contienen un importante número de proteínas (6133). Sin embargo, debido a que existe redundancia entre los conjuntos de datos CB6133 y CB513, se hace uso de una versión filtrada de CB6133 eliminando las secuencias que tengan una similitud superior al 25% (S. Wang et al., 2016; Zhou & Troyanskaya, 2014), lo cual garantiza una mayor independencia de los conjuntos de datos de entrenamiento y prueba.

Adicionalmente, para llevar a cabo un redimensionamiento de los datos a través de un proceso de ventaneo se eliminaron las secuencias con un número de aminoácidos menor a 20. Como resultado de esta etapa de filtrado el conjunto de datos CullPDB fue reducido a 5527 proteínas, donde 5250 proteínas se utilizan para entrenamiento y 277 proteínas para validación. Finalmente, los conjuntos de datos **CB513** (513 proteínas), **CASP10** (123 dominios de proteínas) y **CASP11** (105 dominios de proteínas), fueron utilizados como datos de prueba.

2.1. Características del conjunto de datos

Los datos de cada conjunto fueron generados bajo las mismas condiciones y por cada proteína se generó una Matriz de Puntuación de Posición Específica (PSSM) (D. Li et al., 2012), utilizando Uniref90 como base de datos para obtener información de referencia (Suzek et al., 2015). Las puntuaciones de cada aminoácido en su respectiva PSSM fue estandarizada mediante una función sigmoide para obtener valores en el rango de [0, 1]. Las clases para cada aminoácido fue obtenido a partir de la estructura nativa de cada proteína. A diferencia de (Angioloni, 2019), el número de características que alimenta la RNC está conformado por el tipo de aminoácido (una característica en formato "One Hot Encoding") y los perfiles de las proteínas creados a partir de la PSSM (21 perfiles). Por consiguiente, el modelo sin ninguna clase de procesamiento básicamente tendrá un total de 22 características ($D = 22$). En este sentido la proyección y la reducción de dimensión de los datos se realizará específicamente en los perfiles de las proteínas como se indica en la Figura 2. Además, en este trabajo se decidió utilizar información topológica de las secuencias como longitud, posición del aminoácido inicial y final para dividir aleatoriamente los datos de entrenamiento y de validación teniendo en cuenta proteínas completas y no aminoácidos individuales. Por lo tanto, los porcentajes de entrenamiento y validación que se mencionan en el documento hacen referencia a un número de proteínas en específico.

2.2. Ventaneo de las secuencias

Con el fin de brindar al modelo de predicción una mayor posibilidad de encontrar patrones químicos presentes en cada una de las secuencias se realiza un proceso de ventaneo adicional al que hace internamente la RNC (Angioloni, 2019). Esto principalmente con el objetivo de obtener una mejor perspectiva de la influencia de los residuos presentes en la proteína para la definición de estructuras secundarias. De este modo, a partir de una

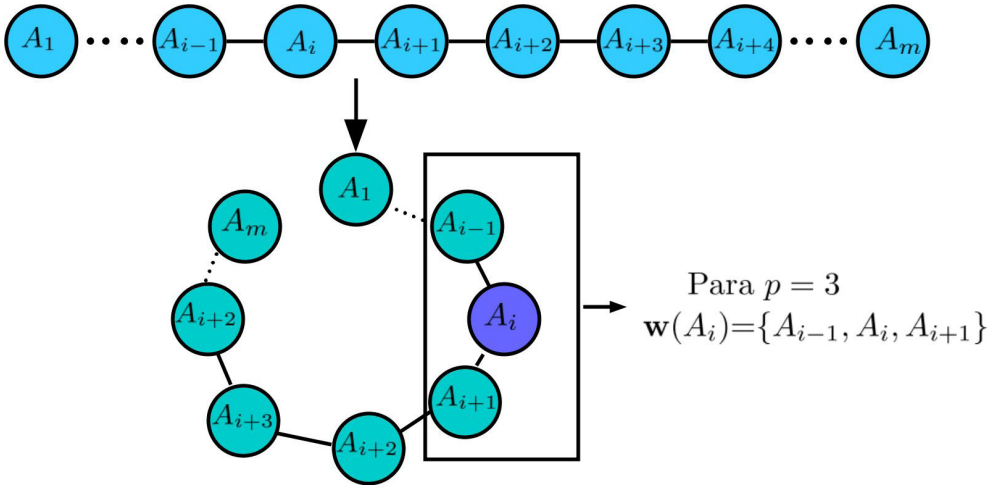


Figura 1 – Diagrama del ventaneo utilizado en una secuencia de aminoácidos de tamaño m, para una ventana w de longitud $p = 3$

secuencia S con una longitud de m aminoácidos donde $m \in \mathbf{Z}$, se realizará un proceso de ventaneo considerando a la secuencia como un anillo continuo (Figura 1). Con el fin de ventanear correctamente los residuos ubicados en los extremos de la cadena debido al desplazamiento de una ventana $\mathbf{w}$ de tamaño p con $p \in \mathbf{Z}$, a lo largo de toda la cadena. En este sentido, el residuo ubicado en el centro de la ventana será el que se va a predecir.

Una vez concluido el proceso de ventaneo se tendrá una matriz de datos tridimensional $\mathbf{X}_{m \times p \times D}$ con $m, p, D \in \mathbf{Z}$, donde m representa el número de aminoácidos, p representa el tamaño de agrupación de la ventana y D representa el número de características que se utilizan como entradas al modelo de predicción**.**

3. Proyección de los datos y reducción de dimensión

Un método de reducción de dimensión puede verse como una función $G(\mathbf{X}_{m \times D})$ que es capaz de obtener una matriz $\mathbf{Y}_{m \times d}$ con $d \in \mathbf{Z}$, que representa un espacio de características reducidas creado a partir de un conjunto de datos de entrada $\mathbf{X}_{m \times D}$ con $(d < D)$, en otras palabras:

$$\mathbf{Y}_{m \times d} = G(\mathbf{X}_{m \times D}), \qquad con\ d < D. \tag{1}$$

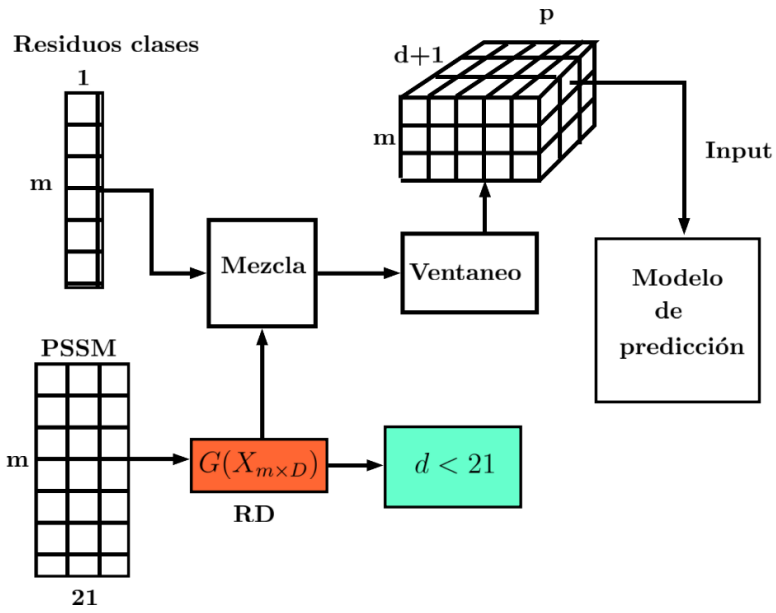


Figura 2 – En este diagrama se representa gráficamente el procesamiento del conjunto de datos de entrada que ingresan al modelo de predicción basado en RNCs. El bloque de RD mapea los perfiles a un espacio más compacto, y se procesa en un bloque de mezcla donde se integra el espacio embebido y el tipo de aminoácido, para finalmente crear vecindades a través del bloque de ventaneo

Donde las dos principales ventajas de $G(\mathbf{X}_{m\times D})$ radican en su capacidad para ser creada a partir de los datos de entrenamiento y en la proyección de nuevos puntos (datos de prueba) al espacio de baja dimensión. Lo anterior abre la posibilidad de que el espacio embebido generado con $G(\mathbf{X}_{m\times D})$, pueda resaltar patrones similares en los datos de entrenamiento y prueba, que probablemente no eran fácilmente diferenciables en el espacio original de alta dimensión. De este modo, se mejoraría el proceso de aprendizaje de los sistemas de clasificación supervisado.

La etapa anteriormente descrita corresponde al bloque resaltado en rojo en la Figura 2 y representa la principal diferencia de trabajos similares como (Angioloni, 2019; S. Wang et al., 2017), los cuales no poseen dicho bloque de procesamiento reflejando así un mayor número de parámetros a ser entrenados.

4. Esquema del modelo de predicción

La arquitectura del modelo de aprendizaje profundo para clasificación de estructura secundaria a partir de secuencias de proteína se muestra en la Figura 3. El modelo utilizado está compuesto de 15 capas, incluida la capa de entrada. La sección convolucional de la arquitectura se encarga del proceso de aprendizaje de características utilizando tres módulos compuestos de una capa convolucional unidimensional, una capa de normalización y una capa de "Dropout" que evita el sobreajuste durante la fase de entrenamiento. La fase de clasificación se lleva a cabo con tres capas densas completamente conectadas que reciben las características procesadas por la capa de aplanado y determina la clase a la que corresponde cada aminoácido en la secuencia de la proteína de entrada. En todas las capas intermedias se usó "relu" como función

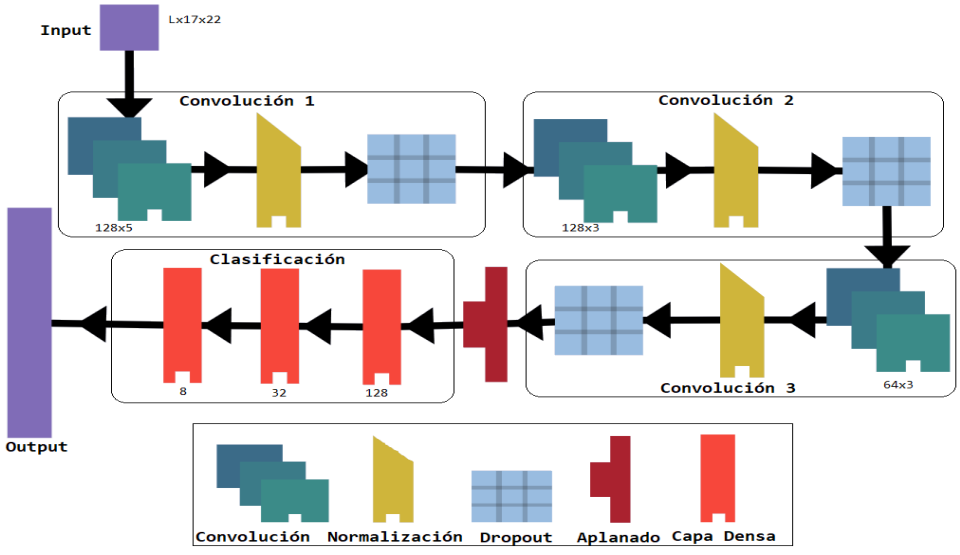


Figura 3 – Arquitectura del modelo propuesto para clasificación de estructura secundaria de proteínas

de activación, mientras que en la capa de salida se utilizó la función sigmoide para determinar la clase de estructura secundaria más probable.

4.1. Selección de hiperparámetros

Teniendo en cuenta que se modificó en cierta medida el enfoque de ventaneo propuesto por (Angioloni, 2019), se realizó una modificación de tres hiperparámetros (Dropout, optimizador y tamaño del batch) de acuerdo a su desempeño con los datos de validación, así como la capacidad de generalización con los datos de prueba (Figura 4). Para la selección de la ventana se utilizaron el 95% de todas las proteínas de CB6133 para entrenamiento y el 5% restante para validación, mientras que para definir la tasa de "Dropout", el optimizador y el tamaño del "batch" se usó el 50% de todas las proteínas de CB6133, respetando la misma división anteriormente descrita (95% y 5%).

En primera instancia se modificó el tamaño de la ventana como se muestra en la Figura 4A, donde se puede observar que los mejores resultados se obtienen con las ventanas de 13 y 17 aminoácidos. No obstante, al no tener una diferencia de costo computacional muy marcada se decidió establecer un valor de $p = 17$ que es el mismo utilizado en (Angioloni, 2019). Una vez definida la ventana se empezó a variar el valor del "Dropout"

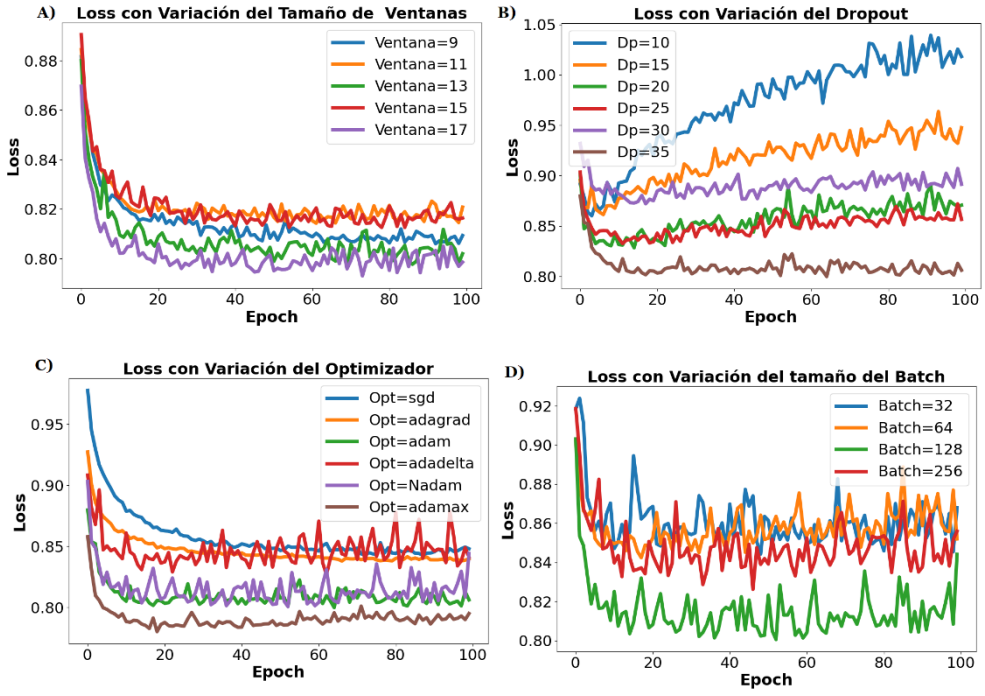


Figura 4 – Afinación de hiperparámetros. En las figuras se indican las diferentes pruebas realizadas que permitieron establecer los hiperparámetros que serán utilizados dentro del marco experimental de este trabajo. A) Variación del tamaño de la ventana; B) Variación de Dropout (Dp); C) Selección del optimizador (Opt) y D) Ajuste del tamaño del Batch

en un rango que va de 0.10 a 0.35 avanzando en pasos de 0.5. La Figura 4B muestra los resultados con los datos de validación, en donde el valor de 0.35 ofrece un mejor aprendizaje. Posteriormente, se probaron diferentes clases de optimizadores, donde la Figura 4C refleja que **Adamax** (Kingma & Ba, 2015), representaba la mejor opción para entrenar el modelo. Sin embargo, Los resultados en los datos de prueba no fueron satisfactorios y fueron superados por el optimizador **Nadam** (Dozat, 2016). Finalmente, con los anteriores hiperparámetros definidos se procedió al variar el tamaño del Batch (Figura 4D) en donde un valor de 128 presenta los mejores resultados.

Por otra parte, se varió el número de capas convolucionales (4, 5 y 6) buscando establecer una mejor arquitectura, sin embargo, no se presentaron cambios importantes en el desempeño de la red neuronal convolucional que justificaran un incremento de las capas convolucionales, teniendo en cuenta que esto implica un mayor costo computacional. Adicionalmente, a diferencia del modelo de (Angioloni, 2019), se incorporaron capas de pooling para probar su desempeño dentro del marco experimental propuesto, no obstante, las métricas de evaluación no alcanzaron los mejores resultados obtenidos.

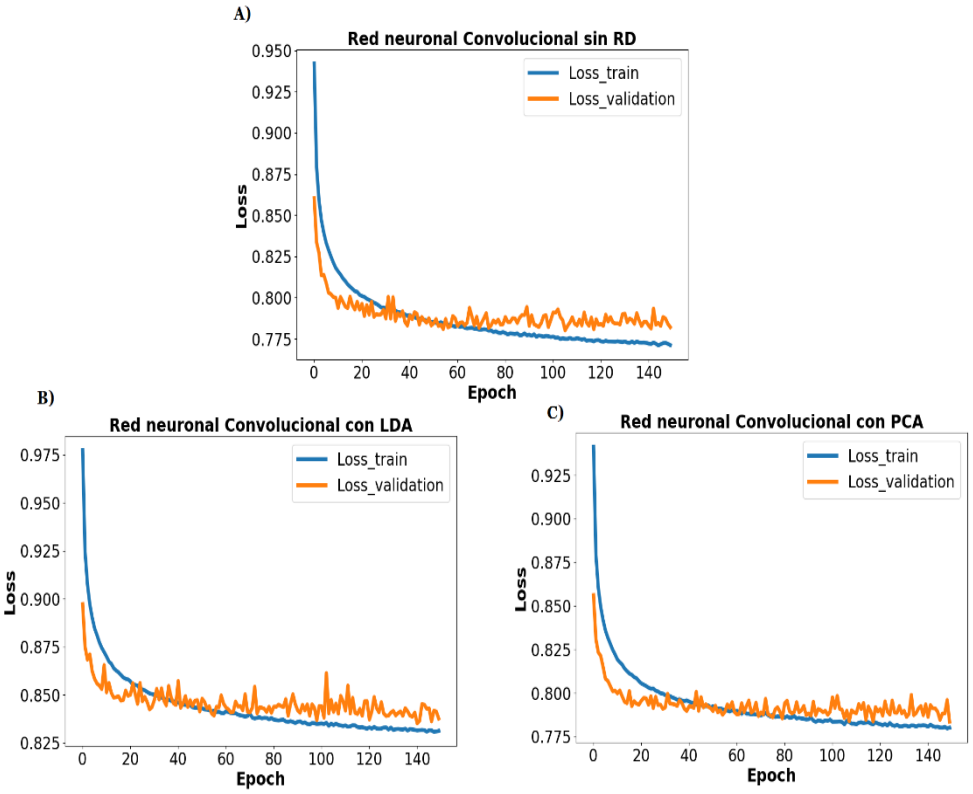


Figura 5 – Comparación del entrenamiento del modelo base vs modelos con técnicas de RD. A) Loss entrenamiento y validación del modelo CCN. B) Loss entrenamiento y validación modelo del CCN+LDA. C) Loss entrenamiento y validación del modelo CCN+PCA

Esto concuerda con (Fang et al., 2018), donde se menciona que el pooling afectará el desempeño de clasificación dado que las secuencias de proteínas tienen longitudes diferentes. Por lo tanto, se tomó la decisión de utilizar la arquitectura de (Angioloni, 2019), con tres capas convolucionales sin pooling. En cuanto, al número de épocas se puede observar en la Figura 5 que a partir de 100 épocas no existe mejoras sustanciales en el proceso de aprendizaje, ni en el valor de precisión para los datos de validación. Los mejores resultados para la tasa de aprendizaje, se presentaron para un valor de 0.001 en el algoritmo de optimización Nadam.

4.2. Análisis de componentes principales (PCA) y Análisis discriminante lineal (LDA)

En este trabajo la función $G(\boldsymbol{X}_{m\times D})$ vendrá dada por la transformación lineal definida en la Ecuación (2).

$$G(\mathbf{X}_{m\times D}) = \mathbf{Y}_{m\times d} = \mathbf{X}_{m\times D}\mathbf{P}_{D\times d} \tag{2}$$

Donde $\mathbf{Y}_{m\times d}$ representa la matriz de baja dimensión y $\mathbf{P}_{m\times d}$ la matriz de proyección la cual es generada a partir de los datos de entrenamiento (CB6133) y que varía de acuerdo al método de reducción de dimensión que se aplique. En el caso de PCA la matriz de proyección $\mathbf{P}_{\mathrm{PCA}}$ se obtendrá con los vectores propios ordenados de forma descendente con respecto al mayor valor propio de la matriz de covarianza (Jollife & Cadima, 2016). Por otro lado, la matriz de proyección $\mathbf{P}_{\mathrm{LDA}}$ asociada a LDA se obtiene con los vectores propios ordenados de acuerdo al mayor valor propio del producto matricial entre la matriz de dispersión inversa de la misma clase y la matriz de dispersión entre las clases (Jenike & Albert, 2009).

Para encontrar el número de componentes principales que mejor representan los datos originales se utilizó la Figura 6 que representa la varianza acumulada y explicativa de cada componente, como resultado se puede observar que la información de los 21 perfiles de las proteínas se encuentran mayoritariamente representadas en los 15 primeros componentes principales con el 94 % de la varianza total. Por su parte el número de dimensiones de la implementación de LDA utilizada se ve limitada por la Ecuación (3).

$$d_{max} = min(D, clases - 1) \tag{3}$$

Donde $min\,()$ denota el valor mínimo entre el número de dimensiones del espacio original D y el número de clases menos uno, de modo que el espacio embebido puede tener una dimensión máxima $d_{max} = 7$.

4.3. Medidas de calidad

Teniendo en cuenta que la base de datos presenta un claro desbalanceo, sobre todo en la clase **I** que cuenta únicamente con 212 instancias de datos en CB6133, se utilizaron varias métricas adicionales de evaluación como: **Recall, Precisión y F1-score**

(Tharwat, 2018), para obtener una perspectiva más amplia de la predicción realizada por cada clase haciendo uso de la matriz de confusión.

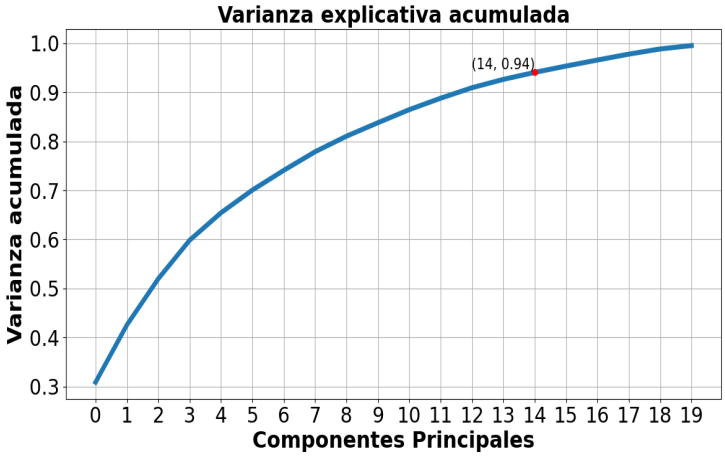


Figura 6 – En esta grafica se representa el valor de la varianza acumulada con respecto al número de componentes principales, donde los primeros 15 (0-14) componentes principales representan el 94 % de la varianza de los datos

5. Resultados

Una vez definidos los parámetros del modelo y los métodos de RD se obtuvieron los resultados presentados en la Figuras 5**,** que reflejan el comportamiento de la RNC cuando se mapean los datos de prueba a un espacio embebido creado a partir de los datos de entrenamiento. Al analizar las Figuras 5A**,** 5B y 5C es posible denotar que la aplicación de métodos de RD mejora el aprendizaje debido a la reducción del desfase existente entre la curva de entrenamiento y la curva de validación, previniendo de este modo el sobre ajuste.

En la Tabla 1 se puede observar una mejora significativa en la precisión promedio del modelo con PCA. Asimismo, teniendo en cuenta el principal aporte y objetivo de esta investigación, se verifica que a pesar que el modelo basado en (Angioloni, 2019) cuenta con un número de 245,352 parámetros, los mejores resultados dentro del marco experimental planteado se obtuvieron con un modelo definido con menos parámetros (241,512). Demostrando así la potencialidad de la aplicación de métodos de RD para abordar los problemas de predicción de la estructura secundaria con un menor número de datos de entrada (menor costo computacional).

En las Tablas 2, 3 y 4 se presentan los resultados de las métricas de evaluación de desempeño obtenidas a partir de la matriz de confusión de cada clase. Estas mediciones son importantes puesto que brindan un contexto más completo de evaluación para establecer un marco de comparación con modelos más complejos disponibles en el estado del arte (Z. Li & Yu, 2016; S. Wang et al., 2016).

En la Tabla 2 se muestran las medidas de calidad calculadas a nivel de clase usando el conjunto de datos de prueba **CB513**. En este conjunto de datos es importante resaltar que el modelo CNN+PCA presenta un mayor valor de precisión para todas las clases excepto L y T, además de tener un **F1-score** superior en la mayoría de las clases denotando que el modelo de características reducidas tiene un mejor balance entre precisión y recall.

Modelo \ Datos	Cb513 Q8	Casp10 Q8	Casp11 Q8	Número de Parámetros
LDA+CNN	64.8	69.2	66.1	236,392
CNN	67.1	70.0	67.7	245,352
PCA+CNN	67.2	71.0	68.2	241,512

Tabla 1 – En esta tabla se puede observar la precisión promedio de los conjuntos de datos utilizados, así como el número de parámetros a ser optimizados en cada modelo

SS8 Cb513	Recall			Precisión			F1-score		
	CNN	CNN LDA	CNN PCA	CNN	CNN LDA	CNN PCA	CNN	CNN LDA	CNN PCA
L	0.615	0.587	0.631	0.524	0.508	0.522	0.566	0.545	0.571
B	0.044	0.031	0.033	0.419	0.493	0.429	0.080	0.059	0.061
E	0.804	0.782	0.800	0.718	0.693	0.725	0.758	0.735	0.761
G	0.295	0.270	0.296	0.414	0.382	0.425	0.345	0.316	0.349
I	0	0	0	0	0	0	0	0	0
H	0.908	0.893	0.913	0.821	0.794	0.821	0.863	0.841	0.865
S	0.239	0.191	0.211	0.477	0.433	0.489	0.318	0.265	0.295
T	0.462	0.451	0.467	0.572	0.541	0.567	0.511	0.492	0.512

Tabla 2 – En esta tabla se muestran los resultados de aplicar las métricas de evaluación planteadas en la Sección 4.3 para cada una de las clases de estructura secundaria en CB513

La Tabla 3 muestra los resultados detallados por clase para el conjunto de prueba **CASP10**. Los conjuntos de datos CASP proporcionan un contexto similar a condiciones de predicción real, dado que contienen únicamente proteínas cuya estructura ha sido recientemente identificada por laboratorios independientes alrededor del mundo y por lo tanto no se encuentran en las bases tradicionales. Esto hace que la evaluación hecha con estos conjuntos de datos sea poco susceptible a los sesgos de entrenamiento. Del mismo modo que en el conjunto de datos anterior, el F1-score de la Tabla 3 es mejor para el modelo entrenado con PCA en seis de las ocho clases de estructura secundaria, lo cual indica que dicho modelo realiza predicciones con un mayor grado de compromiso entre precisión y recall. En cuanto a la precisión CNN+PCA domina en las clases L, B, E y S, mientras que las clases L, G, H, T denotan un mejor recall. En este conjunto de datos la

clase I obtiene nuevamente un resultado que indica que los modelos no logran identificar esta clase, probablemente por la escasez de instancias de este tipo en los conjuntos de datos y en las proteínas en general.

SS8 Casp10	Recall			Precisión			F1-score		
	CNN	CNN LDA	CNN PCA	CNN	CNN LDA	CNN PCA	CNN	CNN LDA	CNN PCA
L	0.648	0.682	0.691	0.542	0.528	0.546	0.590	0.595	0.610
B	0.066	0.066	0.053	0.484	0.652	0.522	0.116	0.120	0.096
E	0.860	0.782	0.819	0.723	0.783	0.784	0.786	0.783	0.801
G	0.290	0.296	0.304	0.458	0.401	0.458	0.355	0.341	0.365
I	0	0	0	0	0	0	0	0	0
H	0.875	0.877	0.903	0.875	0.844	0.860	0.875	0.860	0.881
S	0.252	0.206	0.224	0.459	0.445	0.465	0.325	0.282	0.302
T	0.541	0.578	0.570	0.687	0.632	0.661	0.605	0.604	0.612

Tabla 3 – En esta tabla se muestran los resultados de aplicar las métricas de evaluación planteadas en la Sección 4.3 para cada una de las clases de estructura secundaria en Casp10

Finalmente, en la Tabla 4 donde se muestran los resultados para el conjunto de datos **CASP11**, se puede observar que los resultados denotan una diferencia en la tendencia de precisión que se venía dando con los dos conjuntos de datos anteriores. No obstante, el patrón del F1 y el recall se conservan presentando superioridad en los mismos tipos de estructura secundaria de las tablas anteriores.

6. Conclusiones y trabajo futuro

Los resultados obtenidos en este trabajo demuestran que el uso de métodos de reducción de dimensión puede potenciar los procesos de aprendizaje en modelos de predicción de estructuras secundarias debido a la mejora en el desempeño de la clasificación con un menor número de características. Esto genera un primer paso en la obtención de modelos de predicción más eficientes que puedan ser combinados en trabajos futuros con otras arquitecturas, para generar mejor calidad en la clasificación e inclusive abordar mejor el problema de desbalanceo presente en esta base de datos. Adicionalmente, sería importante la aplicación de métodos de reducción de dimensión no lineales que generen mejores espacios embebidos con un número menor de características reduciendo los parámetros que definen a los modelos. Finalmente, teniendo en cuenta el considerable volumen de información que se maneja para esta tarea de clasificación es necesario que los métodos de RD utilizados tengan la característica de mapear nuevos puntos de un conjunto de datos determinado.

SS8 Casp11	Recall			Precisión			F1-score		
	CNN	CNN LDA	CNN PCA	CNN	CNN LDA	CNN PCA	CNN	CNN LDA	CNN PCA
L	0.604	0.641	0.635	0.506	0.480	0.494	0.551	0.549	0.556
B	0.060	0.035	0.050	0.571	0.467	0.476	0.109	0.065	0.090
E	0.849	0.746	0.795	0.706	0.770	0.765	0.771	0.758	0.779
G	0.209	0.200	0.224	0.355	0.291	0.387	0.263	0.237	0.284
I	0	0	0	0	0	0	0	0	0
H	0.859	0.871	0.895	0.852	0.821	0.840	0.855	0.845	0.866
S	0.220	0.184	0.195	0.441	0.393	0.436	0.293	0.251	0.270
T	0.465	0.485	0.489	0.597	0.544	0.568	0.523	0.513	0.525

Tabla 4 – En esta tabla se muestran los resultados de aplicar las métricas de evaluación planteadas en la Sección 4.3 para cada una de las clases de estructura secundaria en Casp11

Agradecimientos

Esta investigación está financiada y apoyada por la segunda convocatoria del año 2019 del proyecto "Formación del talento humano de alto nivel para el fortalecimiento de necesidades estratégicas de CTei del departamento de Nariño" del Programa "Bécate Nariño" de la fundación CEIBA. Además de ser respaldada por el Grupo de Investigación en Inteligencia Computacional (GICO) del Departamento de Sistemas de la Universidad del Cauca.

Referencias

Angioloni, L. (2019). Protein Secondary Structure-CNN. *University of Florence*, 1–6. Retrieved from: https://n9.cl/4qy5.

Cuff, J. A., & Barton, G. J. (1999). Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins: Structure, Function, and Genetics*, *34*(4), 508–519.

Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. *4rd International Conference on Learning Representations (ICLR), Workshop*, (pp. 2013–2016).

Dunitz, J. D., & Scheraga, H. A. (2004). Exercises in prognostication: Crystal structures and protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(40), 14309–14311. https://doi.org/10.1073/pnas.0405744101

Fang, C., Shang, Y., & Xu, D. (2018). MUFOLD-SS: New deep inception-inside-inception networks for protein secondary structure prediction. *Proteins: Structure, Function, and Bioinformatics*, *86*(5), 592–598. https://doi.org/10.1002/prot.25487

Guo, Y., Li, W., Wang, B., Liu, H., & Zhou, D. (2019). DeepACLSTM: Deep asymmetric convolutional long short-term memory neural models for protein secondary structure prediction. *BMC Bioinformatics*, *20*(1), 1–12. https://doi.org/10.1186/s12859-019-2940-0

Guo, Y., Wang, B., Li, W., & Yang, B. (2018). Protein secondary structure prediction improved by recurrent neural networks integrated with two-dimensional convolutional neural networks. *Journal of Bioinformatics and Computational Biology*, *16*(5), 1–16. https://doi.org/10.1142/S021972001850021X

Jenike, M. A., & Albert, M. S. (2009). The Elements of Statistical Learning. In *Journal of the American Geriatrics Society,* 32(6). https://doi.org/10.1007/b94608

Jiang, Q., Jin, X., Lee, S. J., & Yao, S. (2017). Protein secondary structure prediction: A survey of the state of the art. *Journal of Molecular Graphics and Modelling*, *76*, 379–402. https://doi.org/10.1016/j.jmgm.2017.07.015

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/rsta.2015.0202

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637. https://doi.org/10.1002/bip.360221211

Khoury, G. A., Smadbeck, J., Kieslich, C. A., & Floudas, C. A. (2014). Protein folding and de novo protein design for biotechnological applications. *Trends in Biotechnology*, *32*(2), 99–109. https://doi.org/10.1016/j.tibtech.2013.10.008

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR)*, (pp. 1–15).

Kumar, P., Bankapur, S., & Patil, N. (2020). An enhanced protein secondary structure prediction using deep learning framework on hybrid profile based features. *Applied Soft Computing Journal*, *86*, 1–10. https://doi.org/10.1016/j.asoc.2019.105926

Li, D., Li, T., Cong, P., Xiong, W., & Sun, J. (2012). A novel structural position-specific scoring matrix for the prediction of protein secondary structures. *Bioinformatics*, *28*(1), 32–39. https://doi.org/10.1093/bioinformatics/btr611

Li, Z., & Yu, Y. (2016). Protein secondary structure prediction using cascaded convolutional and recurrent neural networks. *IJCAI International Joint Conference on Artificial Intelligence*, (pp. 2560–2567). AAAI Press.

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., & Tramontano, A. (2014). Critical assessment of methods of protein structure prediction (CASP) - round x. *Proteins: Structure, Function and Bioinformatics*, *82*, 1–6. https://doi.org/10.1002/prot.24452

Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., & Wu, C. H. (2015). UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, *31*(6), 926–932. https://doi.org/10.1093/bioinformatics/btu739

Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2018.08.003

Wang, G., & Dunbrack, R. L. (2003). PISCES: A protein sequence culling server. *Bioinformatics*, *19*(12), 1589–1591.

Wang, S., Peng, J., Ma, J., & Xu, J. (2016). Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. *Scientific Reports*, *6*(1), 1–12. https://doi.org/10.1038/srep18962

Wang, S., Sun, S., Li, Z., Zhang, R., & Xu, J. (2017). Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLOS Computational Biology*, *13*(1), 1–34. https://doi.org/10.1371/journal.pcbi.1005324

Zhang, B., Li, J., & Lü, Q. (2018). Prediction of 8-state protein secondary structures by a novel deep learning architecture. *BMC Bioinformatics*, *19*(1), 1–13. https://doi.org/10.1186/s12859-018-2280-5

Zhou, J., & Troyanskaya, O. G. (2014). Deep supervised and convolutional generative stochastic network for protein secondary structure prediction. *31st International Conference on Machine Learning*, (pp. 1121–1129). JMLR.org.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 14/09/2020
Aceitação/Acceptance: 25/11/2020

Computación Cuántica Adiabática aplicada a la solución del Problema de la Mochila Binaria

Danilo López-Sandoval, Carlos-Alberto Cobos-Lozada

dlopezs@unicauca.edu.co; ccobos@unicauca.edu.co

Grupo de I+D en Tecnologías de la Información (GTI), Universidad del Cauca, Calle 5 # 4-70, 190003, Popayán, Colombia.

Pages: 214-227

Resumen: Día a día los problemas que enfrenta la humanidad son más complejos, y las herramientas que nos brinda la computación clásica empiezan a ser insuficientes para resolverlos, como por ejemplo los problemas de optimización donde el espacio de búsqueda crece exponencialmente. En este contexto, la computación cuántica ha empezado a mostrar su dominio y en los últimos años empresas como IBM empiezan a ofertar servicios de cómputo cuántico para tareas de optimización y análisis de datos. Este artículo presenta y detalla los pasos seguidos para construir una solución al problema de la mochila binaria utilizando un algoritmo de Computación Cuántica Adiabática haciendo uso de las librerías de Qiskit y su implementación en Python. Por último, presenta una comparación del algoritmo cuántico obtenido con una versión clásica de recocido simulado. Los resultados obtenidos muestran la efectividad del algoritmo cuántico y una leve superioridad frente a su contraparte clásica.

Palabras-clave: Computación cuántica; Problema de la mochila; Computación cuántica adiabática; Modelo de Ising; Problemas de optimización.

Adiabatic Quantum Computing applied to the solution of the Binary Knapsack Problem

Abstract: The problems humanity faces are day by day ever more complex and the tools provided by classical computing are beginning to be insufficient to solve certain of these, such as optimization problems in which the search space grows exponentially. Quantum computing has begun to show its dominance in this realm and in recent years companies like IBM have begun to offer quantum computing services for data analysis and optimization tasks. This paper presents and details the steps followed to build a solution to the binary knapsack problem using an Adiabatic Quantum Computing algorithm making use of the Qiskit libraries, and their implementation in Python. Finally, a comparison is presented of the quantum algorithm obtained with a classic simulated annealing version. The results show the effectiveness of the quantum algorithm and the slight superiority it offers compared to its classical counterpart.

Keywords: Quantum computing; Knapsack problem; Adiabatic quantum computing; Ising model; Optimization problems.

1. Introducción

El problema de la mochila es un conocido problema de optimización combinatoria que pertenece a la clase de problemas NP-completos (Karp, 2010). Existen diferentes variantes de este problema que se utilizan como guía para resolver una gran variedad de problemas, entre ellos, problemas de empaque y reducción de existencias, toma de decisiones financieras, titulización respaldada por activos, subastas combinatorias (Coffey, 2017), control de presupuestos, toma de decisiones y corte de material (H. Wang, Ma, Zhang, & Li, 2009).

Entre las variantes más conocidas del problema de la mochila se encuentran: la mochila binaria (Binary Knapsack Problem, BKP), el problema de la suma de subconjuntos ($P_i = W_i$), el problema de la mochila sin límite, el problema de la mochila acotada (Martello & Toth, 1987), el problema de la mochila d-dimensional (d-KP), el problema de las múltiples mochilas (MKP) (Gurski, Rehs, & Rethmann, 2019), el problema de la mochila de elección múltiple multidimensional y el problema de la mochila cuadrática (Assi & Haraty, 2019). A la fecha se han desarrollado una variedad de técnicas para resolver estas variantes, las cuales se pueden agrupar de la siguiente manera: (i) exactos, (ii) programación dinámica, (iii) programación entera, (iv) métodos metaheurísticos, (v) métodos lagrangianos, (vi) métodos basados en árboles de búsqueda con back tracking y (vii) enfoques de red (Salkin & de Kluyver, 1975).

El problema de la mochila binaria se define formalmente en la Ecuación (1), y se describe así: dada una mochila con una capacidad limitada $\mathbf{C} \in Z^+$, y un conjunto de $\boldsymbol{n}$ elementos (artículos), cada uno con un beneficio $\mathbf{P_i} \in Z$ y un peso $\mathbf{W_i} \in Z$ (i = 1, 2, ..., n), se debe seleccionar un subconjunto de $\boldsymbol{m}$ elementos ($m \leq n$) de modo que se genere la mayor (máxima) ganancia o beneficio posible sujeto a una restricción principal, que los pesos totales de los elementos seleccionados no excedan la capacidad C de la mochila (H. Wang et al., 2009). En este artículo se trabajará el problema de la mochila binaria, donde x_i es un valor binario {0,1} que indica si el elemento *i* debe ir o no en la mochila.

$$Max\ f(x) = \sum_{i=1}^{n} P_i \times x_i$$
$$sujeto\ a \sum_{i=1}^{n} W_i \times x_i \leq C \qquad (1)$$
$$x_i \in \{0,1\}, i = 1, 2, \dots, n$$

La dimensión del problema se puede definir de acuerdo con el número *n* de artículos y en la literatura se puede encontrar como pequeñas si $n < 100$, medianas si $100 \leq n < 1000$, o grandes si $n \geq 1000$, aunque esta definición no es del todo aceptada, ya que otros autores la definen como pequeñas si $n < 20$, medianas si $n < 200$ y grandes si $n \geq 200$ (Blado & Toriello, 2019). Una forma más aceptada de clasificación se hace basado en la complejidad de las instancias, esto es, la relación o no de los pesos de cada ítem con su ganancia, y se definen como: instancias No correlacionadas, débilmente correlacionadas, casi fuertemente correlacionadas, fuertemente correlacionadas, Inversa y fuertemente

correlacionadas, e instancias con pesos y beneficios iguales (subconjuntos de instancias de suma) (Pisinger, 2005).

Debido a la importancia y el reto que representa el problema de la mochila binaria, en los últimos años se han reportado un gran número de algoritmos que buscan su solución, estos se agrupan en: exactos, de programación dinámica, basados en back-tracking (incluidos ramificación y poda), y metaheurísticos. Entre los algoritmos metaheurísticos más destacados se encuentran los algoritmos genéticos, el recocido simulado, la optimización por enjambre de partículas (PSO) y la búsqueda tabú (Vickram, Krishna, & Srinivas, 2016; H. Wang et al., 2009). Además recientemente se trata con algoritmos cuánticos como el algoritmo evolucionario cuántico (Li & Li, 2019), el algoritmo genético cuántico (R. Wang, Guo, Xiang, & Mao, 2012) y el algoritmo cuántico Variational Quantum Eigensolver (VQE) (Bolos, 2019).

La computación cuántica nace como una alternativa al paradigma computacional clásico basado en máquinas de Turing y de Von Neumann, la cual ha demostrado su superioridad ante la computación clásica para algunos problemas específicos (Albash & Lidar, 2018; Nielsen & Chuang, 2010). Con este nuevo paradigma se pueden estudiar problemas de alta complejidad que requieren gran cantidad de operaciones y manejan gran cantidad de variables, para esto, se hace uso de algunas propiedades de la física cuántica como el entrelazamiento cuántico o la superposición cuántica, con las cuales se pueden realizar más operaciones en una misma unidad de tiempo disminuyendo radicalmente los tiempos de respuesta (Vega Fernández & Ramírez Celis, 2017).

A la fecha se han definido algoritmos cuánticos eficientes para problemas discretos como la factorización entera, la simulación cuántica, la estimación del valor propio, la integración, la solución de ecuaciones diferenciales parciales y la solución a problemas numéricos de álgebra lineal (Hadfield, 2018), además la dificultad especial de resolver problemas de optimización ha generado mucha expectativa sobre la posibilidad de usar dispositivos cuánticos para resolverlos aproximadamente (menos tiempo con mejores soluciones). Un algoritmo que ha ido evolucionando mediante la combinación de metaheurísticas y conceptos cuánticos particularmente destacado es el algoritmo de Computación Cuántica Adiabática (Adiabatic Quantum Computing, AQC) (Hadfield, 2018), el cual utiliza un enfoque equivalente al modelo de circuito de computación cuántica, adecuado para problemas de optimización combinatoria, incluyendo particiones, coberturas, particionado de árboles y grafos.

Una búsqueda en la literatura reveló que existen pocos artículos publicados sobre algoritmos cuánticos que se apliquen al problema de la mochila binaria, debido a esto, el propósito de este documento es detallar los pasos a seguir para solucionar problemas de optimización mediante computación cuántica a través del enfoque AQC, solucionando el problema de la mochila binaria mapeándolo a un modelo de Ising y realizando su implementación en el lenguaje de programación Python, haciendo uso de la librería Qiskit Aqua. Se utiliza específicamente el algoritmo VQE desarrollado por el equipo de IBM para solucionar problemas cuánticos.

Los aportes principales del presente documento son: 1) detallar la forma como se aborda un problema de optimización mediante el enfoque AQC, 2) dar a conocer el funcionamiento de un algoritmo adiabático cuántico aplicado al problema de la mochila

binaria, y 3) comparar los resultados obtenidos frente a una implementación clásica de recocido simulado en instancias de baja complejidad (n < 20).

A continuación, en la sección 2 se introducen algunos conceptos básicos de computación cuántica. En la sección 3, se presenta la solución del problema de la mochila binaria realizada con el enfoque AQC. Luego, en la sección 4 se describe brevemente la experimentación y los resultados obtenidos junto con la comparación de dichos resultados frente a un recocido simulado clásico. Finalmente, en la sección 5 se presentan las conclusiones del trabajo realizado y el trabajo futuro que el grupo de investigación espera desarrollar en el futuro cercano.

2. Conceptos Cuánticos aplicados al problema de la mochila

2.1. Computación Cuántica

La computación cuántica es un nuevo paradigma de computación que surgió como resultado de la fusión de la informática y la mecánica cuántica. El origen de la computación cuántica se remonta a principios de los 80's cuando Richard Feynman observó que algunos efectos de la mecánica cuántica no se pueden simular de manera eficiente en una computadora clásica. Durante la última década, la computación cuántica ha atraído un interés generalizado y ha inducido intensivas investigaciones, ya que ha demostrado su superioridad en ciertos tipos de algoritmos a su contraparte clásica, en especial destaca su paralelismo innato que reduce la complejidad algorítmica. Tal capacidad de procesamiento en paralelo se puede utilizar para resolver problemas de optimización combinatoria que requieren la exploración de grandes espacios de soluciones (Layeb, 2011).

Qubit: Un bit cuántico (Quantum bit) es la unidad de información más pequeña almacenada en una computadora cuántica de dos estados (Hey, 1999). Al contrario del bit clásico el cual tiene dos valores posibles, "0" o "1", un qubit puede estar en el estado "1", en el estado "0" o en cualquier superposición de los dos. El estado de un qubit se puede representar mediante la notación de corchetes (Bra-Ket) presentada en la Ecuación (2).

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{2}$$

Donde $|\Psi\rangle$ denota más de un vector $\vec{\Psi}$ aen algún espacio vectorial y representa una superposición lineal de la partícula dados los vectores de estado cuántico individuales. $|0\rangle$ y $|1\rangle$ representan los valores de bit clásicos 0 y 1 respectivamente. α y β son números complejos tales que $|\alpha|^2 + |\beta|^2 = 1$, los cuales especifican las amplitudes de probabilidad de los estados correspondientes. Cuando se mide el estado del qubit, se puede obtener cero con una probabilidad $|\alpha|^2$ o uno con una probabilidad $|\beta|^2$; Esta medición se realiza al observar el estado cuántico de un qubit, momento en el que colapsa a un solo estado, cero (0) o uno (1) (Narayanan, 1999).

Entre los algoritmos cuánticos más famosos se encuentran el algoritmo de Shore's (Shor, 1994) utilizado para factorización numérica y el algoritmo de Grover's (Grover, 1996) utilizado para búsquedas en una base de datos no ordenada. Ambos algoritmos

redujeron la complejidad de la solución al problema (Laboudi & Chikhi, 2012). Un sistema de n-qubits puede representar 2^n estados al mismo tiempo. Este crecimiento exponencial del espacio de estado con el número de qubits es lo que sugiere una aceleración exponencial de la computación en las computadoras cuánticas sobre las computadoras tradicionales.

2.2. Computación Cuántica Adiabática (AQC)

La computación cuántica adiabática (Adiabatic Quantum Computation, AQC) es un enfoque equivalente al modelo de circuito de computación cuántica, adecuado para problemas del tipo de optimización combinatoria, incluyendo particiones, coberturas y particionado de árboles y grafos (Coffey, 2017).

Debido a sus inicios, la AQC puede considerarse como una clase particular de Recocido cuántico (Quantum Annealing, QA), la cual utiliza los principios de la mecánica cuántica para resolver problemas de optimización sin restricciones (Albash & Lidar, 2018). Desde la propuesta inicial de QA, ha habido mucho interés en la búsqueda de problemas prácticos donde pueda ser ventajoso con respecto a los algoritmos clásicos, particularmente el recocido simulado (Simulating annealing, SA). Muchos de estos enfoques transforman un problema computacional en un problema donde se debe encontrar el estado fundamental de un modelo de Ising Spin Glass (ISG) cuántico, el cual, en el peor de los casos es NP-completo (Cao, Jiang, Perouli, & Kais, 2016).

El modelo Ising (Una clase conveniente, restringida y ciertamente no universal de hamiltonianos) tiene la versatilidad de codificar eficientemente muchos problemas NP y ha motivado la realización física de QA. En general las computadoras cuánticas universales no pueden resolver problemas NP-hard de manera eficiente, pero se ha encontrado evidencia en los sistemas experimentales de Ising cuántico que sugieren una aceleración cuántica sobre la computación clásica debido al efecto del túnel cuántico[1] (Cao et al., 2016).

Optimización Cuántica Adiabática: En un modelo de circuito de computación cuántica, un cálculo puede evolucionar en todo el espacio de Hilbert[2] y está codificado en una serie de puertas de lógica cuántica unitarias (Albash & Lidar, 2018). Por otro lado, en un modelo AQC el cálculo se realiza mediante un Hamiltoniano inicial (H_0), cuyo estado fundamental codifica la solución a un problema de interés, y otro hamiltoniano (H_p), cuyo estado fundamental es trivial. Entonces, si se prepara un sistema cuántico para estar en el estado fundamental de H_0, y luego se cambia adiabáticamente el hamiltoniano por un tiempo T de acuerdo con la Ecuación (3) y T es lo suficientemente grande, al final, el estado cuántico en T devolverá una solución al problema de interés (Albash & Lidar, 2018)(Lucas, 2014)(Cao et al., 2016).

$$H(t) = (1 - s)H_0 + sH_p \tag{3}$$

[1] En algoritmos cuánticos se refiere a pasar de un estado A a un estado B no contiguo para evitar estancarse en óptimos locales.
[2] Espacio de Hilbert: espacio vectorial complejo con un producto interno que también está completo.

Donde s ∈ [0, 1], es un parámetro de tiempo considerado como una función dependiente del tiempo s(t) = $\frac{t}{T}$ para un tiempo total de evolución T (en general se puede considerar cualquier función que satisfaga s(0)=0 y s(1)=1), y H_0 y H_p no conmutan[3]. Con esta definición y debido al teorema adiabático de la mecánica cuántica el sistema cuántico permanecerá en el estado fundamental todo el tiempo.

2.3. Modelo Ising

Uno de los modelos más utilizados en física se llama el modelo Ising. Propuesto entre 1920 y 1930 por Ernst Ising y Wilhelm Lenz como una forma de entender el funcionamiento de los materiales magnéticos. El enfoque modela un material magnético como una colección de moléculas, cada una de las cuales tiene un espín que puede alinearse o anti-alinearse con un campo magnético aplicado h_i, y que interactúan entre sí con base en un campo de interacción J_{ij} (Brush, 1967).

En la Ecuación (4) se representa el modelo clásico de Ising, el cual se puede escribir como una función cuadrática de un conjunto de *n* giros, donde $s_i \in \{-1, +1\}$ representa el spin de la i-ésima partícula.

$$H(s_1, \ldots, s_N) = -\sum_{i=1}^{n} J_{ij} s_i s_j - \sum_{i=1}^{n} h_i s_i \tag{4}$$

La Ecuación (5) representa un Hamiltoniano (H_p) como la versión cuántica del modelo Ising, donde s_i se reemplaza por σ_i^z en la Ecuación (4).

$$H_p = H\left(\sigma_1^z, \ldots, \sigma_N^z\right) = -\sum_{i<j} J_{ij} \sigma_i^z \sigma_j^z - \sum_{i=1}^{n} h_i \sigma_i^z \tag{5}$$

$\sigma_i^z = \mathbb{I}^{\otimes(i-1)} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \mathbb{I}^{\otimes(n-i)}$ es una matriz de Pauli[4] que actúa sobre el i-ésimo spin en un espacio de Hilbert de N qubits $\{|+\rangle, |-\}^{\otimes N}$ donde $\mathbb{I}$ es una matriz identidad de 2 x 2 y h_i, J_{ij} Î R son coeficientes (Lucas, 2014)(Cao et al., 2016). h_i es la fuerza del campo aplicado sobre el i-ésimo spin y J_{ij} actúa como el campo de interacción entre los spines vecinos i, j (Bian, Chudak, Macready, & Rose, 2010). Un estado fundamental H_0 es una superposición de todos los estados posibles en la base propia de H_p (Lucas, 2014) donde σ_i^x es la puerta NOT sobre el i-ésimo qubit (Coffey, 2017).

3. Solución cuántica al problema de la mochila binaria

En esta sección se presentan dos modelos Ising para solucionar el problema de la mochila (Lucas, 2014). Los dos modelos se explican en (Lucas, 2014), el segundo además se detalla en (Coffey, 2017) y su hamiltoniano cuántico se implementa en (Bolos, 2019).

[3] Dos operadores no conmutan cuando se cumple: [A, B] = AB - BA ≠ 0 (Benenti, Casati, & Strini, 2007).

[4] Matriz de 2x2 cuyo primo $\left[\left(\mathbb{I} + \sigma_i^z\right) / 2\right]$ tiene vectores propios $|0,1\rangle$ con valores propios 0,1 (Lucas, 2014)

3.1. Solución con (n + C) qubits

Teniendo en cuenta la definición del problema de la mochila propuesta en la Ecuación (1) y trabajando con la versión binaria de esta, se tienen *n* variables x_i donde $1 \leq i \leq n$, las cuales denotan valores binarios que se fijan con uno (1) si un determinado elemento debe ir en la mochila o cero (0) en caso contrario, y *j* variables y_j donde $1 \leq j \leq C$ que denotan una variable binaria la cual es 1 si el peso de la mochila es j, o cero (0) en caso contrario. La cantidad de qubits requerida para modelar la solución es (*n* + C) -> $[x_1, \ldots, x_n, y_1, \ldots y_C]$.

La solución al problema consiste en construir un modelo de Ising H = $H_A + H_B$, con H_A como el término que controla el cumplimiento o no de la restricción del problema, cómo se define en la Ecuación (6).

$$H_A = A\left(1 - \sum_{j=1}^{C} y_j\right)^2 + A\left(\sum_{k=1}^{C} k y_k - \sum_{i=1}^{n} W_i x_i\right)^2 \tag{6}$$

Con H_A se obliga a que el peso solo pueda tomar un valor y que el peso de los objetos en la mochila sea igual al valor que se desea. Por ejemplo, si una solución da un valor igual a 10 para la variable peso, solo $y_{10} = 1$, el resto de $y_k = 0$ para los otros valores de k. Si la capacidad de la mochila (C) fuese de 20, esta solución es factible y en ese caso H_A sería igual a $A(1-1)^2 + A(10-10)^2 = 0$, eliminando así a H_A. Pero si C tuviera u valor igual a 8, esta solución no sería factible y en este caso H_A sería igual a $A(1-0)^2 + A(0-10)^2 = 101A$ haciendo que el valor de H_A sea grande (positivo, alejando la solución del mínimo), y entre más se supere el peso máximo de la mochila, mayor será el valor de H_A. Finalmente se tiene H_B en la Ecuación (7).

$$H_B = -B\sum_{i=1}^{n} P_i x_i \tag{7}$$

El término H_B sirve para maximizar el beneficio (a mayor beneficio más se acerca al optimo o mínimo del Hamiltoniano) y el término H_A asegura que se cumpla la restricción de peso total. Debido a que se requiere que no sea posible encontrar una solución donde H_A no se cumpla a expensas de que H_B se vuelva más negativo, se requiere fijar la condición: 0 < B * Max (P_i) < A. El número de qubits requeridos para esta formulación es de n + C, número de artículos más capacidad (Lucas, 2014).

3.2. Solución con (n + ⌊log2 C⌋ + 1) qubits

Realizando una modificación a la solución planteada anteriormente, la cual requiere (*n* + C) qubits, se desea reducir drásticamente la cantidad de giros y_j adicionales que deben agregarse. Así, codificando la magnitud del peso máximo (C) como un valor binario, se define una variable entera M tal que $2^M \leq C < 2^{M+1}$, con M = $\lfloor \log_2 C \rfloor$; De esta manera solo se necesitan (M+1) variables binarias: $[y_0 \ldots y_M]$, en lugar de C variables binarias: $[y_1 \ldots y_C]$, para codificar una variable que puede tomar C valores (C >> M).

Aplicando la nueva configuración, para resolver el problema de la mochila, se requieren (n+$\lfloor \log_2 C \rfloor$+1) qubits (Lucas, 2014). En la Ecuación (8) se presenta H_A construido de acuerdo a la nueva representación de y_j (Coffey, 2017).

$$H_A = A\left(\sum_{j=0}^{M-1} 2^j y_j + (C + 1 - 2^M)y_M - \sum_{i=1}^{n} W_i x_i\right)^2 \quad (8)$$

Si se tiene una solución con un peso de 10 y una capacidad máxima C = 20, entonces la cantidad de qubits requeridos para representar el peso de la solución es M=$\lfloor \log_2 20 \rfloor$+1=5, por lo tanto se tiene el vector $y = [0\ 1\ 0\ 1\ 0]$ = 10, que corresponde a la representación de 10 en binario de izquierda a derecha; Así, se cumple la condición $16 \leq C = 20 < 32$. Esta solución es factible, por lo que $H_A = A\ (10 + (20 + 1 - 32)0 - 10)^2 = 0$, quedando eliminado H_A. Por otro lado, si C = 9, se tendría el vector $y = [0\ 1\ 0\ 1] = 10$ con $H_A = A\ (2 + (9 + 1 - 8)1 - 10)^2 = 36\ A$, que representa un valor alto de contrapeso ya que se supera la capacidad máxima permitida. Este valor crecerá entre más se rebase la capacidad de la mochila.

Este segundo hamiltoniano es el que se toma en el presente trabajo como solución del problema de la mochila binaria. En resumen, el Hamiltoniano H = $H_A + H_B$ esta dado por la Ecuación (9), que utiliza una representación en base cero (0) para los vectores.

$$H = A\left(\sum_{j=0}^{M-1} 2^j y_j + (C + 1 - 2^M)y_M - \sum_{i=0}^{n-1} W_i x_i\right)^2 - B\sum_{i=0}^{n-1} P_i x_i \quad (9)$$

Los diferentes valores implicados se representan de la siguiente manera (Bolos, 2019):

- La solución se representa como un vector $\vec{X}$ = [x_0 ... x_{n-1}] ∈ {0, 1} en donde cada elemento se representa en un índice del vector. Si el valor es 1, entonces el elemento se almacena en la mochila y si es cero (0) el elemento queda fuera.
- El peso de la solución se representa en binario de izquierda a derecha como un vector $\vec{Y}$ = [y_0 ... y_m] ∈ {0, 1}.
- El valor del peso de los elementos se representa como un vector W= [W_0 ... W_{n-1}].
- El beneficio de los elementos se representa como un vector P = [P_0 .. P_{n-1}].
- La variable C representa la capacidad de la mochila.
- La variable A representa un número lo suficientemente grande para dominar la multiplicación del valor B por la suma de los pesos de todos los elementos.

Para construir una función de costo C($\vec{K}$) se debe tener en cuenta que se desea maximizar la Ecuación (1) mientras se cumple con su respectiva restricción (Bolos, 2019). De la definición de H, se tienen M+1 nuevas variables binarias que se representarán en el vector y = [y_0 ... y_M], donde los elementos pueden tomar valores diferentes de cero simultáneamente (Coffey, 2017). La solución en la función de costo se representará en un vector K compuesto por los vectores $\vec{X}$ y $\vec{Y}$, con $\vec{K}$ = [x_0 ... x_{n-1},y_n ... y_{n+M-1}]. En la Ecuación (10) se presenta la función de costo C(K) definida por (Bolos, 2019) y su polinomio expandido.

$$C(\vec{K}) = L\left(C - \sum_{i=0}^{n-1} W_i x_i - S\right)^2 - \sum_{i=0}^{n-1} P_i x_i;\ \text{donde } S = \sum_{i=n}^{n+M-1} 2^j y_j$$
$$C(\vec{K}) = L\left(C^2 + \left(\sum_{i=0}^{n-1} W_i x_i\right)^2 + S^2 - 2C\sum_{i=0}^{n-1} W_i x_i - 2CS + 2S\sum_{i=0}^{n-1} W_i x_i\right) - \sum_{i=0}^{n-1} P_i x_i \quad (10)$$

Se puede observar que para llevar el primer término de la Ecuación (10) a cero, se debe cumplir que $C \geq \sum_{i=0}^{n-1} W_i x_i$ y se tenga un $S = C - \sum_{i=0}^{n-1} W_i x_i$. El valor de L debe dominar la suma de los valores de P, esto es, $L > \sum_{i=0}^{n-1} P_i$. Si por el contrario $C < \sum_i W_i x_i$ el primer término nunca será cero que multiplicado por L evitará seleccionar soluciones no factibles. El valor mínimo de la función de costo será uno donde la restricción es respetada y la suma de los valores se maximice (Bolos, 2019).

3.3. Mapeo de la función de costo a su Hamiltoniano cuántico

Al realizar el mapeo de las variables x_i por $(\mathbb{I} - Z_i)/2$, y_j por $(\mathbb{I} - Z_j)/2$ y eliminar el primer término que es constante, se obtiene la Ecuación (11), donde $\mathbb{I}$ es una matriz identidad de orden $n+m$ y Z_i es la matriz resultante de tensorizar $n+m$ matrices donde en cada posición diferente de i los factores del tensor son matrices identidad y la posición i es una matriz de Pauli Z.

La matriz de Pauli se compone de dos vectores, un vector X y un vector Z, para este caso se modifica el vector Z únicamente. Para hallar los coeficientes de la matriz se debe resolver cada sumatoria resultante en la Ecuación (11), encontrar el respectivo índice para cada valor de la sumatoria e ir adicionándolo a una lista. Por ejemplo, para el término $\sum_{i=0}^{n-1} W_i \left(\frac{\mathbb{I} - Z_i}{2}\right) \times \sum_{i=0}^{n-1} W_i \left(\frac{\mathbb{I} - Z_i}{2}\right)$, se muestra su algoritmo en la Figura 1.

$$C(\vec{K}) = L\left(\begin{array}{c}\left(\sum_{i=0}^{n-1} W_i\left(\frac{\mathbb{I} - Z_i}{2}\right) \times \sum_{i=0}^{n-1} W_i\left(\frac{\mathbb{I} - Z_i}{2}\right)\right) \\ + \left(\sum_{j=n}^{n+M-1} 2^j\left(\frac{\mathbb{I} - Z_j}{2}\right) \times \sum_{j=n}^{n+M-1} 2^j\left(\frac{\mathbb{I} - Z_j}{2}\right)\right) - 2C\sum_{i=0}^{n-1} W_i\left(\frac{\mathbb{I} - Z_i}{2}\right) \\ -2C\sum_{j=n}^{n+M-1} 2^j\left(\frac{\mathbb{I} - Z_j}{2}\right) + 2\left(\sum_{j=n}^{n+M-1} 2^j\left(\frac{\mathbb{I} - Z_j}{2}\right) \times \sum_{i=0}^{n-1} W_i\left(\frac{\mathbb{I} - Z_i}{2}\right)\right)\end{array}\right) - \sum_{i=0}^{n-1} P_i x_i \quad (11)$$

Hay que tener en cuenta que no es necesario calcular los tensores por cuenta propia, ya que Qiskit tiene incorporado un objeto que toma los índices de la matriz de Pauli Z en el tensor y realiza los cálculos restantes.

Como resultado se obtiene una lista de objetos de Pauli, cada una con sus respectivos índices, y un valor shift que se encarga se acumular la diferencia de las magnitudes de la operación de cada sumatoria. Esos valores se envían al algoritmo Variational Quantum Eigensolver (VQE) con el objetivo de minimizar la función y retornar una solución X' donde los primeros n términos son la solución al problema de la mochila. La solución completa está a disposición de la comunidad en https://bit.ly/3kA4dJl.

4. Resultados

Para realizar la evaluación del algoritmo, se contó con un conjunto de 11 problemas disponibles en archivos de texto. Cada archivo contiene el número de elementos (n), la capacidad de la mochila (C), los pesos y valores para cada elemento y el valor óptimo conocido para el problema. Los valores de n, C y el resultado óptimo se muestran en la parte izquierda de la Tabla 1.

```
#term for sum(x_i*w_i)^2
for i in range(n):
    for j in range(n):
        coef = -1 * 0.25 * weights[i] * weights[j] * M

        xp = np.zeros(num_values, dtype=np.bool)
        zp = np.zeros(num_values, dtype=np.bool)
        zp[j] = not zp[j]
        pauli_list.append([coef, Pauli(zp, xp)])
        shift -= coef

        xp = np.zeros(num_values, dtype=np.bool)
        zp = np.zeros(num_values, dtype=np.bool)
        zp[i] = not zp[i]
        pauli_list.append([coef, Pauli(zp, xp)])
        shift -= coef

        coef = -1 * coef
        xp = np.zeros(num_values, dtype=np.bool)
        zp = np.zeros(num_values, dtype=np.bool)
        zp[i] = not zp[i]
        zp[j] = not zp[j]
        pauli_list.append([coef, Pauli(zp, xp)])
        shift -= coef
```

Figura 1 – Implementación del término $(\sum_{i=0}^{n-1} W_i x_i)^2$ para incluir operadores de Pauli

Para realizar las pruebas con un algoritmo cuántico se partió del algoritmo presentado en (Bolos, 2019), realizado en el marco del Qiskit Camp Asia 2019 que se celebró en el mes de noviembre, se migro a programación orientada a objetos y se le realizaron las adiciones pertinentes para poder ejecutar las instancias de los problemas dados. El algoritmo hace uso de la librería Qiskit-Aqua, la cual contiene una biblioteca de algoritmos cuánticos de dominios cruzados sobre los que se pueden construir aplicaciones para computación cuántica.

n	C	Beneficio Optimo	Exhaustivo	Solución cuántica			Recocido Simulado		
			Tiempo (segundos)	Beneficio	Rata de éxito	Tiempo (segundos)	Beneficio	Rata de éxito	Tiempo (segundos)
4	11	23	0,00013	23	100,0%	0,00160	23	100,0%	0,01034
4	20	35	0,00000	35	100,0%	0,00169	35	100,0%	0,00000
5	26	51	0,00003	51	100,0%	0,00180	51	100,0%	0,00000
5	80	130	0,00003	118	0,0%	0,00223	130	100,0%	0,00755
6	190	150	0,00003	150	100,0%	0,00495	150	100,0%	0,03012
7	50	107	0,00013	107	100,0%	0,00323	81,867	3,3%	0,06885
7	170	1735	0,00007	1735	100,0%	0,00894	1735	100,0%	0,00123

n	C	Beneficio Optimo	Exhaustivo	Solución cuántica			Recocido Simulado		
			Tiempo (segundos)	Beneficio	Rata de éxito	Tiempo (segundos)	Beneficio	Rata de éxito	Tiempo (segundos)
8	104	900	0,00020	900	100,0%	0,00848	881,4	36,7%	0,22148
10	60	52	0,00156	52	100,0%	0,01732	52	100,0%	0,00003
10	165	309	0,00105	309	100,0%	0,09495	309	100,0%	0,00000
10	269	295	0,00103	295	100,0%	0,21795	295	100,0%	0,00140
Promedio			**0,00039**	**343,182**	**90,9%**	**0,03301**	**340,297**	**85,5%**	**0,03100**

Tabla 1 – Resultados de la ejecución de los algoritmos

Cada problema se resolvió 30 veces (repeticiones del mismo experimento) con el algoritmo para obtener el valor óptimo promedio encontrado, el tiempo promedio de ejecución en segundos y la tasa promedio de éxito (porcentaje de las veces que encontró el óptimo ideal o conocido). Los resultados se encuentran en la parte central de la Tabla 1. En esta tabla también se incluye una columna de tiempo de ejecución para la solución del problema con un algoritmo exhaustivo "clásico", que en este caso es posible encontrar ya que son pocos objetos o elementos en cada problema.

En la parte derecha de la Tabla 1 se muestran los resultados para una implementación de recocido simulado clásico que usa decremento geométrico de la temperatura y una mutación basado en deseleccionar (sacar) al azar un elemento que está en la mochila e incluir aleatoriamente tantos como sea necesario para llenar la mochila (solución del problema evolucionando soluciones factibles).

Los resultados muestran que la solución cuántica tiene un 5,4% más de tasa de éxito (resolviendo 10 de los 11 problemas en todas las repeticiones del experimento) y el tiempo de ejecución es aproximadamente el mismo que el requerido por el recocido simulado. El problema donde la solución cuántica se queda atrapado en un óptimo local, es una donde la mochila desperdicia la cuarta parte de su capacidad para obtener el mayor beneficio, esto debe ser considerado para proponer una mejora a la formulación del problema.

Dado que la cantidad de qubits necesarios para solucionar el problema de la mochila mediante lógica cuántica es proporcional al peso de la mochila y a la cantidad de elementos, actualmente se hace imposible tratar problemas con más de 15 elementos en entornos simulados cuánticos sobre computadores personales; esto debido a que un sistema con n-qubits puede representar 2^n estados al mismo tiempo, lo cual se convierte en una tarea difícil de realizar con la actual capacidad de procesamiento en estas computadores. Las pruebas se realizaron sobre un equipo personal con 16 Gigabytes de memoria RAM y un procesador Intel(R) Core (TM) i7-6700HQ CPU @ 2.60 GHz.

5. Conclusiones

En este documento se ha detallado una estrategia para dar solución al problema de optimización de la mochila binaria (un problema NP-completo), haciendo uso del

algoritmo de Computación Cuántica Adiabática, detallando la manera como se mapea el problema a un modelo de Ising.

El algoritmo cuántico encuentra la misma solución en cada ejecución (repetición) del experimento, lo que demuestra la estabilidad del proceso que se ejecuta. De los 11 casos de prueba utilizados, se encontró la solución óptima en 10 de ellos, obteniendo resultados de tiempo similares a los del recocido simulado con una sutil mejora en la tasa de éxitos, lo que alienta el desarrollo de nuevas investigaciones en el área. Es preciso comentar que se está usando un entorno de computación cuántica simulado y no una computadora cuántica real donde el procesamiento paralelo natural de estos computadores debe demostrar su superioridad.

Observando los valores de los promedios arrojados en el experimento, se puede ver que no existe una diferencia significativa entre los resultados de los dos algoritmos comparados, esto esencialmente debido a que se esta trabajando con instancias de baja dimensionalidad, con lo cual a los algoritmos les toma muy poco tiempo en resolver el problema; pero alienta el desarrollo de investigaciones debido a los resultados prometedores obtenidos por el algoritmo cuántico, el cual obtiene una mejor tasa de éxito.

Para abordar problemas de mayor complejidad (mayor número de elementos) se hace necesario utilizar un computador clásico de mayor capacidad o idealmente un computador cuántico con una gran cantidad de qubits. Para tomar el segundo camino, se hace necesario establecer alianzas con empresas como IBM o Google quienes están desarrollando y probando los computadores cuánticos de mayor capacidad que existen hoy en día.

Teniendo en cuenta que el algoritmo cuántico mostró una debilidad concreta en la solución de un problema donde el óptimo se encuentra cuando la mochila queda desocupada en un 25% de su capacidad, se hace necesario que el Grupo se centre en mejorar la formulación de la función de costo y ampliar los tipos de problemas (instancias) que se usan en la evaluación para poder tener un panorama completo del rendimiento real del algoritmo.

Referencias

Albash, T., & Lidar, D. A. (2018). Adiabatic quantum computation. *Reviews of Modern Physics*, *90*(1), 015002–1, 015002–015064. https://doi.org/10.1103/RevModPhys.90.015002

Assi, M., & Haraty, R. A. (2019). A Survey of the Knapsack Problem. *ACIT 2018 - 19th International Arab Conference on Information Technology*, 1–6. https://doi.org/10.1109/ACIT.2018.8672677

Benenti, G., Casati, G., & Strini, G. (2007). *Principles of quantum computation and information: Volume II: Basic tools and special topics. Principles of Quantum Computation and Information - Volume II: Basic Tools and Special Topics* (Vol. I). https://doi.org/10.1142/5838

Bian, Z., Chudak, F., Macready, W., & Rose, G. (2010). The Ising model: teaching an old problem new tricks. *D-Wave Systems*, 1–32. Retrieved from https://www.dwavesys.com/sites/default/files/Bian.pdf

Blado, D., & Toriello, A. (2019). Relaxation Analysis for the Dynamic Knapsack Problem with Stochastic Item Sizes. *SIAM Journal on Optimization*, *29*(1), 1–30. https://doi.org/10.1137/16M1101209

Bolos, S. (2019). GitHub - sorin-bolos/QiskitCampAsia2019.

Brush, S. G. (1967). History of the Lenz-Ising model. *Reviews of Modern Physics*, *39*(4), 883–893. https://doi.org/10.1103/RevModPhys.39.883

Cao, Y., Jiang, S., Perouli, D., & Kais, S. (2016). Solving Set Cover with Pairs Problem using Quantum Annealing. *Scientific Reports*, *6*(September), 1–15. https://doi.org/10.1038/srep33957

Coffey, M. W. (2017). Adiabatic quantum computing solution of the knapsack problem, 1–22. Retrieved from http://arxiv.org/abs/1701.05584

Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. *Proceedings of 28th Annual ACM Symposium on the Theory of Computing*, *41*(3), 212–221. https://doi.org/10.1007/BF01913177

Gurski, F., Rehs, C., & Rethmann, J. (2019). Knapsack problems: A parameterized point of view. *Theoretical Computer Science*, *775*, 93–108. https://doi.org/10.1016/j.tcs.2018.12.019

Hadfield, S. (2018). *Quantum Algorithms for Scientific Computing and Approximate Optimization*. Columbia University. https://doi.org/10.7916/D8X650C9

Hey, T. (1999). Quantum computing: an introduction. *Computing Control Engineering*, *10*(3), 105–112. https://doi.org/10.1049/cce:19990303

Karp, R. (2010). Reducibility Among Combinatorial Problems. *50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art*, 1–804. https://doi.org/10.1007/978-3-540-68279-0

Laboudi, Z., & Chikhi, S. (2012). Comparison of genetic algorithm and quantum genetic algorithm. *International Arab Journal of Information Technology*, *9*(3), 243–249.

Layeb, A. (2011). A novel quantum inspired cuckoo search for knapsack problems. *International Journal of Bio-Inspired Computation*, *3*(5), 297–305. https://doi.org/10.1504/IJBIC.2011.042260

Li, J., & Li, W. (2019). *A new quantum evolutionary algorithm in 0-1 knapsack problem*. *Communications in Computer and Information Science* (Vol. 986). Springer Singapore. https://doi.org/10.1007/978-981-13-6473-0_13

Lucas, A. (2014). Ising formulations of many NP problems. *Frontiers in Physics*, *2*, 1–14. https://doi.org/10.3389/fphy.2014.00005

Martello, S., & Toth, P. (1987). Algorithms for Knapsack Problems. *North-Holland Mathematics Studies*, *132*(C), 213–257. https://doi.org/10.1016/S0304-0208(08)73237-7

Narayanan, A. (1999). Quantum computing for beginners. *Proceedings of the 1999 Congress on Evolutionary Computation, CEC 1999*, *3*, 2231–2238. https://doi.org/10.1109/CEC.1999.785552

Nielsen, M., & Chuang, I. (2010). *Quantum Computation and Quantum Information* (Vol. 52). Cambridge, MA, USA: Cambridge University Press. https://doi.org/10.1080/00107514.2011.587535

Pisinger, D. (2005). Where are the hard knapsack problems? *Computers and Operations Research*, *32*(9), 2271–2284. https://doi.org/10.1016/j.cor.2004.03.002

Salkin, H. M., & de Kluyver, C. A. (1975). The knapsack problem: a survey*. *Naval Research Logistics*, *22*(1), 127–144. https://doi.org/10.1002/nav.3800220110

Shor, P. W. (1994). Algorithms for Quantum Computation: Discrete Logarithms and Factoring. *Proceedings 35th Annual Symposium on Foundations of Computer Science*, *59*(3), 124–134. https://doi.org/10.1016/0024-3205(96)00287-1

Vega Fernández, C. A., & Ramírez Celis, J. S. (2017). *Computación Cuántica: Implementación de Algoritmos de Short y Grover en el Computador Cuántico de IBM*. Escuela Colombiana de Ingeniería Julio Garavito.

Vickram, P., Krishna, A. S., & Srinivas, V. S. (2016). A Survey on Design Paradigms to solve 0/1 Knapsack Problem. *International Journal of Scientific & Engineering Research*, *7*(11), 112–117. Retrieved from https://www.ijser.org/researchpaper/A-Survey-on-Design-Paradigms-to-solve-0-1-Knapsack-Problem.pdf

Wang, H., Ma, L., Zhang, H., & Li, G. (2009). Quantum-inspired ant algorithm for knapsack problems. *Journal of Systems Engineering and Electronics*, *20*(5), 1012–1016. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6074538

Wang, R., Guo, N., Xiang, F., & Mao, J. (2012). An improved quantum genetic algorithm with mutation and its application to 0-1 knapsack problem. In *Proceedings of 2012 International Conference on Measurement, Information and Control* (Vol. 1, pp. 484–488). https://doi.org/10.1109/MIC.2012.6273347

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 15/09/2020
Aceitação/Acceptance: 03/11/2020

Generación automática de resúmenes extractivos de múltiples documentos basados en SGVNS e IHEA

Juan-Manuel Díaz C.[1], Anderson Buitrón P.[1], Martha-Eliana Mendoza B.[1]

juanmos@unicauca.edu.co; andersonbuitron@unicauca.edu.co; mmendoza@unicauc.edu.co

[1] Universidad del Cauca, Calle 5 # 4-70, 190003, Popayán, Colombia.

Pages: 228-244

Resumen: Debido a la gran cantidad de información existente en internet, para investigar sobre un tema es necesario leer varios documentos, lo que implica tiempo y esfuerzo para abarcar la información importante. El área de investigación de generación automática de resúmenes extractivos de múltiples documentos-GARMD, aborda este problema, con el objetivo de extraer la información más relevante de una colección de documentos. En este artículo se proponen dos algoritmos para GARMD basados en las metaheurísticas, Búsqueda de Vecindad Variable General Sesgada-SGVNS y Exploración de Hiperplanos Iterado-IHEA, que presentan buenos resultados en el problema de la mochila cuadrática. La evaluación se realizó con medidas ROUGE sobre los conjuntos de datos DUC2006 y DUC2007. En los dos métodos de decisión multicriterio IHEA-GARMD quedó en el segundo lugar, siendo un algoritmo competitivo en el estado del arte; mostrando que las metaheurísticas siguen siendo una línea de investigación prometedora en esta área de investigación.

Palabras-clave: generación de resúmenes de múltiples documentos; problema de la mochila cuadrática; búsqueda de vecindad variable; resúmenes extractivos; algoritmo de exploración de hiperplanos Iterado.

Automatic generation of extractive multi-document summarization based on SGVNS and IHEA

***Abstract*:** Due to the large amount of information on the internet, researching any topic requires reading a number of documents, which takes time and effort to cover the important information. The research area of automatic generation of extractive summaries of multiple documents-GARMD addresses this problem, with the aim of extracting the most relevant information from a document set. In this article, two algorithms for GARMD based on metaheuristics are proposed, Skewed General Variable Neighborhood Search-SGVNS and Iterated Hyperplane Exploration Algorithm-IHEA, which present good results in the quadratic backpack problem. The evaluation was performed with ROUGE measurements on the DUC2006 and DUC2007 data sets. In a unified classification, the IHEA-GARMD and SGVNS-GARMD algorithms took second and fourth place, respectively, showing that metaheuristic-based extraction of multiple document abstracts continues to be a promising line of research.

Keywords: multi-document summarization; quadratic knapsack problem; variable neighborhood search; extractive summaries; iterated hyperplane exploration algorithm.

1. Introducción

Cuando un usuario desea conocer sobre un tema específico puede encontrar muchos documentos, por lo que resulta necesario contar con un resumen que contenga la información más relevante (Ledeneva & García-Hernández, 2017), tarea que resulta compleja y costosa para un humano. El área de generación automática de resúmenes extractivos de múltiples documentos-GARMD, aborda este problema seleccionando la información más relevante contenida en una colección de documentos. En esta área se encuentran aplicaciones como: noticias extraídas de páginas web (Berlanga, Sainz, & Morales, 2011), organiza en grupos un tema en común y genera los resúmenes de cada uno de estos grupos; información extraída del dominio biomédico (Moradi, 2018), extrae conceptos biomédicos de documentos para identificar los temas principales y agrupa las oraciones de temas similares; aprendizaje automático orientado al sentimiento (Abdi, Shamsuddin, Hasan, & Piran, 2018), genera un resumen con la opinión del usuario expresada en las revisiones.

Para generar resúmenes de textos existen dos técnicas: extractiva que selecciona las oraciones más relevantes de los documentos fuente para conformar el resumen; abstractiva que obtiene una secuencia de términos que pueden provenir del texto fuente o ser creados a partir de él, pero requieren un análisis más complejo de semántica y sintaxis del contenido original (Gambhir & Gupta, 2017). Dependiendo de la cantidad de documentos a resumir, se clasifican en: un solo documento y múltiples documentos (Gambhir & Gupta, 2017). Los resúmenes de múltiples documentos son considerados como una extensión de los de un solo documento, sin embargo, el espacio de búsqueda es mucho más extenso, lo que dificulta aún más la extracción de las oraciones más relevantes.

Los algoritmos metaheurísticos abordan el problema de resúmenes automáticos como un problema de optimización global, que busca seleccionar la mejor combinación de oraciones que formarán el resumen. En los últimos años en la literatura se encuentran algoritmos de optimización para generar resúmenes de múltiples documentos con buenos resultados, como: Optimización de enjambre de gatos-CSO, (Rautray & Balabantaray, 2017); Colonia de abejas artificiales multi-Objetivo-MOABC, (Sanchez-Gomez, Vega-Rodríguez, & Pérez, 2017); Optimización de enjambre de partículas-PSO (Asgari, Masoumi, & Sheijani, 2014); cuya función objetivo usa principalmente características como cobertura, redundancia, legibilidad o cohesión.

En (Verma & Om, 2019) proponen un enfoque de optimización para GARMD con resultados prometedores, sin embargo, la función objetivo definida es la similitud de coseno entre las oraciones del resumen generado y el de referencia, lo cual permite obtener buenos resultados pero especializa el algoritmo para generar resúmenes muy similares a los de referencia. En este sentido, los resultados se sobre-ajustan ya que el entrenamiento y la prueba se realizan sobre el mismo conjunto de datos. Por esta razón en la comparación de resultados con los algoritmos propuestos no se incluye este algoritmo.

Recientemente, se propone un algoritmo de búsqueda Cucú para múltiples documentos-MDSCSA (Rautray & Balabantaray, 2018), que se ubica en los primeros puestos del estado del arte. Este algoritmo está inspirado en el comportamiento de reproducción de la especie de pájaros Cucú, que depositan sus huevos en los nidos de otras aves. Un huevo es puesto por cada Cucú en un nido aleatorio, el cual representa una solución. El número de nidos disponibles es fijo y el huevo infiltrado puede ser descubierto y eliminado por el ave anfitrión con una determinada probabilidad. Finalmente, los mejores huevos contenidos en los nidos se trasladan a la siguiente generación. La función objetivo está compuesta por características de cobertura y redundancia.

Por otro lado, el Problema de la mochila cuadrática-QKP tiene como objetivo encontrar la mejor combinación de los elementos que deben ser incluidos en una mochila con capacidad restringida. El beneficio de incluir un elemento en la mochila depende del aporte que este hace de manera individual y en pareja con cada uno de los demás elementos ya contenidos en ella, conformando de esta manera la matriz de beneficio. El problema de GARMD y QKP tienen características similares, ya que las oraciones representan los posibles elementos candidatos a ser introducidos en la mochila (resumen), y deben ser seleccionados aquellos que más aporten en conjunto a su calidad; la mochila está restringida a una cierta capacidad (cantidad máxima de palabras del resumen) con relación a los pesos de cada elemento (cantidad de palabras por oración).

En QKP de alta dimensionalidad (300, 1000 y 2000 objetos), los algoritmos metaheurísticos de Búsqueda de vecindad variable general sesgada-SGVNS (Toumi, Cheikh, & Jarboui, 2015) e Exploración de Hiperplanos Iterado-IHEA (Chen & Hao, 2017) han obtenido buenos resultados. A pesar de la similitud de QKP con GARMD, no han sido utilizados para abordar este problema. En este artículo se presenta la adaptación de los algoritmos SGVNS e IHEA al problema de GARMD.

El resto de este artículo detalla en la sección 2 el planteamiento del problema y las características de la función objetivo. La sección 3 presenta los algoritmos base usados para la adaptación al problema de GARMD. En la sección 4 se describen los algoritmos propuestos. En la sección 5 se presentan los resultados de evaluación de calidad de los resúmenes generados por los dos algoritmos propuestos y la comparación de los resultados con otros métodos del estado del arte. Finalmente, en la sección 6 se muestran las conclusiones y trabajo futuro.

2. Planteamiento del problema y función objetivo

2.1. Representación de documentos

La colección de documentos (D) es representada por el conjunto de todas las oraciones, es decir, $D = (O_1, O_2, \dots O_i, \dots, O_n)$, donde O_i es la oración en la posición i de la colección y n la cantidad de oraciones. A cada término que pertenece a la oración se le asigna un valor de acuerdo a su relevancia dentro de la oración (Manning, Raghavan, & Schutze, 2008). De esta forma, en un espacio *m*-dimensional, como se observa en la Figura 1, cada oración es representada por el vector $O_i = (t_{i1}, t_{i2}, \dots t_{ij}, \dots, t_{nm})$, donde t_{ik} es el *j*-ésimo término de la *i*-ésima oración, y m es el total de términos distintos en la colección de documentos.

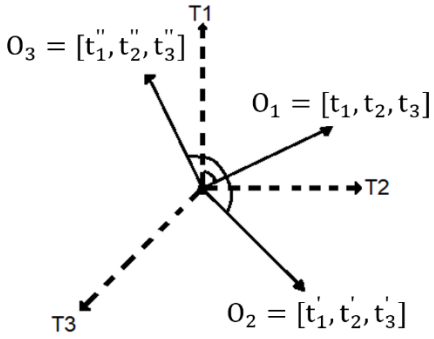


Figura 1 – Representación Modelo Espacio Vectorial

De esta representación se obtiene la matriz de términos por oración de la Ecuación (1), donde las columnas representan cada uno de los términos existentes en la colección de documentos, las filas representan las oraciones y *w* representa el peso de cada término en la oración. El valor de peso es cero cuando este término no pertenece a la oración correspondiente.

$$Matriz_{oración,término} = \begin{pmatrix} w_{1,1} & \cdots & w_{1,m} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,m} \end{pmatrix} \quad (1)$$

El esquema de ponderación TF-ISF (Manning et al., 2008), usa un factor de ponderación w_{ij} para determinar la relevancia de un término en una oración de la colección de documentos, dado por la frecuencia del término TF_{ij} multiplicado por la frecuencia inversa de la oración ISF_i, como se observa en la Ecuación (3).

$$w_{ij} = TF_{ij} * ISF_i \quad (2)$$

En este artículo se utilizó una variante del esquema de ponderación que usa el logaritmo para TF_{ij} (Lan, Tan, Su, & Lu, 2009) presentada en la Ecuación (3). La aplicación del logaritmo busca suavizar los valores obtenidos para conjuntos de gran tamaño.

$$w_{ij} = log_2(1 + f_{ij}) * log\left(\frac{N}{n_j}\right) \quad (3)$$

Donde f_{ij} es la frecuencia del término *j* en la oración *i*, *N* es cantidad de oraciones en la colección de documentos y n_j es la cantidad de oraciones donde aparece el término *j*.

2.2. Función objetivo

La función a optimizar es la de QKP (Gallo, 1980), que maximiza el beneficio de los objetos seleccionados para llenar la mochila, teniendo en cuenta que el peso total

de estos no exceda la capacidad de la misma. QKP se puede modelar de la siguiente manera: una mochila X de capacidad restringida de peso c, y n elementos candidatos $N = \{1,2,3,\ldots,n\}$ que se pueden introducir en ella, cada uno de ellos con un peso asociado p_i, un beneficio individual b_{ii} y un beneficio en pareja con cada uno de los demás elementos, b_{ij} $(1 \leq i \neq j \leq n, donde, i,j \in N)$. Un elemento en la posición i es introducido a la mochila cuando $x_i = 1$, y es removido cuando $x_i = 0$. De esta forma, la función objetivo a maximizar es la de la Ecuación (4), sujeta a la restricción de la Ecuación (5).

$$f(X) = \sum_{i=1}^{n}\sum_{j=i}^{n} b_{ij}\, x_i\, x_j \tag{4}$$

Sujeta a la restricción:

$$Peso(X) = \sum_{i=1}^{n} p_i x_i \leq c, \qquad x \in \{0,1\} \tag{5}$$

Así, la matriz de beneficio $M = \{b_{ij}\}$ es una matriz positiva de tamaño $n \times n$, simétrica, es decir, $b_{ij} = b_{ji}$.

Para transformar el problema de GARMD en QKP, la matriz de beneficio de la Ecuación (4), almacena en su diagonal $b_{ij,\ \forall i=j}$ la suma de la longitud y la cobertura de la O_i, como se muestra en la Ecuación (6); mientras que los valores $b_{ij,\ \forall i\neq j}$ corresponden a la redundancia existente entre la O_i y la oración O_j, de acuerdo a la Ecuación (7). Además, cada característica tiene relacionado un peso, que le da importancia conforme al aporte que esta hace a la calidad del resumen.

$$b_{ij} = \alpha * cobertura * (1 + \beta * longitud), \qquad i = j \tag{6}$$

$$b_{ij} = \gamma * redundancia, \qquad i > j \tag{7}$$

La Ecuación (6) define la relación de la longitud con respecto a la cobertura ($cobertura * longitud$); la cual permite dar importancia a la longitud de una oración en proporción a la cobertura de la misma.

Con respecto a la Ecuación (5), p_i corresponde a la longitud en palabras de cada O_i del resumen y c es la cantidad máxima de palabras que puede tener el resumen.

2.3. Cobertura

Esta característica busca medir la información más relevante de la colección de documentos contenida en el resumen (Gambhir & Gupta, 2017). Para esto, se usa la similitud de coseno (Manning et al., 2008) en el modelo de espacio vectorial, que permite medir la similitud entre cada oración del resumen y las oraciones de todos los documentos (Alguliev, Aliguliyev, Hajirahimova, & Mehdiyev, 2011), como se muestra en la Ecuación (8).

$$Cob(i) = Sim_{cos}(D, O_i) \quad (8)$$

Donde D es el vector de pesos de los términos de todas las oraciones de la colección de documentos; O_i es el vector de pesos de los términos de la oración i que pertenece al resumen. Cuando Cob tiene un valor cercano a uno significa que la oración i cubre gran porcentaje de información con respecto a la colección de documentos, mientras que si el valor de Cob es cercano a cero, significa que cubre poca información.

2.4. Redundancia

Esta característica busca determinar la relación existente entre las oraciones de un resumen (Alonso i Alemany & Fuentes Fort, 2003), con el objetivo de identificar que tanto se parece la información entre dos oraciones; entre más similares sean mayor será la redundancia entre ellas. Para esto, se usa la medida de similitud de coseno como muestra la Ecuación (9).

$$Red(i,j) = Sim_{cos}(O_i, O_j) \quad (9)$$

Donde *Red*(*i, j*) corresponde a la redundancia que las oraciones *i* y *j* aportan al resumen, es la similitud de cosenos entre las oraciones O_i y O_j. Mientras más similares sean las oraciones dentro del resumen, mayor será la redundancia del mismo, por esto en la matriz de beneficio este valor se resta, para de esta forma penalizar los resúmenes con oraciones que sean más redundantes.

2.5. Longitud

Esta característica considera el número de palabras de una oración para determinar su importancia. Para cada oración, la longitud normalizada se calcula como muestra la Ecuación (10).

$$longitud(i) = \frac{np_i}{npMax} \quad (10)$$

Donde np_i representa el número de palabras que componen la oración *i*, y *npMax* es el número de palabras que componen la oración más larga de toda la colección de documentos. La longitud es una proporción que dará más importancia a las oraciones más largas.

3. Algoritmos base

3.1. Algoritmo de Exploración de Hiperplanos Iterado

El Algoritmo de exploración de hiperplanos iterado-IHEA (Chen & Hao, 2017), presenta un enfoque de descomposición para resolver QKP, el cual divide el espacio de búsqueda original en un conjunto de subespacios disjuntos de hiperplanos restringidos. Un

hiperplano está definido por el conjunto de todas las soluciones que tienen la misma cardinalidad (cantidad de elementos seleccionados dentro de una solución). IHEA también recorta el espacio de búsqueda dejando fijos los elementos de mayor aporte dentro de la solución actual, para posteriormente hacer uso de un algoritmo de Búsqueda Tabú. Además, se aplica una estrategia de perturbación guiada que permite una forma global de diversificación para posiblemente examinar hiperplanos aún no explorados.

IHEA está compuesto por cuatro fases, que se observan en la Figura 2, las cuales se ejecutan un número determinado de iteraciones antes de entregar la mejor solución encontrada: (1) *Construcción*, se encarga de definir la primera solución del algoritmo usando GRASP; (2) *Descenso*, la cual se encarga de mejorar la solución con un tipo de búsqueda local de descenso de colina usando secuencialmente los operadores de cambio e intercambio. En los cuales se quita y/o introduce de forma aleatoria un elemento de la mochila, teniendo en cuenta que el que se introduce no supera la capacidad de esta. El operador cambio quita un elemento si está en la mochila, de lo contrario lo introduce; e intercambio quita e introduce un elemento. (3) *Exploración de Hiperplanos*, su principal objetivo es realizar una explotación usando un tipo de búsqueda tabú, esta fase examina cada hiperplano de forma independiente, partiendo desde el hiperplano *k* determinado por la solución entregada por descenso, y si encuentra una mejora, se pasa al siguiente hiperplano superior, si no se encuentra una mejora, se entrega la mejor solución encontrada; (4) *Perturbación*, en este método se eliminan algunos de los elementos de peor calidad, su objetivo es lograr una diversificación global permitiendo al algoritmo salirse de algún óptimo global si es alcanzado.

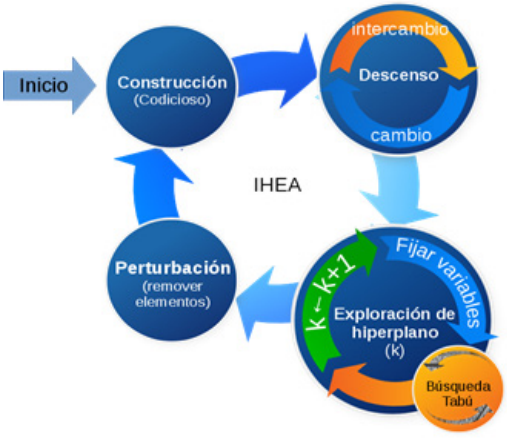


Figura 2 – Esquema general de IHEA

3.2. Búsqueda de Vecindad Variable General Sesgada

El algoritmo de Búsqueda de vecindad variable general sesgada-SGVNS (Toumi et al., 2015) mezcla un enfoque estocástico llamado sacudida, con un descenso de vecindad variable secuencial (seq-VND) de forma anidada para QKP. Este método utiliza un conjunto de estructuras de vecindario y explora los vecindarios distantes de la solución actual buscando una mejor solución y abandonando un óptimo local si es alcanzado.

SGVNS utiliza dos estructuras de vecindario como en IHEA: intercambio y cambio; se empieza con el vecindario de intercambio y cuando no encuentra una mejora se pasa al vecindario de cambio, si mejora se pasa nuevamente a intercambio, repitiendo este proceso hasta que en el vecindario cambio no se encuentre mejora. En la Figura 3 se presenta SGVNS, con sus cuatro fases: (1) *Solución inicial*, construida usando una función codiciosa sujeta a la restricción de capacidad de la mochila; (2) *Sacudida*, aplica la estructura de vecindario de cambio a la solución; (3) *Seq_VND*, aplica a la solución actual de forma iterativa, la estructura de vecindario intercambio y luego cambio, buscando una mejora; (4) *Evaluación*, usa una función de distancia para penalizar la calidad de la solución actual en proporción a la distancia de esta con la solución encontrada por seq-VND, y reemplaza la solución actual por la de seq-VND, si su valor de calidad es mejor. Finalmente, la solución actual es la entrada de la siguiente iteración.

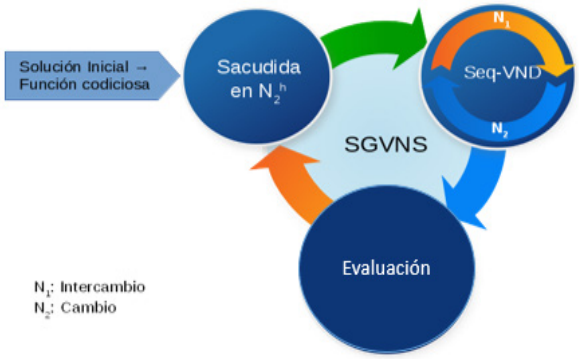


Figura 3 – Esquema general de SGVNS

4. Algoritmos propuestos

En los algoritmos propuestos, los métodos que se encuentran en negrita en el seudocódigo son las modificaciones que se realizaron a los algoritmos base.

4.1. IHEA-GARMD

En la Figura 4 se presenta la adaptación de IHEA para GARMD, llamado IHEA-GARMD, el cual ejecuta GRASP para *construir* la primera solución (línea 2), al igual que en reconstrucción (línea 29) pero se *modifica* asignando pesos en la matriz de beneficio, teniendo en cuenta que en este caso, las características de cobertura y longitud tienen mayor importancia (0.8) que la redundancia (0.2) para GARMD. Para mejorar la solución inicial y la actual, se *incluye* el algoritmo de búsqueda local *VNS* (línea 3 y línea 30) usado en (Toumi et al., 2015), utilizando los operadores cambio e intercambio como estructuras de vecindario como en SGVNS, seleccionando la primera estructura de forma aleatoria. Buscando reducir aún más el espacio de búsqueda, se *modificó* la fórmula del método *fijar los elementos* (línea 8) de mayor aporte dentro de la solución actual (resumen candidato), para aumentar el número de oraciones fijas, y se *incluyó* un método para *ignorar un porcentaje de las peores oraciones* (línea 9 y línea 23).

El ordenamiento de las oraciones en estos dos métodos se realizó calculando la densidad de cada oración, dada por su valor de calidad dividida en su peso. Luego, se hace uso de un algoritmo de *Búsqueda Tabú* (línea 12), en el cual, la condición que existía para buscar mejores soluciones con los elementos (objetos) más pesados se *cambió* por oraciones más cortas (menor peso medido por la cantidad de palabras de la oración) dando la posibilidad de que un mayor número de oraciones pueden entrar a formar parte del resumen. Posteriormente, se aplica una perturbación a la solución actual para luego ser reconstruida y entregada a la siguiente iteración. Cuando se terminan las iteraciones del algoritmo se entrega la mejor solución encontrada, usada para generar el resumen final, que estará conformado por las oraciones con mayor valor en su función de calidad.

```
1    INICIO IHEA
2      sol_global = construir()        /* construcción de la solución inicial */
3      sol_global = descender(sol_global)        /* VNS  */
4      sol_tabu = sol_local = sol_global
5      PARA iter=0 HASTA maxIter:
6        HACER: /* fase de exploración */
7          k = definir_hiperplano(sol_tabu)
8          fijar_mejores_variables(sol_tabu)
9          ignorar_peores_variables(sol_tabu)
10         PARA iter2=0 HASTA maxIter2:   /* inicio tabú */
11           PARA todos los vecinos (intercambio) no tabú de la solución actual:
12             sol_local = solución de mayor calidad y de mayor espacio
13           FIN PARA;
14           SI f(sol_local) > f(sol_global) y factible (sol_local) ENTONCES:
15             sol_global = sol_local
16           SINO
17             definir vecinos tabúes (pasos).  /* fin tabú */
18           FIN SI
19         FIN PARA;
20         SI f(sol_tabu) > f(sol_local) ENTONCES:
21           añadir_elemento(sol_tabu)       /* k = k+1 */
22         FIN SI;
23         liberar variables fijas y variables ignoradas
24       MIENTRAS f(sol_tabu) > f(sol_local) FIN MIENTRAS;
25       SI f(sol_local) > f(sol_global) ENTONCES:
26         sol_global = sol_local
27       FIN SI;
28       sol_local = perturbar(sol_local)    /* fase de perturbación */
29       sol_local = construir(sol_local)   /* fase de re-construcción */
30       sol_local = descender(sol_local)  /* fase de descenso */
31     FIN PARA
32   FIN IHEA
```

Figura 4 – Pseudocódigo algoritmo IHEA-GARMD

4.2. SGVNS-GARMD

El algoritmo SGVNS-GARMD es la adaptación de SGVNS a GARMD, como se observa en Figura 5, el cual después de generar una solución inicial incluye el mismo proceso de *búsqueda tabú* (línea 4) de IHEA-GARMD, que realiza una búsqueda exhaustiva para encontrar soluciones de mayor calidad. Esta solución es entregada a un ciclo iterativo para ser procesada, usando el *método sacudida*, en el cual se modifican los pesos de la

matriz de beneficio para evaluar la calidad de la solución, dando mayor importancia a la cobertura y longitud (diagonal de la matriz) con un peso de 0.8, que a la redundancia (parte inferior de la matriz) con 0.2. Este método aplica el operador de cambio aleatoriamente *h* veces (una iteración) para explorar soluciones en otros espacios de búsqueda, pero se adiciona una condición que permite retornar la *h*-ésima solución si esta mejora a la solución global (línea 10) sin tener que terminar la iteración, y no retornar solo la solución final de la iteración como ocurre en SGVNS. Luego el método seq-VND aplica una búsqueda local de forma secuencial, usando los operadores de intercambio y cambio respectivamente, a la solución entregada por sacudida. Finalmente, compara la mejor solución global con la obtenida por seq-VND y entrega la de mejor calidad al método sacudida en la siguiente iteración.

```
1   INICIO SGVNS-GARMD ()
2     y = solucionInicial ()
3     y_best = y
4     yp = busquedaTabu (20, y, y_best)
5     SI (yp > y best)
6       y_best = yp
7     FIN SI
8     PARA iter=0 HASTA iterMax
9       MIENTRAS que h sea menor o igual que hMax HACER
10        yp2 = sacudida (yp, cambio, h)
11        yp3 = Seq_VND (yp2)
12        SI f(yp3) > f(y_best) ENTONCES
13          y_best = yp3
14        FIN SI
15        /*Comparación de distancia*/
16        SI f(yp3) > (1- ∞ * dist (y, y")) * f(y)
17          y = y"
18          h = 1
19        SINO
20          h = h + 1
21        FIN SI
22      FIN MIENTRAS
23    FIN PARA
24  FIN SGVNS
```

Figura 5 – Pseudocódigo algoritmo SGVNS-GARMD

5. Resultados experimentales

Previo al proceso de GARMD, se realizó un preprocesamiento al texto de la colección de documentos, para limpiarlos y segmentarlos; aplicando técnicas lingüísticas, como la segmentación de oraciones, eliminación de palabras vacías, lematización, indexación y la eliminación de mayúsculas y signos ortográficos. Los algoritmos fueron implementados en C# y Java, y ejecutados en una CPU Intel Core i7 con 8 núcleos y 8 GB de RAM.

Las medidas de ROUGE (Recall-Oriented Understudy for Gisting Evaluation) (Ledeneva & García-Hernández, 2017), son aceptados por la comunidad científica como métricas para la evaluación de resúmenes automáticos. En esta investigación se utilizaron las métricas Rouge-1 y Rouge-2, sobre los conjuntos de datos de DUC2006 y DUC2007.

Los pesos de las características de la función objetivo para cada algoritmo son presentados en la Tabla 1.

Algoritmo	Redundancia	Cobertura	Longitud
	α	β	δ
IHEA	0.38	0.38	0.24
SGVNS	0.35	0.40	0.25

Tabla 1 – Definición de pesos de características de la función objetivo

Además del esquema de ponderación TF-ISF presentado en la Ecuación (3), se revisaron otras dos representaciones que son modificaciones de esta ecuación, estas son: *frecuencia bruta* del término TF_{ij} (Lan et al., 2009) que cuenta el número de veces que el término *j* aparece en la oración *i*, como se observa en la Ecuación (11); y *frecuencia normalizada* (Manning et al., 2008) que divide la frecuencia de aparición del término *j* en la oración *i*, sobre la frecuencia máxima de los demás términos en la oración *i*, como se presenta en la Ecuación (12).

$$W_{ij} = f_{ij} * \log\left(\frac{N}{ni}\right) \tag{11}$$

$$W_{ij} = \frac{f_{ij}}{Max\ (f_i)} * \log\left(\frac{N}{n_i + 1}\right) \tag{12}$$

Donde f_{ij} es la frecuencia de aparición del término *j* en la oración *i* de la colección de de documentos, *N* es el número total de oraciones de la colección de documentos, n_i es el número de oraciones en las que aparece el término *j* y $max\ (f_i)$ es la máxima frecuencia de aparición de términos en la oración *i*.

También se usó la representación ATIRE BM25 (Trotman, Puurula, & Burgess, 2014), que es una función de clasificación usada para evitar números negativos, observada en la Ecuación (13).

$$W_{ij} = \frac{(k + 1) * f_{ij}}{k\left(1 - b + b\left(\frac{|O_i|}{AVG\ (Coleccion)}\right)\right) + f_{ij}} * ln\left(\frac{N}{n_i}\right) \tag{13}$$

Donde O_i es la oración que contiene al término *j*, $AVG\ (Coleccion)$ es la longitud promedio de las oraciones en la colección de documentos; las variables $k = 2$ y $b = 0.75$ permiten ajustar la función a las características específicas de cada colección de documentos.

Los resultados obtenidos de los algoritmos propuestos con los diferentes esquemas de ponderación (Esq. Pond.) utilizados sobre los conjuntos de datos de DUC2006 y DUC2007, se muestran en la Tabla 2 y Tabla 3, respectivamente. Por lo tanto, se escoge ATIRE BM25 para el algoritmo SGVNS y F-Logaritmo para IHEA.

Algoritmo	Esq. Pond.	DUC2006					
		Rouge-1			Rouge-2		
		Recuerdo	Precisión	F	Recuerdo	Precisión	F
SGVNS	BM25	**0,41984**	**0,41557**	**0,41765**	**0,08816**	**0,08725**	**0,08769**
	F-Log.	0,40586	0,40074	0,40325	0,08124	0,08026	0,08074
	F-Norm.	0,41757	0,41352	0,41549	0,08599	0,08522	0,0856
	F-Bruta	0,41532	0,41079	0,41299	0,08685	0,08594	0,08638
IHEA	BM25	0,4134	0,4156	0,0859	0,08512	0,085529	0,14304
	F-Log.	**0,42057**	**0,41569**	**0,41807**	**0,08787**	**0,08689**	**0,08737**
	F-Norm.	0,41925	0,41447	0,4168	0,08735	0,08637	0,08685
	F-Bruta	0,41471	0,40972	0,41215	0,08617	0,08515	0,08565

Tabla 2 – Resultados de esquemas de ponderación DUC2006

Algoritmo	Esq. Pond.	DUC2007					
		Rouge-1			Rouge-2		
		Recuerdo	Precisión	F	Recuerdo	Precisión	F
SGVNS	BM25	**0,44027**	**0,43308**	**0,43652**	**0,11013**	**0,10843**	**0,10924**
	F-Log.	0,43966	0,43219	0,43575	0,11124	0,10949	0,11032
	F-Norm.	0,43714	0,42931	0,43304	0,10681	0,105	0,10586
	F-Bruta	0,43776	0,43004	0,43374	0,11012	0,10828	0,10916
IHEA	BM25	0,44069	0,4336	0,43697	0,10651	0,10503	0,10574
	F-Log.	**0,44458**	**0,43695**	**0,4406**	**0,11048**	**0,10874**	**0,10957**
	F-Norm.	0,43935	0,43121	0,4351	0,10422	0,10244	0,10329
	F-Bruta	0,43962	0,43197	0,43563	0,10683	0,10512	0,10594

Tabla 3 – Resultados de esquemas de ponderación DUC2007

Los algoritmos propuestos fueron comparados con los algoritmos metaheurísticos del estado del arte MDSCSA, PSO y CSO. La Tabla 4 y Tabla 5 muestran los resultados obtenidos en las medidas ROUGE por los algoritmos propuestos IHEA-GARMD y SGVNS-GARMD y algoritmos del estado del arte sobre los conjuntos de datos DUC 2006 y DUC 2007, respectivamente. Para cada métrica ROUGE-1 y ROUGE -2 se presentan las medidas de recuerdo, precisión y medida F, las cuales se dividen en dos columnas: la primera con el valor obtenido por cada algoritmo, y la segunda con la clasificación de cada algoritmo, colocando en negrita el valor que obtuvo el mejor resultado.

Algoritmo	ROUGE-1						ROUGE -2					
	Recuerdo		Precisión		F		Recuerdo		Precisión		F	
SGVNS-GARMD	0.41984	5	0.41557	3	0.41765	4	0.08816	3	0.08725	2	0.08769	3
IHEA-GARMD	0.42057	4	0.41569	2	0.41807	3	0.08787	4	0.08689	3	0.08737	4
MDSCSA	0.43655	2	**0.4258**	**1**	**0.4311**	**1**	**0.12346**	**1**	**0.16129**	**1**	**0.13986**	**1**

Algoritmo	ROUGE-1						ROUGE -2					
	Recuerdo		Precisión		F		Recuerdo		Precisión		F	
PSO	**0.44151**	**1**	0.38491	5	0.41127	5	0.08255	5	0.07469	5	0.0784	5
CSO	0.43098	3	0.41520	4	0.4229	2	0.09950	2	0.08271	4	0.09033	2

Tabla 4 – Resultados DUC2006

Algoritmo	ROUGE -1						ROUGE -2					
	Recuerdo		Precisión		F		Recuerdo		Precisión		F	
SGVNS-GARMD	0.44027	5	0.43308	2	0.43652	2	0.11013	2	0.10843	2	0.10924	2
IHEA-GARMD	0.44458	4	**0.43695**	**1**	**0.4406**	**1**	**0.11048**	**1**	**0.10874**	**1**	**0.10957**	**1**
MDSCSA	0.45830	2	0.39510	3	0.42430	3	0.10930	3	0.09824	3	0.10340	3
PSO	0.44679	3	0.37825	5	0.40967	5	0.08410	5	0.06970	5	0.07620	5
CSO	**0.46158**	**1**	0.38662	4	0.42070	4	0.09240	4	0.08590	4	0.08903	4

Tabla 5 – Resultados DUC2007

El algoritmo IHEA-GARMD en el conjunto de datos DUC2007 ocupa el primer puesto en todas las medidas, excepto ROUGE-1 Recuerdo (ocupando el cuarto). Y en el conjunto de datos DUC2006, MDSCSA ocupa el primer puesto en todas las medidas, excepto ROUGE-1 Recuerdo (ocupando el segundo).

Debido a que ningún algoritmo ocupa el primer puesto en todas las medidas, se hace uso de dos métodos de decisión multicriterio. El primero y más utilizado en el área de generación automática de resúmenes es la clasificación unificada (Aliguliyev, 2009) y, el método de ponderación lineal usado en (Ríos, Sánchez, & González, 2019; Zamudio-Sánchez, Romo-Lozano, Borja-De La Rosa, Martínez-Gómez, & Ávalos-Vargas, 2017).

La clasificación unificada tiene en cuenta la posición ocupada por cada algoritmo en cada medida independiente. Esta clasificación se realiza de acuerdo a la Ecuación (14).

$$Clasificacion(algoritmo) = \sum_{r=1}^{N} \frac{(N - r - 1) * R_r}{N} \tag{14}$$

Donde R_r indica el número de veces que el algoritmo clasifica en la posición r y N representa la cantidad de algoritmos a comparar, en este caso $N = 5$.

Los resultados de la clasificación unificada se muestran en la Tabla 6 , el valor de la columna clasificación se calcula de acuerdo con la Ecuación (14). El mayor puntaje ocupa la primera posición, el siguiente la segunda posición y así sucesivamente.

Algoritmo	1	2	3	4	5	Clasificación
MDSCSA	5	3	4	0	0	9,600
IHEA-GARMD	5	1	2	4	0	8,600
CSOS	1	3	1	7	0	6,800

Algoritmo	1	2	3	4	5	Clasificación
SGVNS-GARMD	0	6	3	1	2	7,400
PSOS	1	0	1	0	10	3,600

Tabla 6 – Clasificación unificada de los algoritmos

El método de ponderación lineal utiliza la Ecuación (15).

$$ponderacion\_lineal(A_i) = \sum_{j=1}^{m} R_{ij} * W_j = \sum_{j=1}^{m} \frac{r_{ij}}{\sum_{i=1}^{n} r_{ij}} * \frac{w_j}{\sum_{j=1}^{m} w_j} \quad (15)$$

Donde W_j es el peso normalizado asignado al criterio j (para este trabajo es el mismo peso W_j para cada criterio); R_{ij} es el valor registrado de la alternativa i con respecto al criterio j; n es la cantidad total de algoritmos; m es la cantidad total de criterios. Así, j representa cada criterio, que en este caso hace referencia: Recuerdo, Precisión y F, de cada medida ROUGE-1 y ROUGE-2 de los conjuntos de datos DUC2006 y DUC2007.

Los resultados obtenidos por medio de la ponderación lineal se muestran en la Tabla 7, el valor de la columna calificación es calculado de acuerdo con la Ecuación (15). El mayor puntaje ocupa la primera posición, el siguiente la segunda posición y así sucesivamente.

Algoritmo (A_i)	Calificación (A_i)
MDSCSA	0,2268
IHEA-GARMD	0,2024
SGVNS-GARMD	0,2019
CSOS	0,1921
PSOS	0,1767

Tabla 7 – Clasificación por ponderación line

Como se observa en la Tabla 6 y Tabla 7, no existe una diferencia considerable entre IHEA-GARMD que ocupa el segundo puesto y MDSCSA que ocupa el primer puesto, mostrando que IHEA-GARMD es un algoritmo competitivo en el estado del arte. El algoritmo IHEA-GARMD ejecuta 1.133 evaluaciones de la función objetivo, pero no fue posible comparar este valor con MDSCSA ya que no presenta la cantidad de iteraciones que ejecuta este algoritmo.

6. Conclusiones

Este artículo propone dos algoritmos de optimización para el problema de generación automática de resúmenes de múltiples documentos-GARMD, llamados IHEA-GARMD y SGVNS-GARMD. En estos algoritmos se usa la función objetivo de QKP, la cual es adaptada a GARMD incluyendo características importantes de este problema en la

matriz de beneficio de la siguiente forma: la cobertura de las oraciones con respecto a la colección de documentos y la longitud de la oración en la diagonal, y en la parte inferior de la matriz la redundancia de cada oración con las demás.

En IHEA-GARMD se realizaron modificaciones para partir de una mejor solución inicial y actual (en las iteraciones): asignando pesos en la matriz de beneficio a las características de cobertura y longitud (0.8), y redundancia (0.2), y aplicando el algoritmo de búsqueda local VNS; también se redujo el espacio de búsqueda aumentando el número de oraciones que se dejan fijas en la solución actual (mayor aporte) e ignorando las de menor peso de acuerdo al valor de densidad de cada oración. Además, la búsqueda tabú que se utiliza para la solución actual se modificó para buscar mejores soluciones con oraciones más cortas (menor peso medido por la cantidad de palabras) para dar posibilidad a un mayor número de oraciones en el resumen. Por su parte, SGVNS-GARMD también se modificó para partir de una mejor solución inicial con la misma búsqueda Tabú de IHEA-GARMD y en el método de sacudida que se usa para la solución actual se modificó el proceso intermedio para que al encontrar una solución que mejore a la global, salir del método con esta solución.

Los resúmenes generados por los algoritmos propuestos son comparados con otros del estado del arte, sobre los conjuntos de datos de DUC2006 y DUC2007, usando las métricas ROUGE-1 y ROUGE-2. Los resultados obtenidos con los dos métodos de decisión multicriterio muestran que IHEA-GARMD ocupan el segundo lugar, siendo un algoritmo competitivo en el estado del arte.

Con respecto a trabajos futuros, se espera incluir nuevas características dentro de la función objetivo, tales como: sustantivos propios u otras características relevantes para el contenido de los documentos. También aplicar estos algoritmos en otros conjuntos de noticias.

Financiamiento

Artículo de investigación derivado del proyecto "Algoritmos para la generación automática de resúmenes extractivos de múltiples documentos basados en los procedimientos de Búsqueda de Vecindad Variable y Exploración de Hiperplanos" financiado parcialmente por la Universidad del Cauca.

Referencias

Abdi, A., Shamsuddin, S. M., Hasan, S., & Piran, J. (2018). Machine learning-based multi-documents sentiment-oriented summarization using linguistic treatment. *Expert Systems with Applications*, *109*, 66–85. https://doi.org/10.1016/j.eswa.2018.05.010

Alguliev, R. M., Aliguliyev, R. M., Hajirahimova, M. S., & Mehdiyev, C. A. (2011). MCMR: Maximum coverage and minimum redundant text summarization model. *Expert Systems with Applications*, *38*(12), 14514–14522. https://doi.org/10.1016/j.eswa.2011.05.033

Aliguliyev, R. M. (2009). Performance evaluation of density-based clustering methods. *Information Sciences*, *179*(20), 3583–3602. https://doi.org/10.1016/j.ins.2009.06.012

Alonso i Alemany, L., & Fuentes Fort, M. (2003). *Integrating cohesion and coherence for automatic summarization*. 1. https://doi.org/10.3115/1067737.1067739

Asgari, H., Masoumi, B., & Sheijani, O. S. (2014). Automatic text summarization based on multi-agent particle swarm optimization. *2014 Iranian Conference on Intelligent Systems, ICIS 2014*, 0–4. https://doi.org/10.1109/IranianCIS.2014.6802592

Berlanga, E. G., Sainz, C. P., & Morales, L. P. (2011). *Agrupación y resúmenes noticias de periódicos web*. Retrieved from https://eprints.ucm.es/13046/1/Agrupación_y_resumenes_multidocumento_de_noticias_de_periódicos_web.pdf

Chen, Y., & Hao, J.-K. (2017). An iterated "hyperplane exploration" approach for the quadratic knapsack problem. *Computers & Operations Research*, *77*, 226–239. https://doi.org/10.1016/j.cor.2016.08.006

Gallo, G. (1980). The Quadratic Knapsack Problem. https://doi.org/10.1007/978-3-540-24777-7_12

Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, *47*(1), 1–66. https://doi.org/10.1007/s10462-016-9475-9

Lan, M., Tan, C. L., Su, J., & Lu, Y. (2009). Supervised and traditional term weighting methods for automatic text categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(4), 721–735. https://doi.org/10.1109/TPAMI.2008.110

Ledeneva, Y., & García-Hernández, R. A. (2017). *Generación automática de resúmenes Retos, propuestas y experimentos* (1a ed.; I. Literario, Ed.). Toluca.

Manning, C. D., Raghavan, P., & Schutze, H. (2008). Introduction to Information Retrieval. In O. edition (c) 2009 C. UP (Ed.), *Endocrinology* (Vol. 38). Cambridge: Cambridge University Press.

Moradi, M. (2018). CIBS: A biomedical text summarizer using topic-based sentence clustering. *Journal of Biomedical Informatics*, *88*(February), 53–61. https://doi.org/10.1016/j.jbi.2018.11.006

Rautray, R., & Balabantaray, R. C. (2017). Cat swarm optimization based evolutionary framework for multi document summarization. *Physica A: Statistical Mechanics and Its Applications*, *477*, 174–186. https://doi.org/10.1016/j.physa.2017.02.056

Rautray, R., & Balabantaray, R. C. (2018). An evolutionary framework for multi document summarization using Cuckoo search approach: MDSCSA. *Applied Computing and Informatics*, *14*(2), 134–144. https://doi.org/10.1016/j.aci.2017.05.003

Ríos, R. C. G., Sánchez, C. N. G., & González, E. L. R. (2019). Application of the scoring method for integral customer classification. *Ingeniare*, *27*(3), 375–382. https://doi.org/10.4067/S0718-33052019000300375

Sanchez-Gomez, J. M., Vega-Rodríguez, M. A., & Pérez, C. J. (2017). Extractive multi-document text summarization using a multi-objective artificial bee colony optimization approach. *Knowledge-Based Systems*, *0*, 1–8. https://doi.org/10.1016/j.knosys.2017.11.029

Toumi, S., Cheikh, M., & Jarboui, B. (2015). 0 – 1 Quadratic Knapsack Problem solved with VNS algorithm. *Electronic Notes in Discrete Mathematics*, *47*(July), 269–276. https://doi.org/10.1016/j.endm.2014.11.035

Trotman, A., Puurula, A., & Burgess, B. (2014). Improvements to BM25 and language models examined. *ACM International Conference Proceeding Series*, *27-28-Nove*, 58–65. https://doi.org/10.1145/2682862.2682863

Verma, P., & Om, H. (2019). MCRMR: Maximum coverage and relevancy with minimal redundancy based multi-document summarization. *Expert Systems with Applications*, *120*, 43–56. https://doi.org/10.1016/j.eswa.2018.11.022

Zamudio-Sánchez, F. J., Romo-Lozano, J. L., Borja-De La Rosa, A., Martínez-Gómez, G., & Ávalos-Vargas, A. (2017). Model of selection and evaluation for graduate applicants in forest sciences. *Revista Chapingo, Serie Ciencias Forestales y Del Ambiente*, *23*(3), 353–367. https://doi.org/10.5154/r.rchscfa.2016.12.074

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 14/08/2020
Aceitação/Acceptance: 28/10/2020

Vari::Ation - De modelos de características a modelos de configuración

Alonso Bobadilla Montoya[1], Samuel Sepúlveda Cuevas[1]

a.bobadilla02@ufromail.cl; samuel.sepulveda@ufrontera.cl

[1] Departamento de Ciencias de la Computación e Informática, Universidad de La Frontera, 4811230, Temuco, Chile.

Pages: 245-258

Resumen: Las líneas de productos de software son un enfoque centrado en la reutilización, basado en la definición de similitudes y variabilidades del software. Los modelos de características (MC) son los más usados entre las notaciones para gestionar la variabilidad. En cambio, los modelos de configuración resultan de seleccionar características siguiendo un conjunto de reglas establecidas. El objetivo de este trabajo es desarrollar una herramienta que permita configurar MC mediante una aplicación web. Se utilizó la metodología iterativa e incremental para el desarrollo, y un conjunto de pruebas funcionales para evaluar la herramienta. Entre los principales resultados se tiene que la herramienta configura de manera satisfactoria los MC importados y pasó la validación efectuada. Es posible concluir que la herramienta cumple los objetivos de diseño, pues permite exportar los archivos necesarios para que sean consumidos en una siguiente etapa.

Palabras-clave: Modelo de Configuración; Modelo de Características; Líneas de Productos de Software; Herramienta.

Vari::Ation - From feature models to configuration models

Abstract: Software product lines are a reuse-centric approach, based on the definition of software similarities and variabilities. Feature models (FM) are the most used among the notations to manage variability. Instead, configuration models result from selecting features following a set of established rules. The objective of this work is to develop a tool that allows FM to be configured through a web application. The iterative and incremental methodology was used for development, and a set of functional tests to evaluate the tool. Among the main results is that the tool successfully configures the imported FM and passed the validation carried out. It is possible to conclude that the tool meets the design objectives, since it allows the necessary files to be exported so that they can be consumed in a next stage.

Keywords: Configuration Model; Feature Model; Software Product Lines; Tool

1. Introducción

Las Líneas de Productos de Software (LPS) son un enfoque que permite aumentar la productividad sin sacrificar la calidad. Las LPS son un conjunto de sistemas de software intensivos que comparten un conjunto común de características y se desarrollan para un dominio específico mediante un proceso definido (Clements & Northrop, 2001). Entre los beneficios del uso de LPS están la reducción en los ciclos de desarrollo del producto, aumento de la productividad en un orden de magnitud y una disminución en los costos de desarrollo (Ahmed & Capretz, 2008) (McGregor, Muthig *et al*., 2010).

El desarrollo de las LPS se divide en dos etapas. Primera etapa, llamada Ingeniería de Dominio, se refiere a las actividades relacionadas con identificar elementos comunes y variantes entre los miembros de la familia de productos e implementar un conjunto de artefactos de software compartidos (p. ej. componentes, requisitos, etc.). Dichos artefactos se construyen de tal manera que los puntos en común se pueden explotar, al mismo tiempo que se preserva la variabilidad. La segunda etapa, la Ingeniería de Productos, se refiere a las actividades relacionadas con la derivación de productos, es decir, la construcción de productos individuales utilizando un subconjunto de los artefactos de software compartidos (Deelstra, Sinnema & Bosch, 2004).

El modelado de requerimientos es una de las partes más importantes de la Ingeniería de Dominio. Este modelado implica el análisis de elementos comunes y variables de una LPS en términos de características del producto. En un trabajo previo (Sepúlveda, Cravero & Cachero, 2016), sobre los lenguajes utilizados para modelar los requerimientos de una LPS, se determinó que el formalismo más utilizado para representarlo son los Modelos de Características (MC).

Por otro lado, un producto individual de una familia está descrito por una selección de características que cumplen las reglas del modelo. Este conjunto de características seleccionadas para un producto individual se le conoce como Modelo de Configuración (Meliá, Pardillo & Cachero, 2010).

En (Deelstra, Sinnema & Bosch, 2004) se indica que la Ingeniería de Dominio recibe mucha atención de la comunidad de LPS, dejando de lado el soporte metodológico para la Ingeniería de Productos. Esto se ve ratificado en (Heradio, Perez-Morago et al., 2016), donde se realiza un análisis de veinte años de las LPS y se puede constatar que existe un aumento en el interés de investigar la Ingeniería de Productos, donde una de las temáticas que destaca es la *derivación de productos*.

El auge de estas temáticas lleva a la confección de una propuesta enfocada en primera instancia en la Ingeniería de Productos, para contribuir al área de *modelado de la configuración* y *derivación de productos*. La propuesta se divide en dos etapas. La primera corresponde a la definición de un dominio problema para la LPS, obteniendo así un MC validado. Luego, se desarrolla la primera herramienta del conjunto de herramientas xx::Ation, llamadaVari::Ation, que permite la configuración de un producto particular mediante la selección de características (respetando las reglas del MC). Finalmente, se exporta dicha configuración mediante un modelo de configuración para ser usado en la siguiente etapa. La segunda etapa corresponde a un trabajo en desarrollo y tiene como objetivo la construcción de aplicaciones web particulares de manera automática,

basándose en las reglas definidas por la configuración de un MC para seleccionar los componentes (previamente construidos) que forman parte del producto final. En la Figura 1 se muestra un resumen de lo descrito anteriormente.

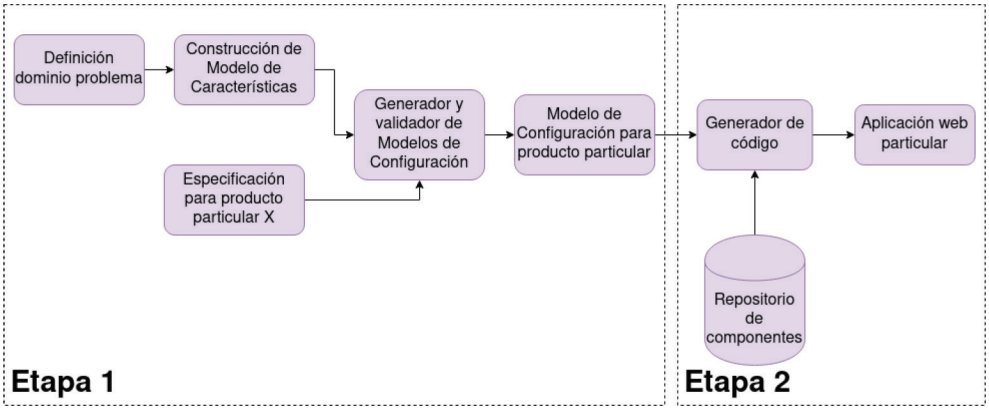


Figura 1 – Etapas de la propuesta.

Los principales aportes de esta investigación están dados por la definición de un MC para tiendas electrónicas adaptado de la literatura, el desarrollo de una aplicación web en el *framework* Angular (Vari::Ation), que permite configurar MC provenientes de una herramienta en particular llamada FMxx[1] y que es la continuación de una línea de trabajo. Además, este trabajo entrega la base para el desarrollo de la siguiente etapa, donde se abordará la ingeniería de dominio mediante la implementación de la variabilidad (repositorio de componentes) y la derivación de productos (generador de código).

El resto del artículo se estructura de la siguiente forma. Sección 2 presenta conceptos relacionados con LPS, MC y Modelos de Configuración. Sección 3 describe la metodología seguida para el desarrollo de este trabajo. La Sección 4 presenta y discute los principales resultados. Sección 5 discute algunos trabajos relacionados. Finalmente, la Sección 6 presenta las conclusiones y trabajos futuros.

2. Antecedentes generales

La gestión de la variabilidad es uno de los conceptos principales en las LPS, ya que corresponde a la definición de las partes del software que están presentes en todos los productos y aquellas partes que pueden cambiar de un producto a otro según el contexto particular (Galster, Weyns *et al.,* 2013).

Para representar la variabilidad, se utilizan los MC. Estos tienen una estructura de tipo árbol, donde su nodo raíz representa a la familia de productos y las características están

[1] http://estocolmo.ceisufro.cl/

organizadas a lo largo del árbol. Dichas características pueden ser ensambladas para dar origen a productos de software particulares (Czarnecki & Wasowski, 2007).

La propuesta dada a conocer en el presente trabajo corresponde a la continuación de una línea de trabajo, que comenzó con la creación de un meta modelo llamado UFM-MM y un conjunto de restricciones OCL para representar MC (Sepúlveda, Cares & Cachero, 2012). Posteriormente, fue desarrollada la herramienta FMxx, utilizando ADOxx (plataforma de desarrollo y configuración de meta modelado para implementar métodos de modelado), que permite representar de forma gráfica los MC respetando tanto las relaciones del meta modelo UFM-MM, como las restricciones OCL definidas (Esperguel & Sepúlveda, 2016). En la Figura 2 se da a conocer un MC creado con esta herramienta y que representa una parte de los lineamientos generales para la presentación de una tesis de pregrado. En este se puede observar que el informe debe contar con un resumen, una introducción y los objetivos, pero la dedicatoria y los agradecimientos son opcionales.

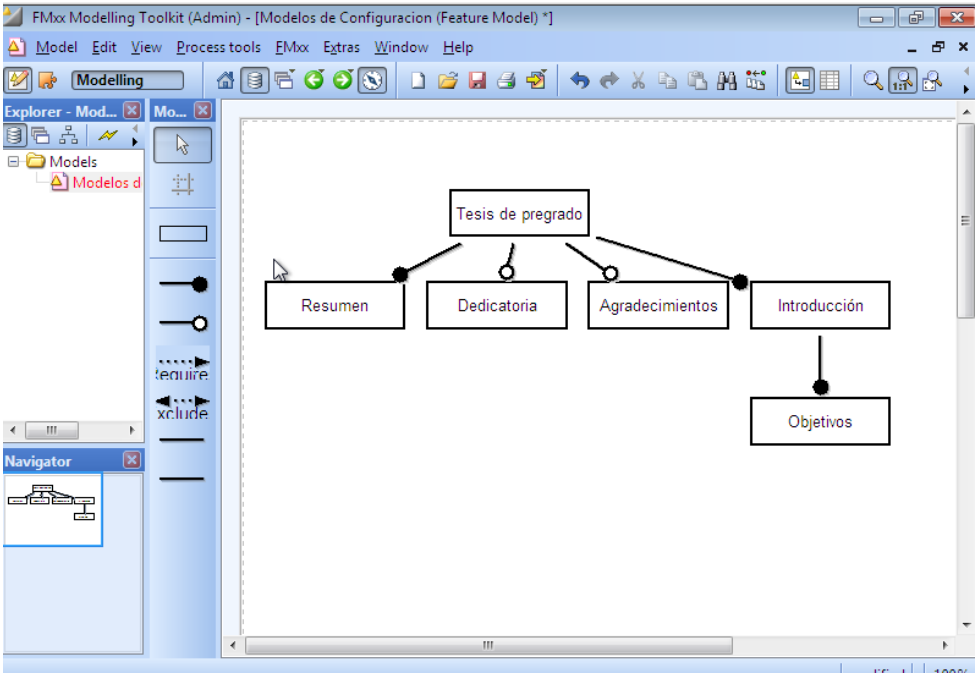


Figura 2 – MC en FMxx

Luego se realizó otra iteración para solucionar problemas de FMxx, como la restricción que implica sólo ser compatible con el sistema operativo Microsoft Windows o no ser capaz de exportar los modelos en formato SVG. De esta manera se crea la herramienta FMxx:Visor, que sirve para trabajar los MC de manera colaborativa y dejando un historial de cambios y comentarios para gestionar su trazabilidad (Esperguel, Sepúlveda & Monsalve, 2017). Para el presente trabajo se utiliza una funcionalidad de esta

herramienta llamada XML2SVG, que permite transformar un MC en formato XML exportado desde FMxx, a una imagen en formato SVG. Esta funcionalidad es importante, ya que es la base para realizar la configuración de MC con Vari::Ation.

3. Metodología

Previo al desarrollo de Vari::Ation, se construye un MC sin seguir una metodología en particular, pero se lleva a cabo siguiendo una estructura sistemática, de reuniones semanales, trabajo autónomo y trabajo colaborativo remoto del equipo de trabajo.

Para el desarrollo de Vari::Ation, se utilizó la metodología de desarrollo iterativa e incremental, que es un enfoque para crear software en el que el ciclo de vida está compuesto por varias iteraciones en secuencia. Cada iteración es un mini proyecto autónomo compuesto por actividades tales como análisis de requerimientos, diseño, programación y pruebas. El objetivo para el final de una iteración es una versión de iteración, un sistema parcialmente completo estable, integrado y probado (Larman, 2004).

El trabajo se desarrolla siguiendo las indicaciones de (Martin, 1999). Martin indica en primer lugar, que, al iniciar un desarrollo iterativo e incremental, el primer objetivo es dividir el proyecto en entregables, llamados *slices* o rebanadas. Luego de realizar un análisis de alto nivel del proyecto completo, se divide el desarrollo de la manera que se ve en la Figura 3. La metodología indica que la rebanada crítica debe ser abordada lo antes posible, en este caso la segunda iteración es la crítica, pero es indispensable realizar la primera para poder avanzar con esta. Por el contrario, la última iteración es la menos crítica.

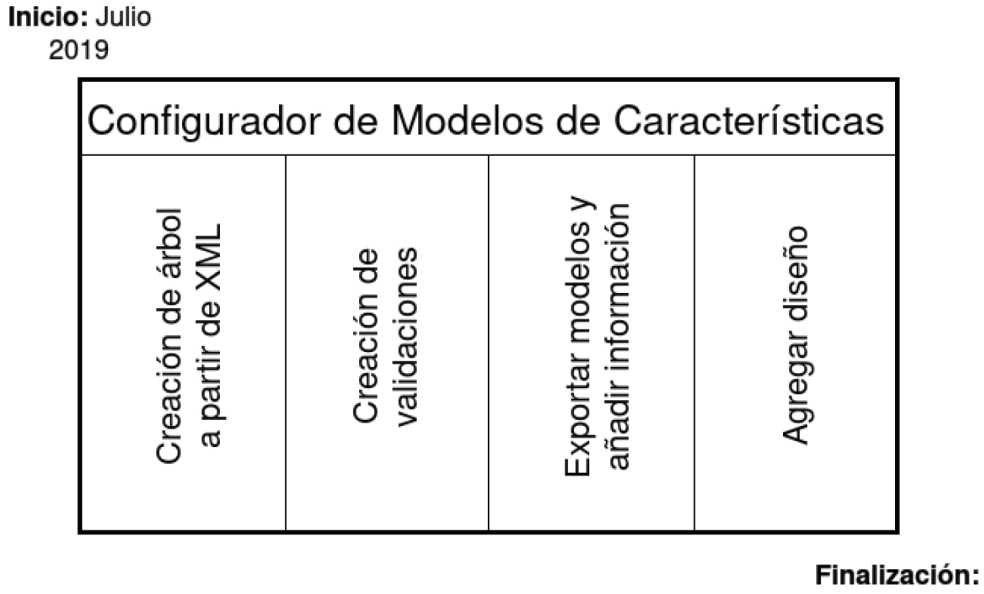


Figura 3 – Etapas de la propuesta

Así como el desarrollo de la herramienta, es importante también su validación. Para esto se realizan una serie de pruebas siguiendo la propuesta de (Di Lucca, Fasolino et al., 2002), para las pruebas funcionales en aplicaciones web. Dicha propuesta plantea la creación de tablas de decisión para probar la aplicación. Una tabla de decisión está dividida en dos partes: la sección de condiciones, que está compuesta de la lista de condiciones y sus combinaciones, y, en segundo lugar, está la sección de acción, que es una lista de respuestas que se producirán cuando las combinaciones de acciones correspondientes sean verdaderas. Siguiendo estas pautas, se probó el MC creado y otros MC de la literatura. Un ejemplo de una tabla de decisión para la primera parte de Vari::Ation es la mostrada en la Tabla 1.

Variante	Sección de entrada		Sección de salida
	Variables de entrada	**Variables de acción**	**Resultados esperados**
1	*MC en formato XML de FMxx*	*Click "Generar árbol"*	*Árbol desplegado con checkbox Modelo en formato SVG* *Resumen de modelo*
2	-	*Click "Generar árbol"*	*No hay resultado, falta seleccionar archivo*
3	*MC en formato de FMxx*	-	*Modelo cargado*
4	*Archivo de otro formato*	*Click "Generar árbol"*	*No hay resultado*
5	*Archivo XML*	*Click "Generar árbol"*	*No hay resultado*

Tabla 1 – Tabla de decisión: Importar MC a la aplicación web

Tanto el desarrollo de Vari::Ation como la realización de las pruebas, tiene el soporte tecnológico de Gitlab[2], donde está el código y las iteraciones realizadas, a modo de documentación sobre cómo fue realizado el proceso hasta llegar a los resultados finales.

4. Resultados y discusión

Los resultados obtenidos están divididos en tres partes. En la Sección 4.1 se presenta el MC creado con la herramienta FMxx. Luego, en la Sección 4.2 se muestra la herramienta Vari::Ation. Finalmente, en la Sección 4.3 se detallan las pruebas realizadas a la herramienta.

4.1. Resultados modelado MC

El principal resultado obtenido en esta etapa es la definición de un MC para tiendas electrónicas usando la herramienta FMxx. Este modelo fue adaptado de la literatura, a partir de las descripciones realizadas por (Mazo, 2018). El modelo resultante cuenta con 21 características, donde 7 de ellas son características núcleo de cada posible producto, y tiene un total de 868 posibles configuraciones, según el análisis realizado con la

[2] https://gitlab.com/abobadilla02/configurationmodel

aplicación S.P.L.O.T.[3]. En la Figura 4 se muestra el MC obtenido y que se usa para el resto de la propuesta.

Junto con el modelado en FMxx, el resultado más importante es el XML exportado desde esta herramienta. Dicho archivo XML contiene el MC que es el parámetro de entrada para la aplicación web desarrollada y se detalla en la sección 4.2.

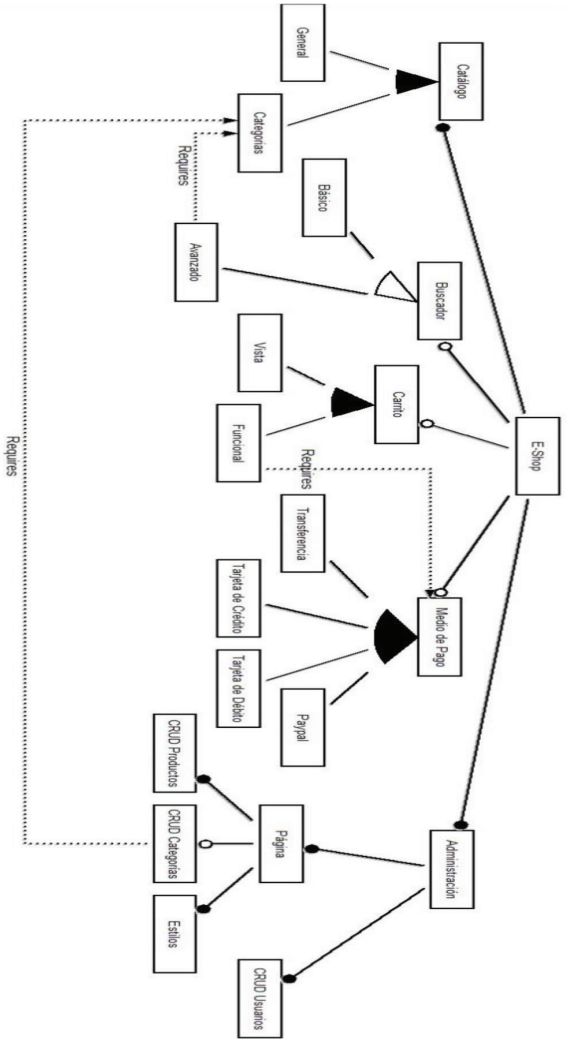


Figura 4 – MC para tiendas electrónicas adaptado de (Mazo, 2018).

[3] http://52.32.1.180:8080/SPLOT/feature_model_edition.html | LPS tienda virtual final | ID: 596 | Alonso Bobadilla

4.2. Resultados herramienta para configurar MC (Vari::Ation)

Luego de realizar tres de las cuatro iteraciones planificadas inicialmente, dónde la última se encuentra actualmente en desarrollo, se logró desarrollar tanto el algoritmo como la implementación de un software que puede importar un MC creado desde la herramienta FMxx, para luego desplegarlo en formato árbol junto a una imagen de referencia en formato SVG. El despliegue en formato árbol viene acompañado de *checkboxs* que permiten seleccionar y deseleccionar las características según la configuración que se busca realizar para un producto en particular. Las configuraciones se van validando en tiempo real con el objetivo de obtener siempre una configuración válida con el mínimo esfuerzo y sin errores. En la Figura 5, se muestran tres posibles configuraciones usando la herramienta, para el MC presentado en la Figura 4.

Otra característica de Vari::Ation es que permite exportar un modelo de configuración en formato JSON con los atributos más importantes de cada característica y si ésta fue seleccionada o no. Además, se puede exportar un segundo JSON, pero esta vez que contiene una versión simplificada del modelo de configuración anterior con las reglas que serán usadas como parámetro de entrada en la siguiente etapa del proyecto, donde se utilizarán las reglas para determinar qué componentes deben ser agregados al proyecto y cuáles deben ser excluidos. En la Figura 6 se muestra un JSON con una configuración válida utilizando Vari::Ation.

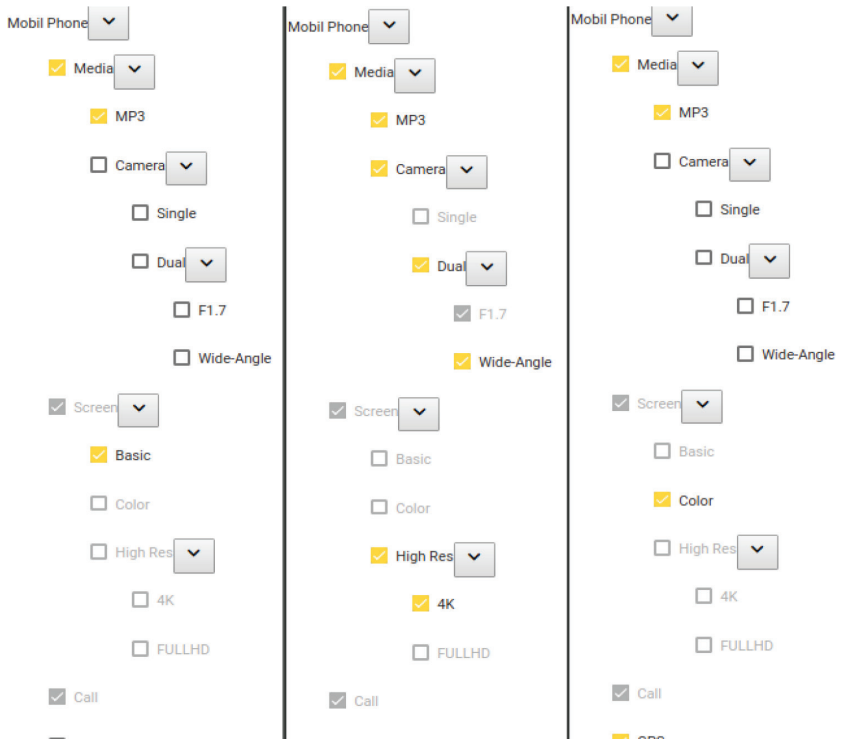


Figura 5 – Configuraciones válidas seleccionadas desde la aplicación web

Finalmente, Vari::Ation cuenta con una característica informativa, que indica algunas características básicas del modelo, como la cantidad de características obligatorias, opcionales o las restricciones del modelo, lo que permite que el usuario sea capaz de ver algunas características generales del modelo de forma sencilla y rápida.


```
[
    {
        "caracteristica": "Mobil Phone",
        "nivel": 0,
        "require": [],
        "excluye": [],
        "tipo": "root",
        "seleccionada": true
    },
    {
        "caracteristica": "Media",
        "nivel": 1,
        "require": [],
        "excluye": [],
        "tipo": "Optional",
        "seleccionada": false
    },
    {
        "caracteristica": "MP3",
        "nivel": 2,
        "require": [],
        "excluye": [],
        "tipo": "OR",
        "seleccionada": false
    },
    {
        "caracteristica": "Camera",
```


Figura 6 – Modelo de configuración en formato JSON

4.3. Resultados de pruebas y validación de Vari::Ation

Esta herramienta fue sometida a diferentes pruebas funcionales. El objetivo de las pruebas es validar que Vari::Ation permite importar modelos en formato XML, realizar la configuración siguiendo las reglas y restricciones propias del MC, y exportar adecuadamente los archivos JSON anteriormente descritos. Las pruebas fueron diseñadas para someter a la aplicación a un proceso de validación con tres MC distintos (incluido el mostrado en la sección 4.1), y la herramienta aprobó la mayoría de los casos propuestos, aunque dependiendo del modelo, no siempre se pudieron realizar todas las pruebas, debido a que los modelos no necesariamente tienen que presentar una estructura similar entre sí, por lo que no todos cumplen con cada caso diseñado, el caso que más se repitió es que no todos los modelos cuentan con restricciones del tipo *exclude*.

Fueron tres los segmentos de la aplicación que se sometieron a la evaluación, primero, importar el MC y la generación de un árbol, luego la selección de características y finalmente exportar el modelo de configuración en formato JSON. Para cada segmento se usaron plantillas como la descrita en la Tabla 1 y se realizaron las pruebas de manera manual. Estas plantillas fueron adaptadas a Gitlab para documentar el proceso y dejar una evidencia visual del correcto funcionamiento de la herramienta. En la Figura 6 se ve un ejemplo de una prueba realizada.

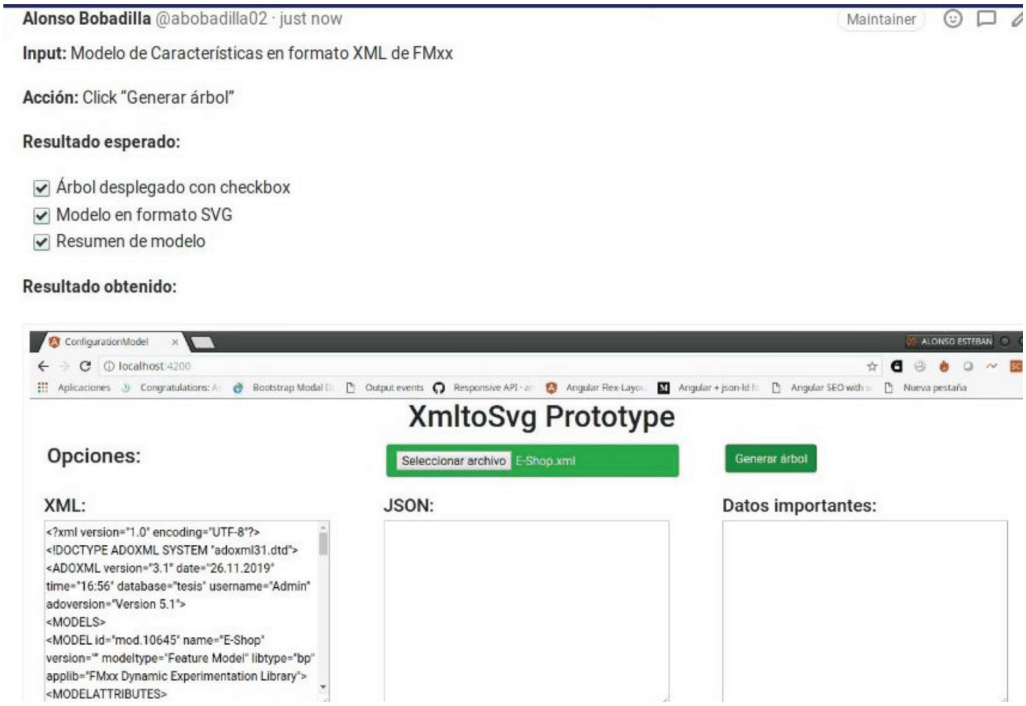


Figura 7 – Prueba realizada y documentada en Gitlab

Luego de realizar 31 pruebas para cada MC, se realizó un resumen con los resultados obtenidos. En la Tabla 2 se muestra el resumen para el MC de tiendas electrónicas, donde se da cuenta de la cantidad de pruebas por segmento abordado, si estas fueron exitosas, fallidas o no se realizaron porque no se cumplían los parámetros mínimos para ejecutarlas.

Segmento	Total pruebas	Pruebas realizadas		Pruebas no realizadas
		Exitosas	Fallidas	
Importar MC y generación de árbol	*5*	*5*	*0*	*0*
Selección de características	*21*	*16*	*0*	*5*
Exportar modelo de configuración y archivo de reglas	*5*	*3*	*0*	*2*

Tabla 2 – Resultados de pruebas en MC de tiendas electrónicas

4.4. Discusión

Vari::Ation ha sido creada inicialmente para dar soporte a los MC, específicamente a MC creados en la plataforma FMxx, por lo que en primera instancia podría parecer una

solución particular para este contexto. Sin embargo, esta primera versión de Vari::Ation corresponde a una prueba de concepto, realizada en la primera etapa de un proyecto de dos etapas, que tiene como objetivo final ser capaz de realizar el ciclo completo de una LPS, desde el modelado hasta la obtención de productos particulares por medio de la derivación de productos. En este sentido, es posible decir que Vari::Ation tiene el potencial para evolucionar desde la configuración de productos para una nomenclatura específica en una herramienta particular, a funcionar con distintos tipos de nomenclaturas provenientes desde distintos tipos de herramientas y formatos, lo que lo desligaría de FMxx y lo haría más potente, conservando la capacidad de configurar los productos, utilizando el mismo tipo de formato árbol o extendiendo su funcionalidad dependiendo de las particularidades de los modelos a importar.

Otro problema de Vari::Ation es que no cuenta con un *backend* ni con una base de datos, por lo que los modelos solo pueden ser importados por el usuario una vez ha ingresado a la aplicación web y sus configuraciones solo las puede guardar exportando los archivos en formato JSON, además hay procesos que no están optimizados por lo que el tiempo de carga al importar un MC podría ser mejorado. Esto es algo que se planea implementar en el futuro, en primer lugar, algunas funcionalidades que se ejecutan directamente en Angular serán migradas al servidor, para agilizar la aplicación en el cliente, también se pretende agregar un sistema de autentificación, para que las personas se registren al sistema y mediante sus usuarios accedan a la aplicación para almacenar sus MC y configurarlos como quieran, guardando las configuraciones realizadas en una base de datos. De esta manera es posible crear las configuraciones requeridas por un usuario, permitiéndole guardar sus progresos en la configuración de un producto, o directamente almacenar una configuración válida, para luego utilizarla, por ejemplo, para gestionar su trazabilidad. Además, como el proceso de configuración de productos, usualmente involucra a más de un *stakeholder* es factible añadir a sus funcionalidades la capacidad de trabajar de manera colaborativa, para que más de un usuario modifique una misma configuración o entregue retroalimentación sobre el proceso que se está llevando a cabo.

Aunque la parte funcional de una aplicación es importante, Vari::Ation tiene un déficit en su interfaz de usuario, por lo que es poco atractiva y puede incluso ser hasta poco intuitiva de utilizar, por lo que es necesario realizar un rediseño, que incluye la facultad de observar que tipo de característica se está seleccionando, ya que actualmente solo se ve el nombre de esta y también se planea modificar la forma en que el árbol en formato de *checkboxs* interactúa con el MC desplegado en formato SVG, ya que actualmente solo se muestra de forma referencial, lo que resulta incómodo especialmente para MC que cuentan con muchas características.

Por último, es necesario señalar que la importancia principal de Vari::Ation es permitir la conexión entre un MC con la siguiente etapa, ya que el modelo de configuración es el que dicta las reglas para que un producto particular sea creado mediante la derivación de productos. Y para la derivación de productos es necesario implementar los componentes de manera que puedan acoplarse según las reglas establecidas, por lo que una mala configuración podría llegar a ser fatal al momento de armar el producto.

5. Trabajos relacionados

A continuación, se presenta un conjunto de trabajos relacionados con el modelamiento o posterior configuración de un MC.

En (Mendonca, Branco & Cowan, 2009), se da a conocer una herramienta llamada S.P.L.O.T. que permite modelar MC mediante un editor en formato árbol y también permite generar y exportar modelos de configuración en formatos CSV y XML, pero no permite modelar MC con pocas características, aunque sean válidos.

En (Bagheri & Gasevic, 2011), modelan una LPS para grafos. Dicha LPS presenta una característica obligatoria formada por dos parejas de características del tipo alternativa. Esto implica que obligatoriamente se debe elegir un hijo, de cada par de hijos. No obstante, esta propuesta no incluye restricciones en el modelo, sino que las expresa sólo en lenguaje natural.

Por otro lado, en (Acher, Cleve et al., 2012), se da a conocer un MC que aparte de tener relaciones del tipo OR y XOR, tiene unas relaciones llamadas *Mutex*, y en lugar de restricciones del tipo *require* tiene restricciones *implies*. Las relaciones Mutex son similares a los XOR, solo que tienen cardinalidad, por lo que se puede indicar la cantidad de características que permite seleccionar, además las restricciones se escriben en lenguaje natural.

Finalmente, si bien FMxx no soporta algunas de las relaciones específicas entre características que muestran los trabajos relacionados, si se incorporan restricciones como parte del MC, lo que previene la generación de modelos con errores y que estos eventualmente se traspasen al modelo de configuración y posterior derivación de productos. Además, es posible visualizar las restricciones y tener una representación del MC mientras se va creando el modelo de configuración.

6. Conclusiones

El artículo presenta la primera etapa de una propuesta que continúa una línea de trabajo y que pretende ser capaz de configurar MC generados con la herramienta FMxx para luego exportar un archivo con reglas de configuración, que será el parámetro de entrada para comenzar el proceso de derivación de productos. Además, se modeló una adaptación de un MC mediante la herramienta FMxx para ser utilizada en las pruebas realizadas a la herramienta. Finalmente, la herramienta fue validada mediante pruebas funcionales que quedaron documentadas.

Actualmente se trabaja en la etapa dos de la propuesta, centrándose en la implementación de la variabilidad mediante el *framework Angular* y su biblioteca *Schematics, para una herramienta llamada Deriv::Ation*. También se trabaja en la conexión de las reglas obtenidas a través del modelo de configuración con los componentes para generar aplicaciones web de manera automática.

Como trabajo futuro se pretende extender la funcionalidad de importar MC de Vari::Ation, para no depender de FMxx y que de esta manera tenga la capacidad para configurar MC con diferentes nomenclaturas. También está planificado el desarrollo del *backend* para Vari::Ation, junto a la implementación de una base de datos, para que las configuraciones puedan ser almacenadas y se pueda gestionar la trazabilidad.

Por último, se esperan agregar distintas funcionalidades a Vari::Ation que añadan información a los modelos de configuración mostrados, en específico, se tratan de datos complementarios como la posibilidad de darle un nombre al proyecto, diferenciar entre generar proyectos con *backend*, *frontend* o ambos, entre otras funcionalidades adicionales.

Agradecimientos

Samuel Sepúlveda agradece al proyecto DI20-0060, Vicerrectoría de Investigación y postgrado, Universidad de la Frontera.

Referencias

Acher, M., Cleve, A., Perrouin, G., Heymans, P., Vanbeneden, C., Collet, P., & Lahire, P. (2012). On extracting feature models from product descriptions. In Proceedings of the Sixth International Workshop on Variability Modeling of Software-Intensive Systems, (pp. 45-54). ACM. https://doi.org/10.1145/2110147.2110153.

Ahmed, F., & Capretz, L. F. (2008, May). Best practices of RUP® in software product line development. In *2008 International Conference on Computer and Communication Engineering* (pp. 1363-1366). IEEE. https://doi.org/10.1109/ICCCE.2008.4580828.

Bagheri, E., & Gasevic, D. (2011). Assessing the maintainability of software product line feature models using structural metrics. *Software Quality Journal*, *19*(3), 579-612. https://doi.org/10.1007/s11219-010-9127-2.

Clements, P., & Northrop, L. (2001). *Software product lines. Practices and patterns (3rd Ed.)*. Boston, MA: Addison-Wesley.

Czarnecki, K., & Wasowski, A. (2007). Feature diagrams and logics: There and back again. In *11th International Software Product Line Conference (SPLC 2007)* (pp. 23-34). IEEE. https://doi.org/10.1109/SPLINE.2007.24.

Deelstra, S., Sinnema, M., & Bosch, J. (2004). Experiences in software product families: Problems and issues during product derivation. In *International Conference on Software Product Lines* (pp. 165-182). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-28630-1_10.

Di Lucca, G. A., Fasolino, A. R., Faralli, F., & De Carlini, U. (2002). Testing web applications. In *International Conference on Software Maintenance, 2002. Proceedings*. (pp. 310-319). IEEE. https://doi.org/10.1109/ICSM.2002.1167787

Esperguel, M., & Sepúlveda, S. (2016). Feature modeling tool: a proposal using ADOxx technology. In *2016 XLII Latin American Computing Conference (CLEI)* (pp. 1-9). IEEE. https://doi.org/10.1109/CLEI.2016.7833363.

Esperguel, M., Sepúlveda, S., & Monsalve, E. (2017, October). FMxx: A proposal for the creation, management and review of feature models in software product lines. In 2017 36th International Conference of the Chilean Computer Science Society (SCCC) (pp. 1-7). IEEE. https://doi.org/10.1109/SCCC.2017.8405152.

Galster, M., Weyns, D., Tofan, D., Michalik, B., & Avgeriou, P. (2013). Variability in software systems—a systematic literature review. *IEEE Transactions on Software Engineering*, *40*(3), 282-306. https://doi.org/10.1109/TSE.2013.56.

Heradio, R., Perez-Morago, H., Fernandez-Amoros, D., Cabrerizo, F. J., & Herrera-Viedma, E. (2016). A bibliometric analysis of 20 years of research on software product lines. Information and Software Technology, 72, 1-15. https://doi.org/10.1016/j.infsof.2015.11.004.

Larman, C. (2004). Agile and iterative development: a manager's guide. Addison-Wesley Professional. https://books.google.cl/books?id=76rnV5Exs50C.

Mazo, R. (Ed.). (2018). Guía para la adopción industrial de líneas de productos de software. Universidad Eafit. https://books.google.cl/books?id=EjCjDwAAQBAJ.

McGregor, J. D., Muthig, D., Yoshimura, K., & Jensen, P. (2010). Guest editors' introduction: Successful software product line practices. *IEEE software*, *27*(3), 16-21. https://doi.org/10.1109/MS.2010.74.

Melia, S., Pardillo, J., & Cachero, C. (2010). Automatic selection of RIA software architectures using quality models. In 2010 Seventh International Conference on the Quality of Information and Communications Technology (pp. 505-510). IEEE. https://doi.org/10.1109/QUATIC.2010.87.

Mendonca, M., Branco, M., & Cowan, D. (2009). SPLOT: software product lines online tools. In Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications (pp. 761-762). https://doi.org/10.1145/1639950.1640002.

Sepúlveda, S., Cares, C., & Cachero, C. (2013). An overview of syntactic and semantic issues in software product lines. In 2013 32nd International Conference of the Chilean Computer Science Society (SCCC) (pp. 11-15). IEEE. https://doi.org/10.1109/SCCC.2013.15.

Sepúlveda, S., Cravero, A., & Cachero, C. (2016). Requirements modeling languages for software product lines: A systematic literature review. Information and Software Technology, 69, 16-36. https://doi.org/10.1016/j.infsof.2015.08.007.

She, S., Lotufo, R., Berger, T., Wąsowski, A., & Czarnecki, K. (2011). Reverse engineering feature models. In Proceedings of the 33rd International Conference on Software Engineering (pp. 461-470). https://doi.org/10.1145/1985793.1985856

Sommerville, I. (2015). Software Engineering. 10th. In Book Software Engineering. 10th, Series Software Engineering. Addison-Wesley.

Vale, T., de Almeida, E. S., Alves, V., Kulesza, U., Niu, N., & de Lima, R. (2017). Software product lines traceability: A systematic mapping study. Information and Software Technology, 84, 1-18. https://doi.org/10.1016/j.infsof.2016.12.004

Weiss, D. M., Li, J. J., Slye, H., Dinh-Trong, T., & Sun, H. (2008). Decision-model-based code generation for SPLE. In 2008 12th International Software Product Line Conference (pp. 129-138). IEEE. https://doi.org/10.1109/SPLC.2008.42.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 02/08/2020
Aceitação/Acceptance: 31/10/2020

DataOps, una alternativa para la gestión de Data Pipelines

Laura Vargas-Rueda[1], Sergio Pongutá-Díaz[1], Juan-Sebastián González-Sanabria[1]

laura.vargas08@uptc.edu.co; sergio.ponguta@uptc.edu.co; juansebastian.gonzalez@uptc.edu.co

[1] Universidad Pedagógica y Tecnológica de Colombia, Av. Central del Norte, 150001, Tunja, Colombia.

Pages: 259-269

Resumen: DataOps es un nuevo enfoque donde la velocidad y la agilidad son la prioridad en el proceso de DataPipelines, desde la recopilación de datos hasta la entrega del análisis. DataOps tiene una diversidad de definiciones debido a que aún se encuentra en una etapa temprana de adopción, sin embargo, los investigadores en el tema coinciden en que DataOps permite mejorar la comunicación entre las partes interesadas y todos los elementos que componen el proceso de flujo de datos para obtener mejores resultados en la entrega final. En este trabajo se presenta un análisis del entorno y la concepción de dicho enfoque según la literatura encontrada sobre el tema, llegando a consolidar una explicación clara sobre el funcionamiento del enfoque a nivel organizacional dentro de los DataPipelines empresariales, así mismo se convierte en un documento base para que investigadores y académicos tengan una aproximación en un lenguaje más claro de las características y consideraciones que se deben tener en cuenta para hacer uso de DataOps.

Palabras-clave: DataOps; Pipeline; Automatización; Ciencia de datos; Análisis de datos.

DataOps, an Data Pipelines Management Alternative

Abstract: DataOps is a new approach where speed and agility are the priority on the Data Pipelines process, from data collection to the delivery of analytics. DataOps is a term with a variety of related concepts, since it is still on an early stage of adoption, however, research groups agree that DataOps improves the communication between stakeholders and all the elements that compose the data flow process, in order to get better results at delivery. In this work, an analysis of the environment and the conception of said approach is presented according to the literature found on the subject, coming to consolidate a clear explanation about the operation of the approach at the organizational level within the Business Data Pipelines, and it also becomes a document basis for researchers and academics to have a clearer approach to the characteristics and considerations that must be taken into account to make use of DataOps.

Keywords: DataOps; Pipeline; Automation; Data science; Data analysis.

1. Introducción

Debido al constante crecimiento en los volúmenes de datos que actualmente almacenan las empresas, en los años recientes se ha propendido por el uso de sistemas de análisis de datos que incorporen funcionalidades ETL (Extracción, Transformación y Carga), principalmente en aquellas empresas que hagan uso de Datawarehouse o de BigData (Bindu & Bonthu, 2017). Estos sistemas de análisis proveen a las empresas una solución integral, que va desde la configuración asistida, hasta la generación de robustos métodos de acceso; sin embargo, tienen diversas dificultades: requerir de esquemas de almacenamiento dedicados, necesidad de construir mapeos, realización de documentar procesos tediosos y comunicarlos posteriormente a los usuarios, entre otros (Pinkel, 2015).

Debido a las dificultades anteriormente mencionadas, han surgido una serie de alternativas como los DataOps, que se consolida como un nuevo enfoque para el análisis de datos, este enfoque tiene notable superioridad respecto a los métodos de analítica tradicionales en términos de flexibilidad, calidad y tiempo de ciclo de desarrollo. DataOps adopta los conceptos clave de Lean Manufacturing (Padilla, 2010), al considerar el análisis de datos como una Pipeline de operación continua, que puede automatizarse, monitorearse y controlarse (Webster & Watson, 2002), por tanto, propone realizar una implementación estilo contenedores que proporcione retroalimentación sobre el rendimiento y la calidad del servicio del proceso de gestión de datos, de esta manera el Pipeline se va "auto-ajustando" en tiempo real según las necesidades y requisitos operativos actuales (Bergh, Benghiat & Strod, 2019).

Si bien existen variadas concepciones, DataOps se podría definir como un conjunto de prácticas, procesos, métodos de ingeniería de software ágil y tecnologías que combina una perspectiva integrada y orientada a los procesos sobre datos con automatización (Ereth, 2018), el cual se enfoca en la gestión de datos, comunicación, colaboración, integración, automatización y la medición continua de la cooperación entre los ingenieros de datos con los científicos de datos, sin descuidar a los demás miembros de los equipos (Palmer, 2015); dicha integración se basa en crear y mantener una relación bidireccional entre los objetivos de la gestión de datos con el consumo de datos (Whitmore, 2017). Sin embargo, es de aclarar que a hoy no hay un conjunto de herramientas de DataOps, ni una definición del alcance de lo que una plataforma debe cubrir, solo se limita a la proposición de marcos que respaldan un enfoque de colaboración y agilidad, cubriendo de extremo a extremo la operación de los datos (Whitmore, 2017).

Pese al auge que ha tenido DataOps, aún no es ampliamente descrito ni mencionado en artículos científicos y, en aquellos que sí lo involucran, lo hacen superficialmente y no explican su funcionamiento. Por lo anterior, en el presente trabajo se explica y desarrolla el uso de DataOps, recurriendo a la revisión de la literatura encontrada en índices y bases de datos de artículos. El tema es desglosado a un nivel de pleno entendimiento para la comunidad académica e investigativa, con el fin de fomentar su uso y evidenciar las ventajas de este enfoque en los procesos de análisis de datos.

2. Metodología

Con el fin de determinar el uso de DataOps a nivel de investigaciones preliminares, se identificó y revisó la literatura en el tema siguiendo el procedimiento de revisión de conocimiento propuesto por Webster y Watson (2002):

1. Primera búsqueda: Una primera búsqueda se hizo con la palabra clave "DataOps" en diferentes bases de datos de artículos científicos y documentos de conferencias, principalmente Springer y Redalyc. Además de libros y documentos de empresas particulares que han realizado investigaciones sobre la metodología.
2. Selección: Al ser un tema muy reciente no existe una gran cantidad de documentos que traten acerca del mismo, así que se seleccionaron todos los artículos encontrados basados en dos criterios: (a) Su título contiene la palabra clave "DataOps", (b) El texto contiene la palabra clave "DataOps" en alguna de sus secciones diferente a la introducción.
3. Revisión: Una vez seleccionados los estudios se procedió a revisar y buscar las referencias citadas en estos que pudiesen aportar información al artículo.

En la Tabla 1 se presenta la cantidad de documentos recuperados en la etapa de selección de los índices y buscadores Google Scholar y Scopus, teniendo en cuenta que estos son los de mayor cobertura (Martín-Martín et al., 2018).

Buscador	Documentos Español	Documentos Ingles
Google Scholar	4	22
Scopus	0	4

Tabla 1 – Número de trabajos encontrados en índices y buscadores académicos

Teniendo en cuenta los resultados tan limitados obtenidos, y al analizar la información que se presentaba en los mismos, se encontró la particularidad que el término no era plenamente de desarrollado en la mayoría de los trabajos, motivo por el cual, se reforzo el objetivo del trabajo de plasmar un documento que recopile el funcionamiento básico del enfoque. Así mismo, pese a que se intentó incluir estudios de caso, a nivel documental la información es limitada, sin embargo, las experiencias de los casos encontrados fueron incorporados en el trabajo para poder plasmar una base documental para los académicos y profesionales que se recién se involucran en el tema.

3. DataOps

En esta sección se presentan los principios encontrados en el manifiesto de DataOps (DataOps Manifesto, 2019), los cuales son explicados haciendo uso de la experiencia de diferentes autores al momento de aplicar un proceso de DataOps dentro de una organización. Cabe resaltar que tras utilizar durante años sistemas de análisis de datos, se llegó a la conclusión de que los factores que se suelen valorar dentro de una organización tienen prioridad de unos sobre otros (Bergh, Benghiat & Strod, 2019). Un sistema de analítica bien aplicado deberá dar principal atención a: la interacción de los individuos,

las analíticas aplicadas sobre el trabajo realizado, las colaboraciones con el cliente, la experimentación, iteración, retroalimentación, y, las propiedades interfuncionales de las operaciones que se realizan; todo esto sobre otros elementos que son menos dinámicos o están aislados como, las responsabilidades particulares de un equipo, o las herramientas y amplios diseños iniciales que no cambian a lo largo de todo el proceso de generación y análisis de datos (Bergh, Benghiat & Strod, 2019), en la Figura 1 se sintetizan los 18 principios de DataOps (DataOps Manifesto, 2019).



Figura 1 – Principios de DataOps. Fuente: elaboración propia, basado en (DataOps Manifesto, 2019).

DataOps se ajusta al flujo continuo de trabajo sobre agilidad en inteligencia de negocios y ciencia de datos. Aquí, la agilidad se ve como "la capacidad de reaccionar a requisitos imprevistos o volátiles con respecto a la funcionalidad o el contenido de una solución de BI" (Ereth, 2018).

3.1. Componentes y proceso de DataOps

DataOps es una metodología basada en DevOps, orientada a procesos que dan soporte para la toma de decisiones comerciales y monetización de los datos, eliminando los silos de datos de tal manera que se obtengan datos de calidad y permitiendo la automatización de los Data Pipelines (Ranjan & Premchand, 2019).

Dado que la intención de DataOps es automatizar tareas manuales existentes y racionalizar el proceso de creación de DataPipelines, en la Tabla 2 se sintetizan los componentes a incorporar dentro de la implementación de DataOps

Componentes	Descripción
Orquestación de los Data Pipelines	Incluye los pasos en la producción de análisis de datos, como el acceso a los datos, la integración, el modelo y la visualización, que deberán ser un flujo de trabajo basado en gráficos dirigidos.
Automatización de garantía	Se centra en las pruebas automáticas y el monitoreo de la calidad de producción de datos y artefactos en el proceso de producción de análisis de datos.
Despliegue continuo y creación	La implementación continua se centra en el movimiento del código, así como en la configuración continúa en la producción desde los entornos de desarrollo de manera transparente.

Componentes	Descripción
Implementación del modelo de ciencia de datos	Los entornos de desarrollo reproducibles son casi espejos de los entornos de producción para garantizar el movimiento libre de errores de los modelos de ciencia de datos en producción.

Tabla 2 – Componentes y proceso de DataOps

El proceso DataOps tiene una serie de pasos, descritos en la Figura 2, que deben ser adecuadamente implementados para obtener los beneficios de la adopción del enfoque. A continuación, se describe cada uno de los pasos (Ranjan & Premchand, 2019):

- Planificación de requisitos comerciales: Incluye la identificación de consumidores, productores o fuentes, proceso de adquisición, modelo y el proceso general de los datos. Se divide en: i) identificación de consumidores comerciales, y, ii) la definición del alcance y los objetivos.
- Adquisiciones de datos: Inicia desde la identificación del objetivo, fuentes de datos e identificación de las herramientas para la creación de un Pipeline.
- Transformación de datos: Se basa en el proceso de modelo, prueba y evaluación.
- Estrategias de transformación: A medida que los datos pasan por el Data Pipeline, deben pasar por los filtros y pruebas correctos para garantizar que los datos sean confiables y su calidad sea incuestionable. Aquí es donde la estandarización introducida por DataOps juega un papel importante pues la

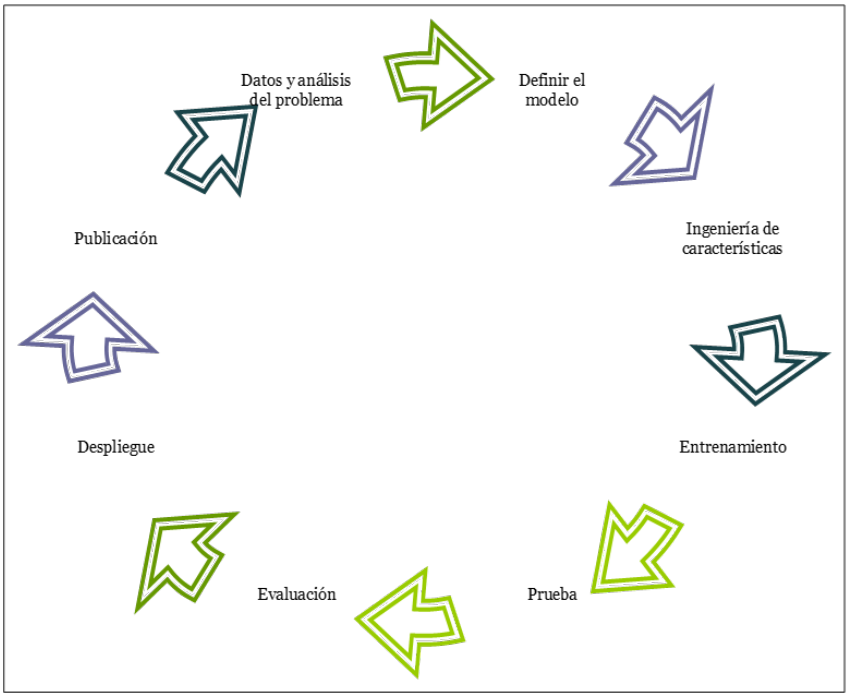


Figura 2 – Proceso de modelo, prueba y evaluación. Fuente: elaboración propia, con base en (Ranjan & Premchand, 2019).

trazabilidad y el flujo continuo son clave para garantizar la entrega de valor empresarial utilizando DataOps.
- Gestión del repositorio de datos: La organización de los datos juega un papel importante, por eso la relevancia de realizar una adecuada selección del repositorio de datos.
- Descubrimiento: Antes de resolver cualquier problema utilizando un nuevo modelo, los modelos existentes y los datos relacionados se pueden redescubrir si hay problemas y soluciones similares disponibles.
- Modelado de datos: Se requiere un enfoque científico para la adopción de DataOps que pueda ayudar a mantener el modelo, fomentar la reutilización, ayudar al crecimiento y permitir la creación de información para ayudar a las empresas.
- Publicación de información: Los modelos publicados se pueden reutilizar a nivel empresarial para obtener diversos análisis de utilidad para las empresas.

3.2. DataOps y Pipelines

Una Data Pipeline es el código fuente responsable de convertir el contenido sin formato en información útil. Esta Pipeline es esencial para el análisis de datos y se puede automatizar de extremo a extremo para producir un código que se pueda consumir en forma reproducible (IBM, 2018). La intención de DataOps es automatizar varias tareas manuales existentes y racionalizar el proceso de creación de Pipelines de datos (Tabla 3).

Componente de software	Componente técnico
Orquestación de los Data Pipelines	Automatización de pruebas
Pruebas y calidad de producción de prueba, e integración y unificación de datos	Repositorio de código
Automatización de implementación	Marcos de orquestación
Implementación del modelo de ciencia de datos y gestión de sandbox	Colaboración y gestión del flujo de trabajo
Virtualización de datos	Gestión de metadatos
Control de versiones	Linaje y análisis de impacto
Gestión de datos	Sistemas gestores de bases de datos; herramientas de integración, preparación y automatización de datos; herramientas de análisis y visualización; y, plataformas de monitoreo e inteligencia de desempeño

Tabla 3 – Componentes clave de una plataforma DataOps

Cuando se deseen considerar herramientas para apoyar una práctica de DataOps dentro de una organización, se debe mirar desde un enfoque holístico que incorporé: Servicios de curación de datos, Gestión de metadatos, Gobierno de datos, Gestión de datos maestros e Interacción de autoservicio (IBM, 2018).

La adaptación del modelo DevOps para DataOps da como resultado un modelo de proceso complejo que consta de dos bucles de prácticas CI/CD interactivos, uno para desarrollar y operacionalizar modelos analíticos, y otro para desarrollar y operacionalizar Data Pipelines (Wells, 2019). Dichas prácticas se describen de la siguiente manera:

- Integración continua (CI): Es una práctica de desarrollo de gran aceptación en la industria del desarrollo de software, esta práctica tiene como objetivo hacer que los miembros de un equipo integren y fusionen el trabajo de desarrollo (la mayor parte de las veces, código) frecuentemente a lo largo del día. CI permite a las compañías de software tener un ciclo más corto para hacer los lanzamientos de avances en el desarrollo, mejorar la calidad del software y aumentar la productividad de sus equipos. Esta práctica incluye la creación y prueba automatizada de software (Thusoo & Sen, 2017).
- Despliegue continuo (CD): Es una práctica que permite el despliegue de la aplicación de forma automática y continua en entornos de producción. No debe haber pasos manuales en CD, en esta práctica tan pronto como los desarrolladores confirmen un cambio, el cambio se implementa en producción a través de una Pipeline de implementación (Thusoo & Sen, 2017).

En la Figura 3, el ciclo superior describe CI/CD para análisis y ciencia de datos, y el ciclo inferior describe CI/CD para Data Pipelines. El punto es que el análisis impulsa la demanda de canalizaciones de datos, no a la inversa. DataOps siempre comienza con análisis.

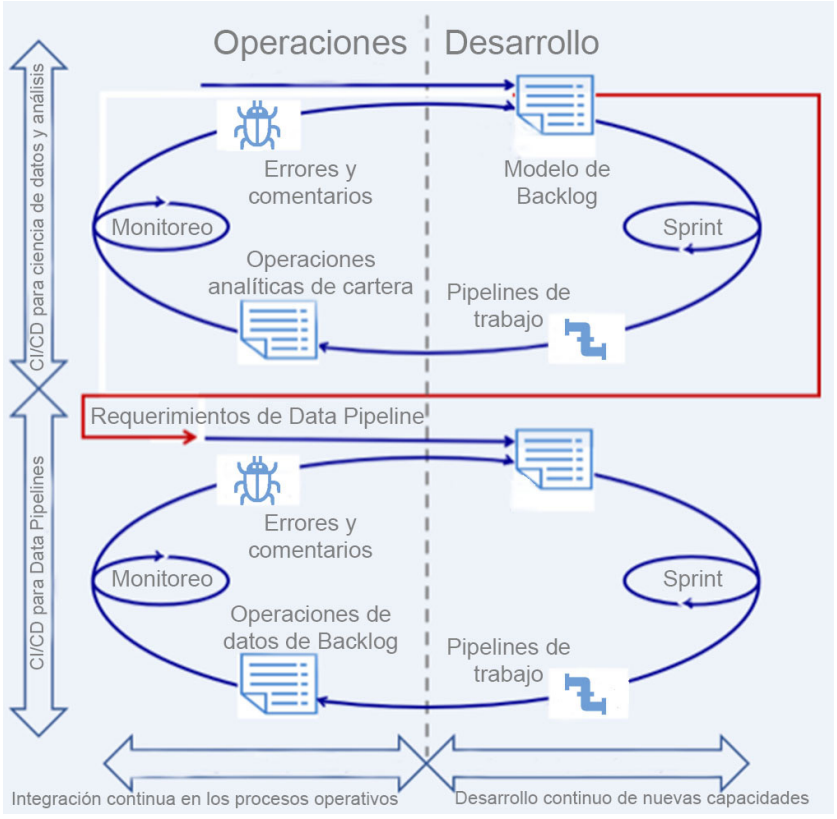


Figura 3 – CI/CD en los Pipelines de DataOps. Adaptado de (Wells, 2019).

El desarrollo de nuevos modelos no es posible hasta que los datos estén disponibles, por lo que nuevamente las canalizaciones de datos son críticas. La línea roja en de la Figura 3 representa una dependencia entre los dos bucles. Cuando se entiende el trabajo atrasado del modelo, se utiliza para identificar las canalizaciones de datos que se requieren para desarrollar y poner en funcionamiento los modelos atrasados. La cartera de pedidos del modelo se convierte en una nueva fuente de requisitos que alimenta la cartera de pedidos (Wells, 2019).

3.3. DataOps, una necesidad

Las empresas funcionan a un ritmo cada vez más rápido, por lo que, si los datos no se mueven a la misma velocidad, se eliminan del proceso de toma de decisiones. Esto es similar a cómo la agilidad en la creación de aplicaciones web condujo a la creación de la cultura DevOps, ahora se necesita la misma agilidad en el lado de los datos, dado que los mismos se vuelven más convencionales, existiendo una proliferación de fuentes de datos debido a los avances en la recopilación, nuevas aplicaciones, sensores en Internet de las cosas (IoT, por las siglas en inglés de Internet of Things) y redes sociales.

DataOps se ha convertido en una disciplina crítica para cualquier organización de Tecnologías de la Información (TI) que quiera sobrevivir y prosperar en un mundo en el que la inteligencia empresarial en tiempo real es una necesidad competitiva. Lo anterior es impulsado principalmente por: los datos no son una "cosa" estática, la tecnología no es suficiente y se necesita mayor agilidad en los procesos (Shahin, Ali Babar & Zhu, 2018).

Hay riesgo de falle el enfoque de DataOps al momento de llevarlo a cabo, por esta razón la organización se deben plantear las siguientes preguntas antes de realizar la implementación. ¿Durante la migración, los datos son cifrados? Si no se cifran los datos, como resultado, estos son vulnerables durante el tránsito y en el otro extremo (en reposo). ¿Se hace uso de datos fabricados? Entrenar los sistemas sin datos reales limita las pruebas y evita generar errores importantes. ¿Se hace uso de datos obsoletos? Los datos antiguos/sucios evitarán que sus científicos de datos vean las señales, evitarán el "aprendizaje de IA" y restringirán la capacidad de sus proyectos para realizar pruebas de manera efectiva (ItUser, 2019). ¿Es confiable la seguridad de la información? En ocasiones pensamos que es suficiente establecer restricciones durante la producción, pero al realizar copias de los datos, estos quedan desprotegidas y vulnerables al robo.

A la hora de manejar recursos humanos para llevar a cabo la implementación de DataOps es necesario tener en cuenta:

- Contratar visionarios de datos.
- Mantener los datos en un almacén de datos de tal manera que todos tengan acceso a este.
- Empoderar a todos los empleados.
- Invertir en herramientas de datos de autoservicio adecuadas.
- Dar responsabilidades a los empleados.

Aunque es un reto mantener los datos y el personal en una misma sintonía (Atwal, 2020), el enfoque de DataOps trae grandes beneficios para las organizaciones, tales como: acelerar el tiempo de producción mediante Pipelines analíticas simplificadas y automatizadas, aumentar la calidad, la fiabilidad y la visibilidad (pues permiten mejorar

la velocidad y la robustez), y aumentar la seguridad al hacer uso de una plataforma de datos unificada (Morris & Opazo, 2018).

Adicionalmente, como lo mencionan en experiencias previas de aplicación en la industria (Atwal, 2020; Ranjan & Premchand, 2019), DataOps será "la columna vertebral de los sistemas de análisis de datos en los próximos años", dado que tiene un alto potencial cuando se implementa correctamente, aportando un elevado valor comercial a las organizaciones que lo adoptan. El ciclo de DataOps es repetible e iterativo lo que permite a la organización un alcance para la evaluación continua y la mejora de los modelos.

4. Conclusiones

Un ecosistema de datos basado en DataOps debe ser escalable, de alto rendimiento y de plena disponibilidad, de tal manera que aporte valor comercial a las organizaciones a la toma de decisiones. Es posible realizar una aproximación basada en metodologías ágiles, DevOps y DataOps en aquellos proyectos de ciencia de datos que hagan uso de herramientas para la automatización de la creación de datos consumibles por diferentes sectores del negocio.

DataOps reconoce un nuevo concepto en el área de desarrollo, donde los proyectos de software requieren que los datos integrados al proceso deben moverse a la misma velocidad que los demás elementos de este. Es de resaltar que esta es una metodología que aparece como una serie de principios para los científicos de datos y desarrolladores que centran su trabajo en los datos, esta les permite manejar un flujo de trabajo ágil sin abandonar los requisitos de gobierno de datos.

DataOps puede ayudar a manejar entornos que presenten datos complejos y generar soluciones analíticas que requieren la coordinación de una gran cantidad de tecnologías, metodologías y partes interesadas en el negocio. Así mismo, permite acelerar los procesos y aumentar la calidad al proporcionar pipelines de análisis de datos optimizadas a través de niveles profundos de automatización y pruebas.

Las organizaciones tienen la necesidad creciente del contar con "Cientificos de Datos", sin embargo, dada la dificultad que existe para contar con los mismos en la actualidad, el combinar los perfiles adecuadamente permitirá mejorar la eficiencia de estas. Tal es el caso de DevOps, donde el contar con perfiles en analistas de datos con conocimientos en gestión de operaciones de software permitirá una mejor integración junto con desarrolladores de software y expertos en machine learning para un proceso eficiente.

DataOps puede ayudar a manejar entornos que presenten datos complejos y generar soluciones analíticas que requieren la coordinación de una gran cantidad de tecnologías, metodologías y partes interesadas en el negocio. Asimismo, permite acelerar los procesos y aumentar la calidad al proporcionar Pipelines de análisis de datos optimizadas a través de niveles profundos de automatización y pruebas.

Referencias

Atwal, H. (2020). Organizing for DataOps. En: *Practical DataOps*. Apress, Estados Unidos.

Bergh, C., Benghiat, G., & Strod, E. (2019). *The DataOps Cookbook*. DataKitchen Headquarters.

Bindu, K. H., & Bonthu S. (2017). Review of Leading Data Analytics Tools. *International Journal of EngineeringTechnology*, 7 (3), e18190. https://doi.org/10.14419/ijet.v7i3.31.18190

DataOps Manifesto (2019). *The DataOps Manifesto*. Disponible en http://dataopsmanifesto.org

Ereth, J. (2018). DataOps - Towards a Definition. En: *Proceedings of the Conference "Lernen, Wissen, Daten, Analysen",* Mannheim, Germany, 104-112. http://ceur-ws.org/Vol-2191/paper13.pdf

IBM. (2018). An introduction to the IBM DataOps methodology and practice. Deliver business-ready data fast with DataOps. Disponible en https://www.ibm.com/downloads/cas/GLNAZBXN

ItUser (2019). Algunos patrones y antipa-trones clave de DataOps. Disponible en https://discoverthenew.ituser.es/devops/2019/10/algunos-patrones-y-antipatrones-clave-de-dataops

Martín-Martín, A. et al. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12 (4), 1160-1177. https://doi.org/10.1016/j.joi.2018.09.002

Morris & Opazo. (2018). Introducción a DataOps. Disponible en https://docplayer.es/152948025-Ebook-introduccion-a-dataops.html

Padilla, L. (2010). Lean manufacturing manufactura esbelta/agil. *Revista Ingeniería Primero*, 15, 64-69.

Palmer, A. (2015). *From DevOps to DataOps*. Disponible en https://www.tamr.com/blog/from-devops-to-dataops-by-andy-palmer/

Pinkel, C., Schwarte, A., Trame, J., Nikolov, A., Bastinos, A. S., & Zeuch, T. (2015). DataOps: Seamlees End-to-End Anything-to-RDF Data Integration. The Semantic Web: ESWC 2015. Editors: Gandon, F., Guéret, C., Villata, S., Breslin, J., Faron-Zucker, C., Zimmermann, A. (Eds.). Springer, 123-127. https://doi.org/10.1007/978-3-319-25639-9_24

Ranjan, A., & Premchand, P. (2019). DataOps in Manufacturing and Utilities Industries. *International Journal of Applied Information Systems*, 12 (23), 1-6. https://doi.org/10.5120/ijais2019451814

Shahin, M., Ali Babar, M., & Zhu, L. (2018). Continuous Integration, Delivery and Deployment: A Systematic Review on Approaches, Tools, Challenges and Practices. *IEEE Access*, 5, 3909-3943. https://doi.org/10.1109/access.2017.2685629

Thusoo, J., & Sen, A. (2017). *Creatinga Data-Driven Enterprise with DataOps*. O'Reilly Media, Inc., Sebastopol, California.

Webster, J., & Watson, R. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS* Quarterly, 26 (2), e4132319. https://doi.org/10.2307/4132319

Wells, D. (2019). DataOps: More Than DevOps for DataPipelines. Disponible en https://www.eckerson.com/articles/dataops-more-than-devops-for-data-pipelines

Whitmore, T. (2017). *DataOps: The Collaborative Framework for Enterprise Data-Flow Orchestration*. Report Blue Hill Research. Disponible en https://bit.ly/2LuKtaV

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 07/09/2020
Aceitação/Acceptance: 05/11/2020

Plataforma Software Extensible para Campus Inteligente Basada en Microservicios

Henry Jiménez, Emilio Cárcamo, Gabriel Pedraza

henry2198161@correo.uis.edu.co; emcartro@saber.uis.edu.co; gpedraza@uis.edu.co

Escuela de Ingeniería de Sistemas e Informática - Universidad Industrial de Santander, Bucaramanga, 680002, Colombia.

Pages: 270-282

Resumen: En el contexto de campus inteligente, gran parte de las investigaciones se han centrado en resolver un problema particular, como monitoreo ambiental o estacionamiento inteligente, pero no presentan una solución que permita la integración, interconexión y optimización de la tecnología IoT desarrollada para cada contexto. Este artículo propone una plataforma software parcialmente autónoma para campus inteligente, que brinda las bases y soporte tecnológico en todos los niveles de la infraestructura IoT. Se presenta la arquitectura de cada componente de la plataforma, sus propiedades de autonomía, escalabilidad y la validación realizada mediante la implementación de un prototipo y un caso de estudio simple.

Palabras-clave: Internet de las Cosas, Campus Inteligente, Microservicios, Computación Autónoma, Escalabilidad.

Extensible Software Platform for Smart Campus Based on Microservices

Abstract: In the smart campus context, large of the research proposals have focused on solving a particular problem , such as environmental monitoring or intelligent parking, but doesn't present a solution that allows the integration, interconnection and optimization of IoT technology developed for each context. This article proposes a partially autonomous software platform for Smart campus, providing the foundations and technological support in all IoT infrastructure levels. This paper introduces the architecture of each platform component, its autonomy properties, scalability, and validation performed by deploying a prototype and a simple case study.

Keywords: Internet of Things, Smart Campus, Microservices, Autonomic Computing, Scalability.

1. Introducción

En el contexto de la era digital y gracias a la proliferación de dispositivos especializados de bajo costo, la estandarización de protocolos de comunicación y la llegada de nuevas

tecnologías como el 5G, cada vez es más común el uso del IoT en ambientes cotidianos. Las tendencias muestran que el IoT continuará brindando nuevas oportunidades para la innovación durante la próxima década. De acuerdo con (Omale, 2018) pasaremos de 14.2 millardos de dispositivos conectados en 2019 a 25 millardos en el 2021.

En el área de Campus Inteligente, el desarrollo IoT ha permitido gestionar contextos específicos de aplicación como la medición del ambiente, seguridad de edificios y áreas del campus, ocupación de salones, gestión de parqueaderos, laboratorios, entre otros. Sin embargo, se encontró que entre las principales dificultades para la implementación del campus inteligente están la falta de una infraestructura software adecuada para tratar aspectos como heterogeneidad de dispositivos, lenguajes, sistemas Operativos, entre otros; costo asociado al uso de plataformas propietarias como Azure, y AWS; alta disponibilidad, tolerancia a fallos, escalabilidad a nivel de dispositivos, aplicaciones y usuarios; y la facilidad de administración.

Por esta razón, desde la ingeniería del software existe el desafío de diseñar e implementar en el contexto del campus inteligente una plataforma escalable para responder adecuadamente a la demanda de los diferentes servicios, extensible para la implementación o evolución de los componentes en cada contexto de aplicación, que sea tolerante a fallos, administrable y con alta usabilidad para todos los actores del campus.

En el presente trabajo se muestra cómo desde un enfoque de microservicios y de herramientas open source se puede lograr la interoperabilidad de plataformas, abordar los anteriores desafíos y dotar de capacidades autónomas al campus inteligente de próxima generación.

El artículo está organizado de la siguiente forma: la sección 2 presenta el trabajo relacionado en el contexto de Smart Campus y Arquitecturas escalables basadas en microservicios, la sección 3 describe el diseño y arquitectura propuesto para la plataforma de campus inteligente, la sección 4 muestra la validación de la plataforma y en la sección 5 las conclusiones y trabajo futuro.

2. Trabajo Relacionado

La noción de Smart Campus es fundamental para numerosas investigaciones en el contexto de IoT y como se presenta en los trabajos de (Pagliaro et al, 2016) y (Adamkó, Kãdek & Kósa, 2014) es equiparable en muchos aspectos a la noción de Smart City, de hecho, es posible hablar de un Campus inteligente como un modelo a una escala más pequeña de una ciudad inteligente; esta noción no está estandarizada o unificada y cada trabajo desarrollado en este contexto presenta una visión propia de lo que puede ser una arquitectura IoT, que satisface determinadas necesidades en un campus universitario.

Uno de los problemas que se pueden abordar es la recolección de información del entorno en tiempo real y su correcta presentación para apoyar la toma de decisiones, como se aprecia en el trabajo de (Monti, Prandi & Mirri, 2018) donde se propone una arquitectura que permite monitorear parámetros de un campus universitario como calidad del aire, temperatura ambiental, imagen de un lugar, y demás datos que posteriormente son tratados y enlazados con información complementaria para ser publicados; este trabajo

se complementa con (Prandi, Monti, Ceccarini & Salomoni, 2019) donde se realiza la implementación de un caso de estudio para visualizar la información de los lugares del campus y se analiza la interacción de los usuarios con la plataforma. Las soluciones presentadas en estos trabajos pueden ser complementadas agregando la posibilidad de actuar sobre el entorno con una comunicación bidireccional y la inclusión de actuadores en la arquitectura propuesta.

Otros enfoques buscan resolver problemas más concretos abarcando uno o diferentes contextos específicos dentro del campus inteligente, como la posibilidad de determinar la cantidad de ocupantes en un edificio inteligente (Lilis, Hoffet & Kayal, 2016), aulas de clase inteligentes (Gerval & Le Ru, 2016), (Bargaoui & Bdiwi, 2014) y (Yuliantoputri, Muhamad & Suhardi, 2019), sistemas de seguimiento inteligente similares al trabajo de (Tang, Shi & Lei, 2016), plataformas de parqueo inteligente (Somani, Periwal, Patel & Gaikwad, 2018), control inteligente del alumbrado del campus (Hussain, Qiu & Yue, 2019) y un gran número de investigaciones que dan solución a problemas particulares. En el trabajo de (Narendrakumar & Pillai, 2017) se presenta una arquitectura para monitorear de manera inteligente la temperatura y el uso del agua en un campus, así como una aplicación para administrar la infraestructura asociada, es interesante la capacidad de extensibilidad de esta última propuesta, que podría ser complementada con componentes para la administración de la infraestructura.

También podemos encontrar trabajos enfocados en las arquitecturas basadas en microservicios que buscan brindar soluciones en el contexto IoT y que no son explícitamente para campus inteligente, en este punto, es importante considerar aspectos como la escalabilidad, alta disponibilidad, tolerancia a fallos y administración del sistema; como el trabajo de (Miladinovic & Schefer-Wenzl, 2018) donde con las capacidades de la virtualización de funciones de red, se propone una visión de una arquitectura IoT para un entorno de sala de operaciones, en ese trabajo algunos casos de uso tienen un tiempo de respuesta crítico y se tienen en cuenta aspectos como la escalabilidad y alta disponibilidad, pero no está enfocado en la tolerancia a fallos o la administración del sistema.

En (Hasselbring & Steinacker, 2017) podemos encontrar una propuesta para E-commerce basada en microservicios, que presenta la descomposición de responsabilidades a través de los componentes, lo que permite comprender la propuesta de escalabilidad, agilidad y dos aspectos muy importantes en este contexto: la tolerancia a fallos y la fiabilidad de las transacciones, adicionalmente menciona la monitorización de los microservicios que es una tarea fundamental para la administración de la plataforma.

Finalmente encontramos trabajos con un enfoque interesante hacia el desarrollo de arquitecturas basadas en microservicios, en (Butzin, Golatowski & Timmermann, 2016) se presenta un estudio sobre las diferentes arquitecturas de microservicios enfocadas en la aplicación para el Internet de las Cosas, las técnicas de desarrollo, despliegue, tecnologías para la implementación, monitorización y orquestación que permitan la escalabilidad y autónoma de los servicios. En (Mayer & Weinreich, 2017) el trabajo está concentrado totalmente en la monitorización de microservicios, exponiendo el tipo de métricas que se pueden usar, la información recopilada y un tablero de visualización para el comportamiento o rendimiento de cada componente desplegado.

Todas estas soluciones presentadas están orientadas a resolver problemas particulares o presentar soluciones más generales en el contexto IoT, el trabajo que se presenta busca integrar las soluciones de cada contexto en una misma infraestructura para mejorar su administración, escalabilidad, autonomía y uso de recursos, a continuación, se presenta la arquitectura de la plataforma campus inteligente propuesta en esta investigación.

3. Plataforma Campus Inteligente

En esta sección se presenta el diseño de la plataforma campus inteligente, inicialmente el modelo conceptual, seguido por los conceptos principales y sus relaciones, finalmente se muestran las propiedades principales de la arquitectura tales como la noción de escalabilidad y autonomía propuestas.

3.1. Modelo Conceptual

La visión de campus inteligente de esta investigación se presenta mediante el diagrama del modelo conceptual en la figura 1.

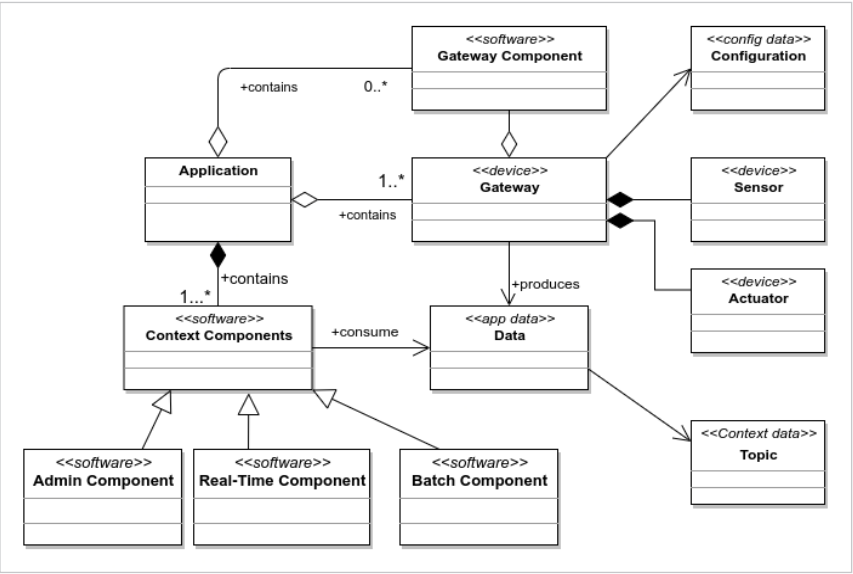


Figura 1 – Modelo Conceptual Plataforma Campus Inteligente

En este modelo una aplicación en la plataforma se entiende como:

- Un conjunto de gateways que integra sensores, actuadores y los componentes software desplegado en los gateways.
- Un backend que integra datos y permite administrar los gateways.
- Un conjunto de componentes de un contexto específico que pueden interactuar con los gateways a través del backend.

Un gateway es un dispositivo hardware con capacidad de comunicación y de cómputo. Los gateways están encargados de integrar sensores y actuadores que no tienen autonomía de comunicación directa, además en el gateway se pueden desplegar componentes software capaces de tareas específicas. Un gateway puede estar compartido entre varias aplicaciones.

El backend está encargado de la comunicación de doble vía con los gateways con tres objetivos principales: (i) administrar los gateways y sus dispositivos, (ii) integrar los datos que se recogen desde los gateways, (iii) permitir la comunicación entre los componentes externos de cada contexto y los gateways.

Los componentes de contexto proveen diversos tipos de funciones a las aplicaciones desplegadas en la plataforma: tomar decisiones en tiempo real basadas en los datos recogidos que impliquen utilizar los actuadores en el entorno, presentar interfaces de visualización de los datos recogidos, agregar una capa de administración específica a la plataforma, etc.

Para desplegar una aplicación en el campus inteligente se necesita entonces: (i) indicar los gateways que conformarán la aplicación, (ii) desplegar componentes software en los gateways, (iii) desplegar las componentes software del contexto específico.

3.2. Infraestructura

En la figura 2 se presenta la arquitectura general de plataforma campus inteligente y sus componentes. El primer componente es el backend compuesto por microservicios que se encargan de procesar las diferentes solicitudes y almacenar a largo plazo la información, también integra un bróker que permite una comunicación asincrónica y se encarga de recibir la información de cada tópico para enviarla a todas las aplicaciones suscritas, otro componente es el servidor frontend que está conformado por una aplicación de una sola página diseñada para el fácil acceso a las diferentes opciones de administración de los dispositivos IoT, el framework gateway es una herramienta que facilita la configuración y comunicación de los diferentes dispositivos, tiene acceso a un almacenamiento limitado y temporal que se borra periódicamente.

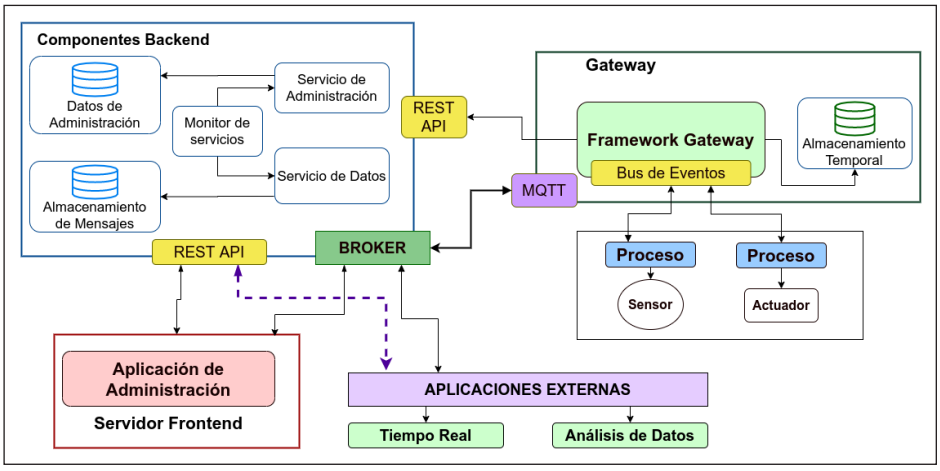


Figura 2 – Arquitectura Campus Inteligente

A continuación, se presenta brevemente cada componente de la arquitectura campus inteligente.

Backend

El backend es el encargado del procesamiento de las solicitudes de administración para componentes como gateways, procesos y dispositivos, también se encarga del almacenamiento a largo plazo de los datos enviados por los diferentes dispositivos de la plataforma. Está conformado por cinco componentes: dos bases de datos que permiten separar la información de administración de los mensajes recibidos en los diferentes tópicos y tres microservicios, el microservicio de administración que se conecta con su respectiva base de datos y es el encargado de registrar, modificar y eliminar elementos del sistema, el microservicio de datos que se encarga de la recepción, envío y almacenamiento de los diferentes datos generados en el sistema y finalmente el microservicio encargado de monitorear los demás microservicios para conocer su estado en tiempo real.

El bróker es un elemento fundamental para las comunicaciones y hace uso de protocolos con bajo consumo energético que son comunes en IoT, su función es recibir los mensajes provenientes de los gateways, almacenarlos temporalmente en una cola para que el backend pueda procesarlos y almacenarlos a largo plazo, también puede enviar los mensajes a las aplicaciones suscritas en tópicos específicos permitiendo la transmisión de datos en tiempo real.

Aplicación de Administración

La aplicación de administración fue diseñada para facilitar las diferentes tareas que se realizan en la plataforma, se contemplan dos tipos de usuarios: a) los administradores del sistema que pueden ver y gestionar su perfil, aplicaciones, gateways, dispositivos, datos y notificaciones, b) los super usuarios poseen las mismas capacidades de un administrador y adicionalmente pueden ver y gestionar las cuentas de los demás usuarios. La función principal de esta aplicación es permitir la administración de los dispositivos de la plataforma, por lo tanto, no se considera un rol de consulta de información y no se incluye la administración de los microservicios desplegados en el servidor backend, también posee la capacidad de recibir mensajes del bróker para actualizar en tiempo real el estado de los diferentes dispositivos desplegados en la plataforma.

Gateway

En la figura 3 se puede apreciar en detalle la arquitectura de los componentes desplegados dentro del gateway, estos componentes se diseñaron pensando en la autonomía y extensibilidad, el framework diseñado hace uso de diferentes microservicios para el control de procesos, almacenamiento temporal de datos, recepción de instrucciones de administración y envío de mensajes al bróker, el bus de eventos es el encargado de la comunicación con los procesos, el servicio web permite la comunicación con el backend y el cliente bróker permite el envío de mensajes, los servicios comunes contienen la lógica para procesar las solicitudes y configuraciones del gateway.

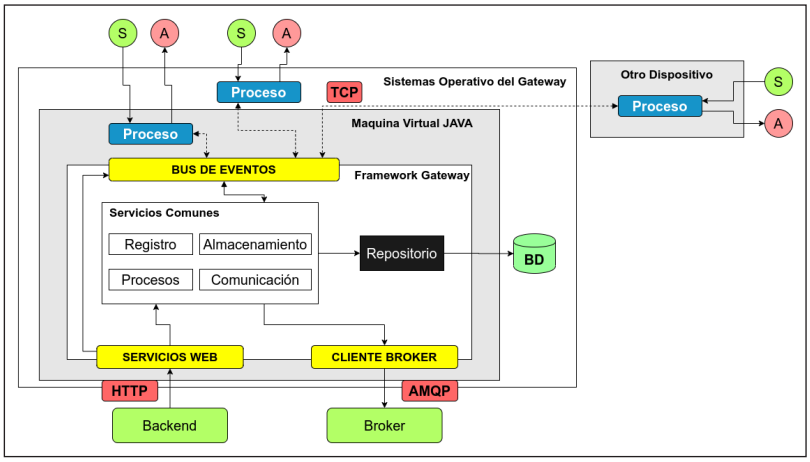


Figura 3 – Arquitectura Gateway.

3.3. Propiedades

A continuación, se exponen las propiedades de escalabilidad y las propiedades de autonomía de la plataforma campus inteligente.

Escalabilidad

La propuesta de escalabilidad para la plataforma campus inteligente está orientada a la infraestructura y aplicaciones, como se aprecia en la figura 4, de acuerdo con las necesidades propias de cada contexto, el primer eje está orientado al nivel de

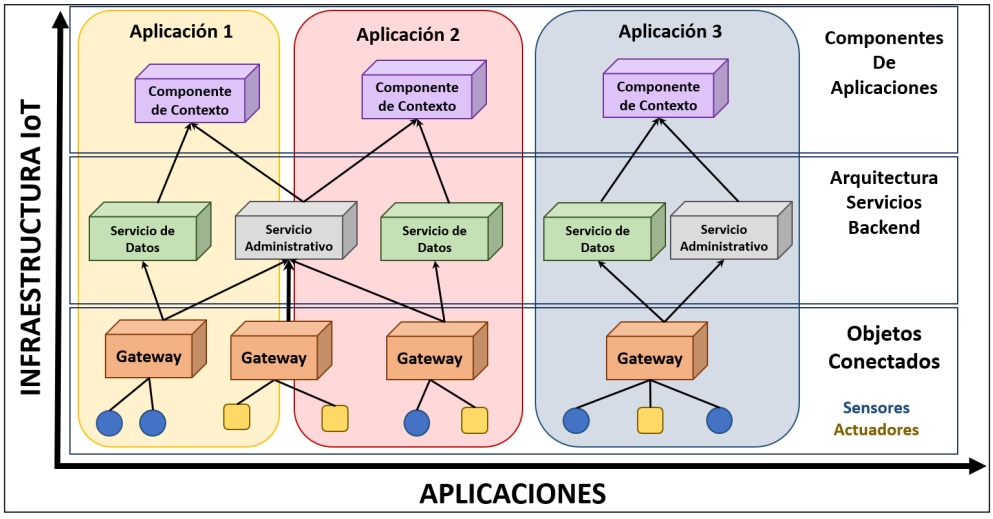


Figura 4 – Escalabilidad Plataforma Campus Inteligente.

aplicaciones, ya que en la plataforma se pueden desplegar diferentes aplicaciones que compartan la misma infraestructura. Un gateway y sus respectivos sensores y actuadores pueden ser compartidos por dos aplicaciones, por ejemplo, los datos provenientes de un sensor de luz podría ser usados tanto por una aplicación de parqueo inteligente, cómo por una aplicación de monitoreo ambiental; adicionalmente los desarrolladores de cada aplicación pueden enlazar componentes específicos de cada contexto para cumplir requerimientos del modelo de negocios, en este sentido la plataforma puede escalar en el número de aplicaciones que soporta dentro del campus inteligente.

El segundo eje está orientado a la escalabilidad de la infraestructura IoT desplegada en el campus, al aumentar el número de aplicaciones, también es necesario soportar el despliegue de nuevos dispositivos (sensores, actuadores o gateways), nuevos componentes software, y el despliegue de diferentes instancias de los servicios que brindan soporte a una o muchas aplicaciones, en este sentido la plataforma permite escalar el número de componentes acorde a la demanda de las aplicaciones desplegadas.

Autonomía

La motivación de trabajar con computación autónoma surge de la necesidad de dotar a los sistemas IoT de cierto grado de autonomía, permitiendo facilitar las labores de gestión, adaptabilidad, corrección de fallos y tareas generales sin la necesidad de intervención humana a medida que aumenta la complejidad y tamaño de los componentes instalados.

Desde la perspectiva arquitectónica del campus inteligente se pueden definir niveles de autonomía en cada uno de sus componentes, en la Figura 5 se presenta el modelo de adopción autonómico propuesto por IBM (Lalanda, McCann & Diaconescu, 2013) para hacer la transición del componente de microservicios existente al nivel de madurez de los futuros sistemas autónomos. Donde el eje x corresponde al aumento de las funciones autónomas y el eje y a los alcances crecientes sobre los cuales se pueden aplicar las funciones autónomas.

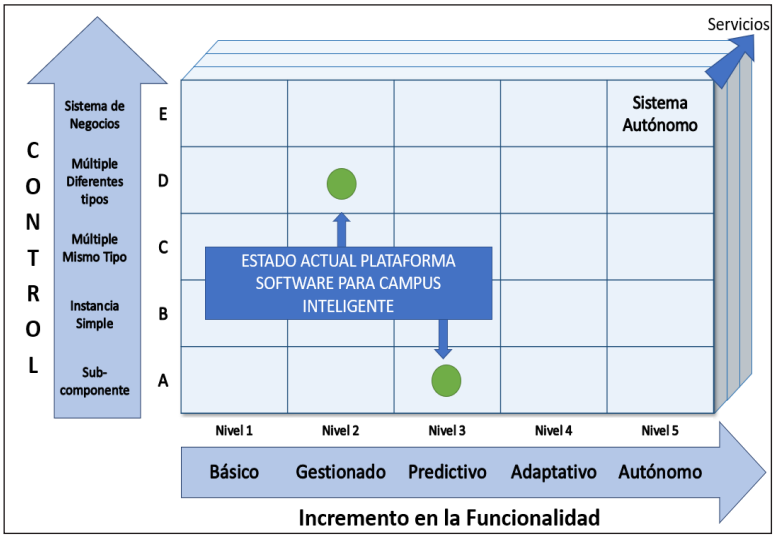


Figura 5 – Modelo para la adopción de computación autónoma de IBM.

Cada nivel de funcionalidad adaptado al control de componentes más complejos permite mejorar la autonomía del sistema, los tres primeros niveles abordados en este trabajo pueden ser definidos con un enfoque a microservicios de la siguiente forma:

Nivel 1 - Básico: los profesionales de TI gestionan cada microservicio de forma independiente. Se considera el punto de partida para la implementación del sistema autonómico.

Nivel 2 - Gestionado: se utilizan sistemas de gestión para recopilar y presentar la información de los microservicios asociados a consulta de estados, almacenamiento y comunicación con gateways. La arquitectura presentada se sitúa en este nivel ya que es capaz de monitorear el estado de los componentes desplegados, independientemente de su tipo.

Nivel 3 - Predictivo: se dota el sistema con capacidades de análisis para correlacionar datos e identificar patrones de los microservicios. Al utilizar contenedores, se pueden monitorizar el rendimiento y el estado del entorno de ejecución mediante las métricas de consumo de energía, CPU, memoria y red, host, estado y periodos de actividad; o métricas a nivel del nodo como almacenamiento, contenedores y versiones.

4. Validación

La validación del diseño propuesto se realizó mediante la implementación de un prototipo de la plataforma campus inteligente, también se desplegó un caso de estudio simple, a continuación, se presentan los resultados obtenidos.

4.1. Prototipo Plataforma Campus Inteligente

La implementación de un prototipo de la plataforma campus inteligente se realizó mediante un conjunto de herramientas de código abierto que permiten la independencia tecnológica y reducción de costos, los protocolos de comunicación usados son HTTP para la transferencia de información hacia el backend, AMQP y MQTT para el envío de mensajes principalmente desde el gateway hacía el bróker y las aplicaciones suscritas, finalmente TCP para la comunicación entre los procesos y el gateway.

En la implementación del backend se usó el lenguaje de programación JAVA y el framework Spring Boot, también se usó MySQL en la base de datos de administración y MongoDB en la base de datos de mensajes el bróker no fue implementado, en su lugar se usó un bróker ya existente dentro de la comunidad open source llamado ActiveMQ.

La aplicación frontend de administración se realizó mediante tecnologías como HTML, CSS, JavaScript y AngularJS para construir una aplicación de una sola página, en la figura 6 se puede ver el menú principal de la aplicación de administración del sistema, donde se pueden gestionar los diferentes gateways, procesos, dispositivos, aplicaciones y (en caso de ser super usuario) usuarios.

La implementación de los microservicios que componen el framework gateway se realizó en el lenguaje de programación JAVA y el framework Spring Boot, la base de datos usada fue MongoDB, además existe la posibilidad de desplegar un proceso de tres formas distintas: la primera, es desplegar el proceso en la máquina virtual de JAVA haciendo

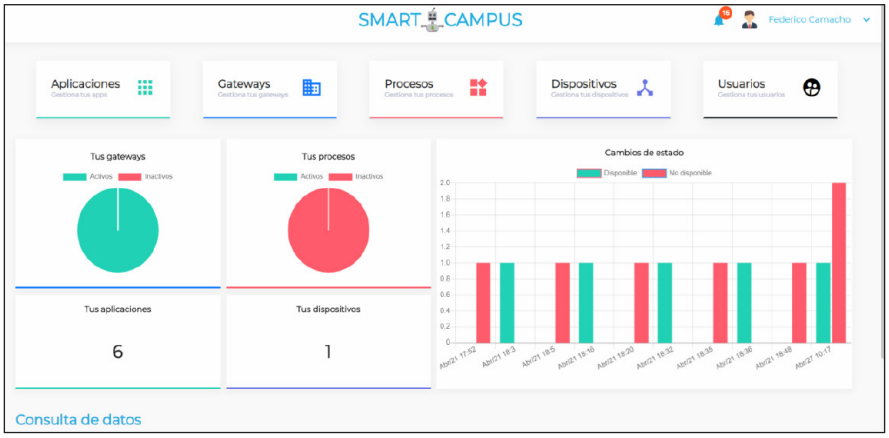


Figura 6 – Menú principal aplicación de administración.

uso del framework implementado VertX , esto permite una comunicación directa con el framework gateway y la posibilidad de desplegar procesos de forma remota, la segunda forma es desplegar el proceso en el mismo sistema operativo del gateway haciendo uso del lenguaje de preferencia y la última forma es desplegar el proceso en un dispositivo externo, en las últimas dos alternativas la comunicación se hace a través del bus de eventos por el protocolo TCP, el código fuente de la plataforma pueden encontrarse en el repositorio del grupo de investigación CAGE – UIS (CAGE, 2020).

4.2. Caso de Estudio Laboratorio Smart Campus

El contexto del caso de estudio es un equipo en un laboratorio del campus donde se debe monitorear constantemente la temperatura al interior y el voltaje de llegada, estos valores deben estar en unos rangos establecidos y en caso de registrar valores anormales se deben tomar acciones lo más rápido posible. Para monitorear y generar alarmas en caso de aumento de temperatura o caída de voltaje se implementaron los dispositivos que se muestran en la figura 7, un dispositivo Intel galileo donde se desplegaron dos procesos programados en Python encargados de recibir y enviar los valores de temperatura y voltaje a un dispositivo Raspberry pi A en el que se encuentra desplegado el framework gateway.

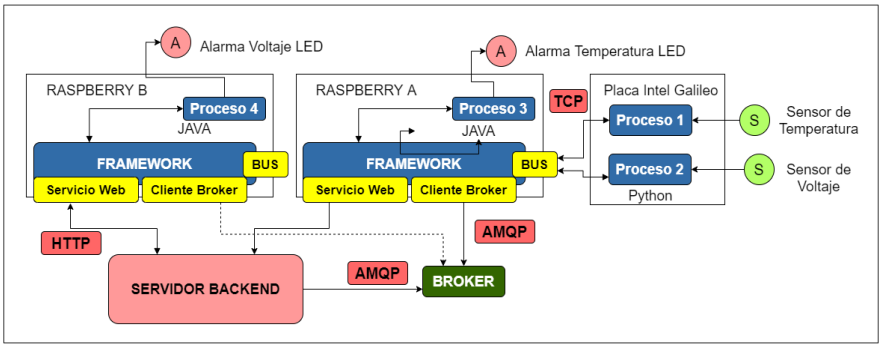


Figura 7 – Caso de estudio equipo de laboratorio.

La tarjeta Raspberry pi A además de recibir y enviar los datos de los dos sensores, tiene conectado el actuador más importante, una alarma de temperatura que activa un LED cuando la temperatura registrada aumenta; adicionalmente existe un gateway Raspberry B que tiene desplegado el proceso para activar la alarma de voltaje, que será disparada cuando el voltaje disminuya del umbral determinado.

Los gateways Raspberry A y Raspberry B solo se comunican a través del bróker y los diferentes servicios del servidor backend; al realizar las pruebas con los sensores variando la temperatura y el voltaje se activaron adecuadamente las alarmas.

5. Conclusiones y Trabajo Futuro

La plataforma Smart Campus presentada realiza dos aportes centrales, el primero un modelo conceptual utilizado para el desarrollo de aplicaciones de Campus Inteligentes basadas en IoT, y el segundo, una infraestructura software que soporta este modelo conceptual y provee un conjunto de propiedades para facilitar el desarrollo de las aplicaciones de este tipo.

La plataforma Smart Campus provee la capacidad de ser utilizada en diversos contextos específicos de un campus universitario inteligente, permitiendo obtener una mayor independencia tecnológica, reducción de costos de implementación y la capacidad de adaptarse a las necesidades específicas de cada universidad.

Tanto la arquitectura del backend, cómo la arquitectura del framework gateway fueron diseñadas pensando en ser extensibles, por lo que es posible introducir nuevos microservicios para soportar más funcionalidades en estos componentes de acuerdo con las necesidades específicas de cada campus.

La validación mediante la implementación de un prototipo permitió comprobar la compatibilidad entre los componentes, métodos y protocolos de comunicación, así como el conjunto de tecnologías seleccionadas para el despliegue de la plataforma.

Como trabajo futuro se puede implementar o desplegar modelos predictivos dentro de la arquitectura, y extender el nivel de monitorización de los elementos no solo a microservicios y procesos, de esta forma se puede alcanzar un mayor nivel de autonomía dentro de la arquitectura propuesta.

Es prudente realizar un plan de pruebas que permita documentar detalladamente el rendimiento, escalabilidad y extensibilidad de la arquitectura en diferentes escenarios y a mayor escala, teniendo en cuenta más métricas de trabajo, y escenarios donde se puedan integrar un mayor número de dispositivos IoT.

Para los administradores sería de utilidad visualizar el estado de los diferentes microservicios desplegados, sus instancias y las instancias de bases de datos, todo esto haciendo uso de herramientas de administración existentes o proponiendo nuevos componentes que aprovechen los monitores autónomos que ya están integrados en la arquitectura.

Referencias

Adamkó, A., Kãdek, T., & Kósa, M. (2014, November). Intelligent and adaptive services for a smart campus. In 2014 5th *IEEE Conference on Cognitive Infocommunications* (CogInfoCom) (pp. 505-509). IEEE. doi: 10.1109/CogInfoCom.2014.7020509.

Bargaoui, H., & Bdiwi, R. (2014, November). Smart classroom: Design of a gateway for ubiquitous classroom. *In 2014 International Conference on Web and Open Access to Learning* (ICWOAL) (pp. 1-4). IEEE. doi: 10.1109/ICWOAL.2014.7009206.

Butzin, B., Golatowski, F., & Timmermann, D. (2016, September). Microservices approach for the internet of things. In 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 1-6). IEEE. doi: 10.1109/ETFA.2016.7733707.

CAGE. (2020, 07 20). *Repositorio grupo de investigación CAGE*. Retrieved from https://github.com/SC3UIS/smart-campus.

Gerval, J. P., & Le Ru, Y. (2016). Smart classroom. In Smart education and e-learning 2016 (pp. 415-422). *Springer, Cham*. doi:10.1007/978-3-319-39690-3_37.

Hasselbring, W., & Steinacker, G. (2017, April). Microservice architectures for scalability, agility and reliability in e-commerce. *In 2017 IEEE International Conference on Software Architecture Workshops* (ICSAW) (pp. 243-246). IEEE. doi: 10.1109/ICSAW.2017.11.

Hussain, B., Qiu, C., & Yue, C. P. (2019, October). Smart Lighting Control and Services Using Visible Light Communication and Bluetooth. In 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE) (pp. 1-2). IEEE. doi: 10.1109/GCCE46687.2019.9014635.

Lilis, G., Hoffet, A., & Kayal, M. (2016, April). GeoAware: A hybrid indoor and outdoor localization agent for smart buildings. *In 2016 18th Mediterranean Electrotechnical Conference* (MELECON) (pp. 1-6). Ieee. doi: 10.1109/MELCON.2016.7495392.

Lalanda, P., McCann, J. A., & Diaconescu, A. (2013). Autonomic computing: principles, design and implementation. *Springer Science & Business Media*. (pp 41-42) ISBN 978-1-4471-5007-7.

Mayer, B., & Weinreich, R. (2017, April). A dashboard for microservice monitoring and management. In 2017 IEEE International Conference on Software Architecture Workshops (ICSAW) (pp. 66-69). IEEE. doi: 10.1109/ICSAW.2017.44.

Miladinovic, I., & Schefer-Wenzl, S. (2018, February). NFV enabled IoT architecture for an operating room environment. *In 2018 IEEE 4th World Forum on Internet of Things* (WF-IoT) (pp. 98-102). IEEE. doi: 10.1109/WF-IoT.2018.8355128.

Monti, L., Prandi, C., & Mirri, S. (2018, November). Iot and data visualization to enhance hyperlocal data in a smart campus context. In Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good (pp. 1-6). doi: 10.1145/3284869.3284878.

Narendrakumar, T., & Pillai, A. S. (2017, July). Smart connected campus. In 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (pp. 1591-1596). IEEE. doi: 10.1109/ICICICT1.2017.8342808.

Omale, G. (2018, November) Gartner Identifies Top 10 Strategic IoT Technologies and Trends. Gartner Symposium/ITxpo 2018. Recuperado desde: https://www.gartner.com/en/newsroom/press-releases/2018-11-07-gartner-identifies-top-10-strategic-iot-technologies-and-trends.

Pagliaro, F., Mattoni, B., Gugliermenti, F., Bisegna, F., Azzaro, B., Tomei, F., & Catucci, S. (2016, June). A roadmap toward the development of Sapienza Smart Campus. In 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC) (pp. 1-6). IEEE. doi: 10.1109/EEEIC.2016.7555573.

Prandi, C., Monti, L., Ceccarini, C., & Salomoni, P. (2019). Smart campus: Fostering the community awareness through an intelligent environment. Mobile Networks and Applications (pp. 1-8). doi: 10.1007/s11036-019-01238-2.

Somani, A., Periwal, S., Patel, K., & Gaikwad, P. (2018, January). Cross Platform Smart Reservation Based Parking System. In 2018 International Conference on Smart City and Emerging Technology (ICSCET) (pp. 1-5). IEEE. doi: 10.1109/ICSCET.2018.8537282.

Tang, H., Shi, J., & Lei, K. (2016, June). A smart low-consumption IoT framework for location tracking and its real application. In 2016 6th international conference on electronics information and emergency communication (iceiec) (pp. 306-309). IEEE. doi: 10.1109/ICEIEC.2016.7589744.

Yuliantoputri, A. R., Muhamad, W., & Suhardi, S. (2019, November). Smart Classroom Services System Design Based on Services Computing System. In 2019 International Conference on ICT for Smart Society (ICISS) (Vol. 7, pp. 1-6). IEEE. doi: 10.1109/ICISS48059.2019.8969809.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 12/08/2020
Aceitação/Acceptance: 26/10/2020

Diseño analítico de reglas de decisión en procesos timetabling para reasignación de espacios académicos

Monica K. Duran-Vaca[1], J. A. Ballesteros-Ricaurte[2], A. M. González-Amarillo[3]

monica.duran@uptc.edu.co; javier.ballesteros@uptc.edu.co; angela.gonzalez@unad.edu.co

[1] Universidad Pedagógica y Tecnológica de Colombia, 152211, Sogamoso, Boyacá, Colombia.

[2] Universidad Pedagógica y Tecnológica de Colombia, 150003, Tunja, Boyacá, Colombia.

[3] Universidad Nacional Abierta y a Distancia, 150001, Tunja, Boyacá, Colombia.

Pages: 283-296

Resumen: Un desafío en la gestión académica de las instituciones educativas es la asignación de espacios académicos. Se requiere contar con una planificación de los recursos escolares que cumpla con todas las restricciones impuestas y que se obtenga una solución aceptable en un tiempo moderado, la cual es considerada como una tarea computacionalmente compleja. Por tal razón, en este artículo se crearon reglas de decisión generadas por un algoritmo de árboles de decisión. Este proceso se realizó a partir de la identificación de variables que se seleccionaron mediante la técnica de clasificación Random Forest, que permitió la toma de decisiones frente a la posible configuración de nuevos parámetros en el software optimizador de espacios contribuyendo a una mejora del proceso, para lo cual se plantearon indicadores estadísticos sustentados con técnicas de análisis multivariado, siendo la base para seleccionar e identificar variables relevantes a partir de técnicas de Machine Learning.

Palabras-clave: Árboles de decisión; Modelo analítico; Random forest; Timetabling.

Analytical design of decision rules in timetabling processes to reassign academic spaces.

Abstract: A challenge in the academic management of educational institutions is the allocation of academic spaces. It is required to have a planning of the school resources that complies with all the imposed restrictions and that an acceptable solution is obtained in a moderate time, which is considered as a computationally complex task. For this reason, this article created decision rules generated by a decision tree algorithm. This process was carried out from the identification of variables that were selected using the Random Forest classification technique, which allowed decision-making regarding the possible configuration of new parameters in the space optimizer software, contributing to an improvement of the process, for which statistical indicators supported by multivariate analysis techniques were proposed, being the basis for selecting and identifying relevant variables from Machine Learning techniques.

Keywords: Decision Trees; Analytical model; Random Forest; Timetabling;

1. Introducción

Las instituciones educativas en Colombia basadas en su infraestructura y la programación previa del horario de clases realizan una asignación apropiada de sus espacios, mediante el uso de sistemas de información para optimizar el proceso. Sin embargo, cuando esta asignación inicial ocurre se presenta un porcentaje considerable de solicitudes de cambio. Por esta razón se toma como caso de estudio una problemática real de asignación de espacios académicos en una universidad privada que tiene 6600 solicitudes iniciales (Durán Vaca, Gómez Moreno, & Ramírez Aguirre, 2019). Se presentan 1900 solicitudes de cambio de espacio físico aproximadamente (Durán Vaca et al., 2019), lo que conlleva a realizar manualmente ajustes sobre dichos cambios.

La gestión de reserva y asignación de espacios para el desarrollo de actividades institucionales, académicas y culturales es considerada como un problema de timetabling (Ferland & Roy, 2018; Paechter, Gambardella, & Rossi-Doria, 2002; Perumal, Tabassum, Norwawi, Narayana, & Perumal, 2018). Las instituciones educativas se enfrentan a estos problemas y se han planteado diferentes formas de resolverlo desde la informática como desde la investigación de operaciones. Este proceso en la teoría de la complejidad computacional se clasifica en la clase de complejidad NP-completo (Even, Itai, & Shamir, 1975) con solución polinómica lo que la convierte de difícil solución. Al ser un problema complejo que requiere la optimización de múltiples objetivos y que utiliza diversos criterios o vectores, se hace necesario el uso de métodos heurísticos de solución como, algoritmos genéticos evolutivos, búsqueda tabú, colonia de hormigas, modelo general de redes, entre otras (Alsuwaylimi & Fieldsend, 2019). La adición de detalles o cambios por la complejidad hacen que los modelos no sean tan precisos, lo que conlleva a que diversas instituciones e investigadores unan esfuerzos para poder dar respuesta a la problemática.

En la revisión de literatura, se encuentra que muchos estudios basados en diferentes estrategias han logrado grandes avances al respecto, dentro de ellos encontramos: a) una propuesta de (Calle-López et al., 2018) que se fundamenta en un sistema experto basado en minería de datos y programación entera lineal para soporte en la asignación de materias y diseño de horarios en educación superior; b) en la Universidad Francisco de Paula Santander Ocaña (Vergel Miranda, 2017), se realizó un algoritmo para la asignación de horarios utilizando técnicas de inteligencia artificial; c) en la Universidad Francisco José de Caldas (Díaz Cortes, García, & Hernandez, 2015), desarrollaron una herramienta basada en técnicas de inteligencia artificial, para la asignación de horarios y recursos académicos, en el proyecto curricular de ingeniería de sistemas y d) en la Universidad de Carabobo (Cuadrado Guevara, 2014), se planteó un modelo matemático para el problema de horarios del Departamento de matemáticas de la Facultad de Ciencias y Tecnología.

En el presente trabajo se realiza el diagnóstico de las variables de mayor incidencia o probabilidad en la generación de cambios en el proceso de asignación de salones al comienzo de un semestre académico. Este análisis se realiza a través de técnicas de Machine Learning utilizando algoritmos de clasificación (Almeida, Medeiros, &

Oliveira, 2016; Wang, Zhou, Li, Zhang, & Dong, 2019) que permiten abordar el problema de disminución del número de solicitudes de reasignación o cambio de salones. Una vez se ha publicado la asignación para un semestre, se genera un alto volumen de cambios que les implica realizar manualmente los ajustes solicitados. Se deben volver a tener en cuenta el conjunto de restricciones duras y blandas (Alsuwaylimi & Fieldsend, 2019; Yang, Ayob, & Nazri, 2018), lo cual hace que se deba tener en cuenta en primera instancia las restricciones duras que son las que permiten mantener una asignación factible y óptimo. En segunda medida tener presentes las restricciones blandas tanto como sea posible ya que son las que dan la calidad de los horarios.

El objetivo principal de este artículo es presentar las reglas de decisión generadas mediante el uso de un modelo analítico y un algoritmo de clasificación. Las variables que influyen en que se presenten solicitudes de cambios son: capacidad del salón, número de inscritos, cupos, profesores, secciones, solicitudes especiales, características de los salones y franjas horarias. Y así, permitir la toma de decisiones frente a la posible configuración de nuevos parámetros en el software optimizador de espacios para contribuir a una mejora del proceso de asignación de espacios académicos.

El documento está estructurado de la siguiente manera: en la sección 2, se describe el análisis de la situación actual donde se explica el proceso de asignación de salones. En la sección 3, se presenta la metodología, se describen las etapas utilizadas en esta propuesta. La sección 4 muestra el diseño de la solución propuesta. La sección 5 informa los resultados y discusión de los resultados de cada proceso de diseño propuesto. La sección 6 permite observar los trabajos relacionados. Por último, la sección 7 da la conclusión del trabajo realizado.

2. Análisis de la situación actual

Después del proceso de asignación de espacios académicos que se realiza al inicio de cada semestre, se genera un alto número de solicitudes de cambio de salón, aproximadamente un 28% del total de peticiones iniciales. Esta situación exige un trabajo adicional de tiempo y personal, ya que se genera otra actividad al proceso estándar de asignación, que consiste en registrar en una herramienta alterna las solicitudes de cambio de salón de aquellos que fueron resultado “Aceptado” en la asignación masiva. Este proceso conlleva a una nueva tarea de negociación y un registro manual realizando un cambio forzoso sobre el sistema automatizado que representa un potencial foco de errores en la calidad de los datos y en el proceso. Se espera que el resultado del proceso generado en el sistema de optimización de recursos de espacio sea el más cercano al resultado requerido por los coordinadores académicos.

En la Figura 1 se muestran las etapas del proceso de asignación de espacios; cada tarea, recurso y herramientas que interactúan en dicho proceso se explican a continuación.

1. Registro Inicial. Tres meses antes de iniciar un nuevo ciclo académico, se inicia el proceso de asignación de salones, habilitando la plataforma a los coordinadores académicos para la solicitud básica de tipología de espacios, de acuerdo con la planeación de secciones que se tenga para un periodo académico.

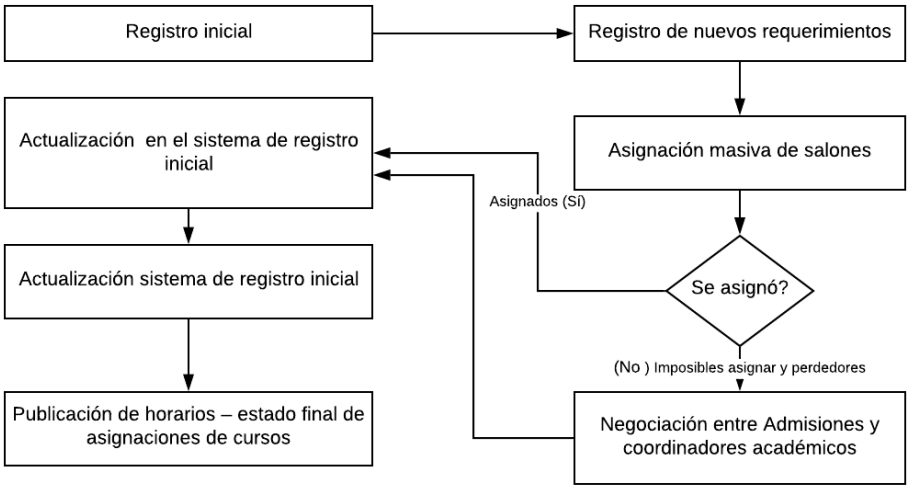


Figura 1 – Diagrama de procesos de asignación de espacios y recursos

2. Registro de nuevos requerimientos a la solicitud inicial. Posterior a la solicitud de salón inicial, se activa una aplicación tipo formulario para las coordinaciones académicas. Se pueden detallar o complementar con mayor exactitud la asignación de salones iniciales para los casos en los cuales se requieran características puntuales y que no fueron incluidas a través del registro inicial.
3. Asignación masiva de salones. De los pasos previos se obtiene una solicitud para que el sistema optimizador de espacios, plataforma utilizada para la reserva de espacios y recursos disponibles, se ejecute buscando cumplir con todas las solicitudes ingresadas. Se busca maximizar el espacio conforme a las parametrizaciones que le han sido configuradas. En este punto se tienen tres grupos principales de salones:
   - Salones asignados: los casos en los cuales se encontró un salón.
   - Salones imposibles de asignar: solicitudes con características de salones que no existen en la universidad. Estos casos solo son detectados por el sistema optimizador de espacios, quien tiene las parametrizaciones referentes a los espacios para asignar un salón - curso.
   - Cursos sin salón o perdedores: solicitudes de salones que existen en la universidad pero que ya han sido asignados a otras secciones.
4. Salones asignados o imposibles de asignar. Esta condición establece que, si los salones pertenecen al grupo de los asignados, se configuran en el sistema de registro inicial realizando una actualización de parámetros a partir de las peticiones realizadas por los coordinadores académicos. Aquellos salones incluidos dentro del grupo imposibles de asignar y perdedores pasarán por un proceso de negociación.
5. Negociación entre el administrador del sistema de horarios y coordinadores académicos. Los casos que quedan registrados en los grupos de salones

imposibles de asignar y perdedores, son reportados a los coordinadores académicos. Durante dos semanas aproximadamente, se encargan de buscar la asignación que mejor se ajuste a su requerimiento hasta dejar definido un salón conforme a la disponibilidad y las necesidades del curso. Este acuerdo es registrado manualmente en el sistema optimizador de espacios.

6. Actualización del sistema inicial. Es necesaria la actualización de los parámetros que involucran aquellos espacios que fueron identificados fuera de los asignados debido a las peticiones de nuevos requerimientos hechas por los coordinadores académicos. Se ajusta el proceso para la posterior inscripción y publicación de horarios.
7. Inscripción de materias. Luego de obtener la asignación de espacios (salón - curso o sección) actualizando los parámetros del sistema inicial, empieza el proceso de inscripción de materias por parte de los alumnos. Durante este periodo, se puede presentar un primer cambio de salones debido a la necesidad de aumento de cupo en las secciones.
8. Publicación de horarios Banner (cursos - salones). Una vez culminado los subprocesos anteriores, se publican finalmente los horarios para los alumnos, profesores y cursos con sus respectivos salones. En este momento se genera un alto volumen de cambios, los cuales deben ser solucionados en lo posible la primera semana de clases.

Esta situación es una oportunidad para realizar un análisis de la información generada por los sistemas de información involucrados en el proceso y poder hacer una caracterización de las variables relacionadas con la tipología y recursos de los salones. Las solicitudes de cambio en el proceso de asignación y reasignación de salones son los insumos para obtener indicadores a partir de un modelo analítico de clasificación que permita definir la eficiencia (menor número de solicitudes de cambio de salón) de dicho proceso.

3. Metodología

El enfoque utilizado para esta investigación implicó cinco etapas principales: 1) Selección de indicadores estadísticos, 2) Análisis descriptivo, 3) Selección de variables, 4) Implementación de árboles de decisión y 5) Reglas de decisión. Todas las herramientas necesarias para el diseño analítico de reglas de decisión en procesos timetabling y para la implementación de un algoritmo de clasificación de modelos de aprendizaje automático, se crean internamente y se implementan en Python. La metodología se muestra en la Figura 2.

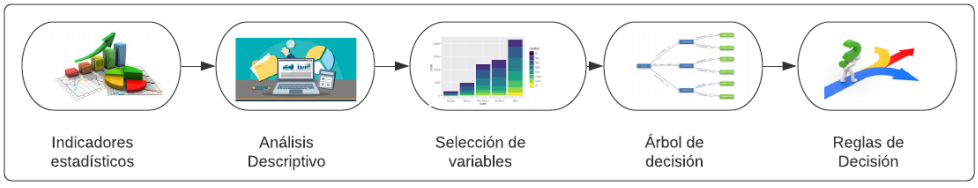


Figura 2 – Etapas de la solución propuesta

La problemática planteada en el presente trabajo se basa en el perfilamiento de tipos de salones, segmentación y caracterización del comportamiento de las solicitudes de salón,

con el fin de disminuir el número de solicitudes de cambio, basados en los resultados generados por el sistema optimizador de espacios y el tipo de solicitudes hechas por los coordinadores del programa. Con base en las fuentes de datos suministradas y el conocimiento de la situación actual, se definen los indicadores estadísticos a partir de la hipótesis sobre las variables que influyen en el proceso de asignación de salones, sustentados sobre análisis bivariado y multivariado, los cuales se mencionan a continuación:

- Distribución del número de inscritos y tipo de solicitud. Esta responde a la idea que el número de inscritos es determinante en el aumento o disminución de una solicitud de cambio.
- Existe una relación entre el cupo original, cupo final y los inscritos. Esto responde a la idea que las diferencias del cupo original frente al cupo final generan altas solicitudes de cambio.
- Existe una relación entre el número de inscritos y la capacidad del salón. Esta idea se sustenta en el hecho que los salones con una capacidad inferior son los que menos cambios generan.
- Existe una relación de facultad por tipo de cambio. Esta idea se sustenta en que existen facultades que solicitan con frecuencia cambios de espacio por una razón particular, ya sea por tipo de mobiliario o por el bloque donde el salón se encuentre.
- Distribución de variables facultad y franja horaria. El cambio de espacio solicitado por una facultad podría darse debido el horario establecido al inicio del proceso de asignación de espacios.

El cupo solicitado por los coordinadores académicos y la capacidad asignada por el sistema optimizador de espacios están relacionados.

Después de la definición de los indicadores, se realiza un análisis descriptivo (Harvill, 2012), el cual permite describir tendencias claves, calcular las medidas simples de composición y la distribución de variables identificando las principales características de un grupo de datos. Esto permite entender y conocer aquellas variables que deben accionarse para abordar el problema que se pretende resolver para posteriormente poder aplicar una técnica de análisis de acuerdo con el objetivo planteado (Adams et al., 2011).

Para soportar la selección de las variables que se tienen en cuenta para validar el resultado del modelo de clasificación, se toman las variables que, para el administrador del sistema, soportado por correlación y tablas de contingencia, son las más relevantes en las solicitudes de cambio de salón. Se aplica el algoritmo basado en árboles extremadamente aleatorios (Extra Trees Classifier) y por medio de la extracción aleatoria se clasifican las variables de importancia. Este es un método de aprendizaje conjunto basado fundamentalmente en árboles de decisión, el cual aleatoriza ciertas decisiones y subconjuntos de datos para minimizar el *sobre-aprendizaje* de los datos y el *sobre-ajuste*.

En la Figura 3 se muestra el cálculo de la importancia de cada variable. El algoritmo Extra Trees Classifier se selecciona ya que es el que permite agrupar varios árboles de clasificación y divide nodos usando subconjuntos aleatorios de características. Además, selecciona la cantidad de variables de manera aleatoria y construye árboles individuales, para luego realizar la clasificación por el orden de importancia resaltando que en la aleatoriedad no existe arranque de datos, sino que proviene de las divisiones aleatorias de todas las observaciones.

```
#Importamos desde sklearn.ensamble, el algoritmo de Random Forest
from sklearn.ensemble import RandomForestClassifier

#Inicializamos el algoritmo Random Forest e indicamos el número de árboles que vamos a construir
classifier=RandomForestClassifier(n_estimators=25)

#Construimos el modelo sobre los datos de entrenamiento
classifier=classifier.fit(pred_train,tar_train)

#Predecimos para los valores del grupo Test
predictions=classifier.predict(pred_test)

from sklearn.metrics import accuracy_score
```

Figura 3 – Uso de algoritmo Extra Trees Classifier

Una vez se tienen las variables más relevantes en el proceso de asignación y cambio de salón, se aplica un modelo de árboles de decisión(extra-árboles) extremadamente aleatorios (Witten, Frank, & Hall, 2011) como se muestra en la Figura 4. En consecuencia, se crean las reglas de decisión que permiten conocer aquellas variables que pueden estar ocasionando las solicitudes de cambio y en un siguiente proceso que no contempla esta propuesta, poder configurar la parametrización del sistema optimizador de espacios. Las medidas tenidas en cuenta para la construcción del árbol de decisión con el fin de obtener la mejor opción son la entropía y la ganancia de información (Patil & Kulkarni, 2019). La entropía es la medida del grado de incertidumbre asociado a una distribución de probabilidad; en una distribución uniforme, todos los valores son igualmente probables Pi = 1/N y por tanto la entropía es máxima, lo cual indica máxima incertidumbre (Pisane, Marée, Wehenkel, & Verly, 2010). La ganancia de información es un criterio para elegir un split: seleccionar el split con la mayor ganancia de información (Patil & Kulkarni, 2019); esta medida ayuda a conocer el valor de una variable aleatoria *X* para conocer el verdadero valor de otra *Y* (Pisane et al., 2010).

```
from sklearn import metrics #Import scikit-learn metrics module for accuracy calculation
from sklearn.metrics import accuracy_score
import sklearn
from sklearn.metrics import classification_report
```

```
sklearn.metrics.confusion_matrix(tar_test,predictions)
```

```
array([[ 9828,  2316],
       [ 1020, 10872]], dtype=int64)
```

```
#Sacamos el índice Accuracy Score, que resume la Matriz de Confusión y la cantidad de aciertos.
sklearn.metrics.accuracy_score(tar_test, predictions)
```

```
0.8612081877184223
```

Figura 4 – Árboles de Decisión

La selección de este tipo de algoritmo (modelo de clasificación) se sustenta debido a que es el que más se ajusta dada la información que aporta, teniendo en cuenta la variable objetivo 'cambio'. Además, los tipos de datos presentes en las variables de análisis, los de mayor presencia son categóricos y este algoritmo es adecuado para esta tipología de los datos. Del resultado obtenido con este modelo se obtienen reglas de la forma

SI (Condición) - *ENTONCES* (Decisiones), que definen las características asociadas al cambio de salón a partir de las variables que más inciden en este proceso.

4. Resultados

La descripción de las etapas de la metodología y resultados se explican a continuación.

4.1. Validación de indicadores estadísticos

Para sustentar cada uno de los indicadores estadísticos definidos como etapa para la selección de variables relevantes en el algoritmo de clasificación, se realiza una matriz de coeficientes de correlación ya que se tienen variables cuantitativas, en donde la medida del grado de asociación lineal entre las dos variables cuantitativas es: r_{ry}=[-1, 1] para relaciones negativas o positivas respectivamente y cero para variables independientes. También se aplican tablas de contingencia para analizar la distribución de porcentajes entre variables.

En la Figura 5 se muestra el resultado de la matriz de correlación para el indicador estadístico, donde los valores de la diagonal principal de la matriz están dados por la unidad: $S_x^2/(S_x \times S_x) = 1$ y el restante de los términos oscilan entre +1 y -1. Este acotamiento de las correlaciones nos permite establecer las relaciones y observar que

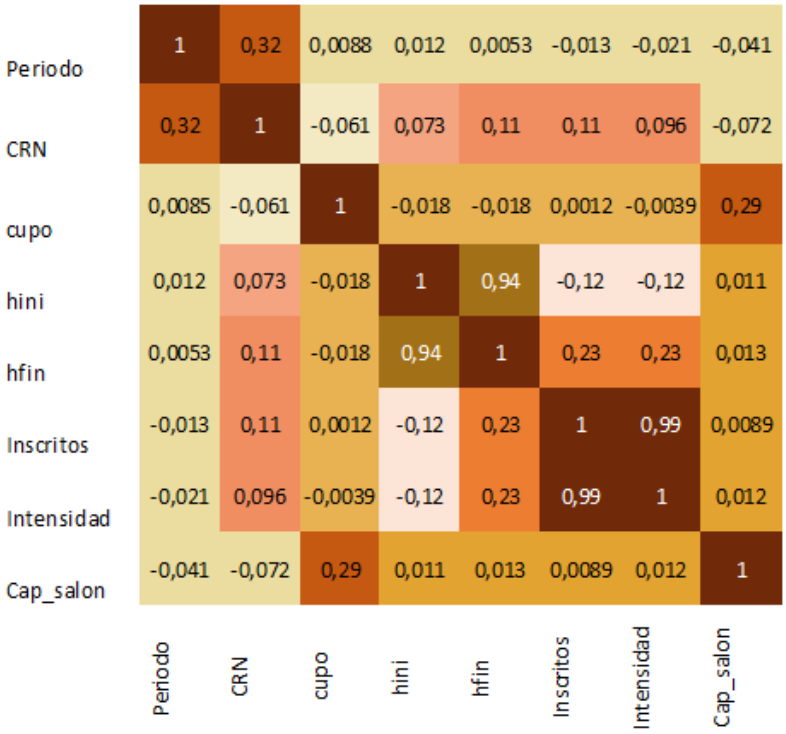


Figura 5 – Matriz de correlación de variables

existe una relación fuerte y directa entre las variables hora de inicio y hora final (0,94). De manera general podría indicarse que existe una baja correlación entre las demás variables. Se observan también relaciones bajas e indirectas entre algunas variables, como por ejemplo entre intensidad y periodo, siendo esta de -0,021, entre Capacidad del salón y periodo (-0,041), entre Capacidad del salón y el código CRN (-0,072).

4.2. Transformación de datos de entrada en variables categóricas

Se realiza un análisis descriptivo para poder perfilar los datos generados por el sistema de optimización de recursos de espacio, teniendo como resultado una transformación de los datos de entrada en variables categóricas obteniendo agrupaciones, como por ejemplo niveles de formación académica y franjas horarias, para reducir los rangos, de tal forma que no interfieran en los resultados finales.

4.3. Random Forest para obtener las variables relevantes

Se utiliza el algoritmo Random Forest (Müller & Guido, 2016), que es un método de clasificación supervisado. Permite obtener la importancia de cada variable. Opera mediante la construcción de árboles de decisión en el proceso de entrenamiento, reduce el overfitting, mejora la precisión del algoritmo de clasificación y reduce el tiempo de entrenamiento. En la Figura 6 se presenta el resultado del algoritmo Extra Trees Classifier con las variables más relevantes para el proceso y su nivel de importancia.

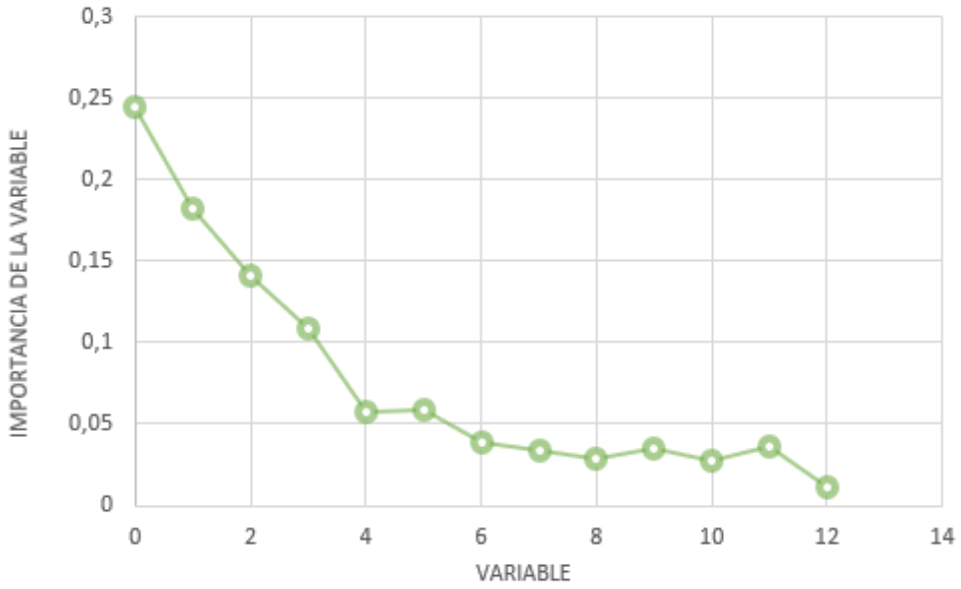


Figura 6 – Resultado del Algoritmo Extra Tree Classifier

Los resultados que se muestran en la Tabla 1 dejan ver un ranking de importancia de cada variable con el fin de poder determinar qué variables juegan un papel importante en

el modelo de clasificación. La de mayor importancia, dada la variable objetivo "cambio", con un valor de 0.2450 es la de facultad, seguida por las variables intensidad, franja horaria, bloque, etc. Los valores asociados a cada una de las variables seleccionadas se obtienen teniendo en cuenta estimadores dentro del algoritmo Extra Trees Classifier, que ajusta un número de árboles de decisión aleatorio en varias sub-muestras del conjunto de datos inicial usando el promedio para mejorar la precisión del modelo y controlar el sobreajuste.

Eje x	Variable	Importancia
0	Facultad	0.2450
1	Intensidad	0.1825
2	Franja horaria	0.1411
3	Bloque	0.1087
4	Tipo mobiliario	0.0577
5	Mobiliario	0.0590
6	Móvil Express	0.0390
7	Viernes	0.0356
8	Miércoles	0.0282
9	Lunes	0.0356
10	Martes	0.0277
11	Jueves	0.0306
12	Sábado	0.0108

Tabla 1 – Importancia de las variables

4.4.Modelo de Árboles de decisión

Mediante el uso del método Cross Validation, que consiste en dividir el conjunto de datos que se disponen para entrenar en dos partes, una que se llama "conjunto de entrenamiento" y otro "conjunto de prueba", se puede evaluar los resultados de un análisis para que así se pueda garantizar que son independientes de la partición entre estos dos conjuntos. Se genera el modelo de árbol de decisión con una división del conjunto de datos del 60% para entrenamiento y un 40% para su validación.

Para obtener mejores resultados se aplica el balanceo de carga con el método up-sampled, con el fin de equilibrar las clases. Proceso de duplicar aleatoriamente las observaciones de la clase minoritaria para reforzar su señal.

La Tabla 2 muestra el reporte de clasificación del árbol de decisión con cada valor que se explica a continuación:

	Precisión	Recall	F1-Score	Support
0 - No	0.91	0.81	0.86	12026
1 - Si	0.83	0.92	0.87	12010
micro avg	0.86	0.86	0.86	24036
macro avg	0.87	0.86	0.86	24036
0 - No	0.91	0.81	0.86	12026
1 - Si	0.83	0.92	0.87	12010

Tabla 2 – Resultado modelo árbol de decisión

En cuanto al desempeño del modelo, y para el interés de la propuesta que está en la identificación de los cambios = Si, se tienen en cuenta los criterios de precisión con un 86% y el recall del 86%. Se refleja que el comportamiento del algoritmo para el caso de la precisión está logrando una clasificación de los casos relevantes mejor que de los irrelevantes y el alto recall, muestra que el algoritmo responde a una alta habilidad de evaluar todos los ejemplos en su correspondiente clase.

De acuerdo con los resultados de la matriz de confusión, que se presentan en la Tabla 3, se observa como 9828 casos están siendo clasificados como verdaderos positivos y los falsos positivos son 1020, lo que logra que el algoritmo de una alta precisión, de igual forma se ve que 2136 casos son falsos negativos que frente a los verdaderos positivos 9828 son pocos haciendo que exista un alto recall.

	Clase Si	Clase No
Clase Si	9828	2316
Clase No	1020	10872

Tabla 3 – Matriz de confusión

De los resultados del árbol de decisión de la Figura 7 se plantean reglas en la solicitud de un salón que son determinantes a la hora realizar un cambio del espacio físico y reflejan los siguientes parámetros: en ninguna regla se contempla el servicio móvil de recursos tecnológicos; la intensidad horaria en rangos de minutos varía entre 270 a 520 minutos, solo una regla incluye un rango mayor a 520 minutos; las facultades que más solicitan cambios son Educación, Sociales, Arquitectura y Economía teniendo en cuenta el tipo de curso, especialización o maestría; el tipo de mobiliario influye a la hora de solicitar cambio si el salón cuenta con sillas móviles y/o mesas fijas; así como el día en que se dicta un determinado curso.

Con las reglas de decisión generadas por el modelo se logra validar el propósito de cada indicador estadístico diseñado, con el fin de disminuir las solicitudes de cambios de espacio físicos generando nuevas estrategias en el proceso de Asignación de salones.

```
SERXPRESS = NA: NO {NO = 15100, SI = 687}
SERXPRESS = NO
|       Intensidad = range1 [-∞ -270] :  NO {NO = 5198, SI = 986}
|       Intensidad = range2 [270 - 520]
|       |       FACULTAD = CIDER
|       |       |       Tipo_curso = Especializaciones: SI {NO = 0, si = 5}
|       |       |       Tipo-curso = Maestría
|       |       |       |       VIERNES = NA
|       |       |       |       |       TIPO_MOBILIARIO = Pupitre universitario: NO {NO = 15, SI = 15}
|       |       |       |       |       TIPO_MOBILIARIO = Silla móvil y mesa fija
|       |       |       |       |       |       JUEVES = J: NO {NO = 5, SI = 0}
|       |       |       |       |       |       JUEVES — NA
|       |       |       |       |       |       |       BLOQUE = B: NO {NO = 5, SI = 0}
|       |       |       |       |       |       |       BLOQUE = O: SI {NO = 0, SI = 5}
|       |       |       |       |       TIPO_MOBILIARION -- Silla móvil, mesa móvil: NO {NO -- 11, SI -- 0}
|       |       |       |       VIERNES = V: NO {NO = 35, SI = 0}
|       |       FACULTAD = DERECHO: NO {NO = 12, SI = 0}
|       |       FACULTAD = EDUCACION
|       |       |       BLOQUE = AU: SI {NO = 0, SI = 8}
|       |       |       BLOQUE = G: SI {NO = 0, SI = 8}
|       |       |       BLOQUE = PU: NO {NO = 2, SI = 0}
|       |       |       BLOQUE = SI : SI {NO = 0, SI = 6}
|       |       FACULTAD = SOCIALES: SI {NO = 0, SI = 6}
|       Intensidad = range3 [520 - ∞]
|       |       SABADO = NA: NO {NO = 18, SI = 0}
|       |       SABADO = S
|       |       |       FACULTAD = ARQUITECTURA Y DISEÑO: SI {NO = 1, SI = 2}
|       |       |       FACULTAD = DERECHO:  NO {NO = 4, SI = 0}
|       |       |       FACULTAD = EDUCACION: NO {NO = 6, SI = 0}
SERXPRESS = SI: NO {NO = 9633, SI = 908}
```

Figura 7 – Resultado Árbol de Decisión

5. Conclusiones

La construcción de la solución presentada en este proyecto le permite al experto del negocio tener una visión global de los datos con los que trabaja día a día, y con el apoyo de los modelos construidos, identificar las principales variables relevantes.

Esta investigación presenta los modelos de aprendizaje automático basados en conjuntos de árboles que pueden mostrar un mejor rendimiento en la clasificación, para lograr resultados positivos en cuanto a la analítica de datos utilizando modelos de aprendizaje supervisado. Es importante realizar un entendimiento del negocio junto a un experto y elaborar un preprocesamiento de los datos, que garanticen una calidad y coherencia para el entrenamiento exitoso del modelo elegido.

El trabajo realizado por el área encargada de asignación de espacios para un periodo académico contempla la programación de horarios que es una de las características del problema general de timetabling y de los cuales se han realizado investigaciones profundas debido a que es considerado un tema complejo. Es posible que, a futuro, con el objetivo de poder encontrar otras alternativas de modelos, se puedan realizar pruebas con algoritmos de Machine Learning. Con el hallazgo de variables significativas en el proceso de solicitudes de cambio y la definición de indicadores de gestión se podría contemplar un estudio de factibilidad de creación de nuevos parámetros que puedan incluirse en un sistema de reserva de espacios y recursos físicos.

Referencias

Adams, L. G., Khare, S., Lawhon, S. D., Rossetti, C. A., Lewin, H. A., Lipton, M. S., … Drake, K. L. (2011). Enhancing the role of veterinary vaccines reducing zoonotic diseases of humans: Linking systems biology with vaccine development. *Vaccine*, *29*(41), 7197–7206. https://doi.org/10.1016/j.vaccine.2011.05.080

Almeida, M. W. S., Medeiros, J. P. S., & Oliveira, P. R. (2016). Solving the academic timetable problem thinking on student needs. In *Proceedings - 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015* (pp. 673–676). https://doi.org/10.1109/ICMLA.2015.184

Alsuwaylimi, A. A., & Fieldsend, J. E. (2019). A New Initialisation Method for Examination Timetabling Heuristics. In *2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019* (pp. 1636–1643). https://doi.org/10.1109/SSCI44817.2019.9002989

Calle-López, D., Cornejo-Reyes, J., Pesántez-Avilés, F., Rodas-Tobar, M., Vásquez-Vásquez, C., & Robles-Bykbaev, V. (2018). Un sistema experto basado en minería de datos y programación entera lineal para soporte en la asignación de materias y diseño de horarios en educación superior. *Enfoque UTE*, *9*(1), 102–117. https://doi.org/10.29019/enfoqueute.v9n1.226

Calle Lopez, D. E., Cornejo Reyes, P. J., Pesantez Aviles, L. F., Rodas-Tobar, M., Vasquez, C. E., & Robles Bykbaev, V. E. (2018). Un sistema experto basado en minería de datos y programación entera lineal para soporte en la asignación de materias y diseño de horarios en educación superior.

Cuadrado Guevara, M. D. (2014). *Un modelo matemático para el problema de horarios del Departamento de matemáticas de la Facultad de Ciencias y Tecnología de la Universidad de Carabobo*. Universidad de Carabobo. https://doi.org/10.13140/RG.2.1.4880.6001

Díaz Cortes, W., García, C., & Hernandez, J. (2015). Herramienta basada en Técnicas de Inteligencia Artificial, para la Asignación de Horarios y Recursos Académicos, en el Proyecto Curricular de Ingeniería de Sistemas de la Universidad Distrital Francisco José De Caldas.

Durán Vaca, M., Gómez Moreno, Y., & Ramírez Aguirre, J. (2019). *Indicadores de solicitudes de cambio en el proceso de asignación de salones, para medir la efectividad del proceso en el área de Admisiones y Registro de la Universidad de los Andes*. (U. de los Andes, Ed.). Bogotá.

Even, S., Itai, A., & Shamir, A. (1975). On the complexity of timetable and multi-commodity flow problems. *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, *1975-Octob*, 184–193. https://doi.org/10.1109/SFCS.1975.21

Ferland, J., & Roy, S. (2018). Timetabling problem for university as assignment of activities to resources. *Computers & Operations Research*, *12*(2), 207–218. https://doi.org/https://doi.org/10.1016/0305-0548(85)90045-0

Harvill, J. L. (2012). *An introduction to mathematical statistics and its applications*.

Müller, A., & Guido, S. (2016). *Introduction to Machine Learning with Python*. (O. Media, Ed.). United States of America.

Paechter, B., Gambardella, L., & Rossi-Doria, O. (2002). Metaheuristic Network.

Patil, S., & Kulkarni, U. (2019). Accuracy prediction for distributed decision tree using machine learning approach. In *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019* (pp. 1365–1371). Tirunelveli, India: IEEE. https://doi.org/10.1109/icoei.2019.8862580

Perumal, S., Tabassum, M., Norwawi, N., Narayana, G., & Perumal, S. (2018). Development of an Efficient Timetable Sustem using AngularJS and Bootstrap 3. In *IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2018)* (pp. 3–8). Malaysia: IEEE.

Pisane, J., Marée, R., Wehenkel, L., & Verly, J. (2010). EXTRA-TREES. In *2010 IEEE Radar Conference* (pp. 1454–1458). Washington D.C.

Vergel Miranda, A. (2017). *Algoritmo para la asignación de horarios académicos en la Universidad Francisco de Paula Santander Ocaña utilizando técnicas de Inteligencia Artificial. Tesis*. Retrieved from http://repositorio.ufpso.edu.co:8080/dspaceufpso/bitstream/123456789/2290/1/32100.pdf

Wang, R., Zhou, M., Li, Y., Zhang, Q., & Dong, H. (2019). A Timetable Rescheduling Approach for Railway based on Monte Carlo Tree Search. In *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019* (pp. 3738–3743). IEEE. https://doi.org/10.1109/ITSC.2019.8916963

Witten, I., Frank, E., & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. United States of America: British Library.

Yang, X. F., Ayob, M., & Nazri, M. Z. A. (2018). An investigation of timetable satisfaction factors for a practical university course timetabling problem. In *Proceedings of the 2017 6th International Conference on Electrical Engineering and Informatics: Sustainable Society Through Digital Innovation, ICEEI 2017* (Vol. 2017-Novem, pp. 1–5). https://doi.org/10.1109/ICEEI.2017.8312409

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 01/08/2020
Aceitação/Acceptance: 09/10/2020

Anna: Modelo para el análisis de requisitos de software y la identificación automática de variables, valores y generación de casos de pruebas funcionales de software

Dilan Mejía[1], Hugo Ordóñez[2], César Pardo[2]

dmejia@greensqa.com; hugoordonez@unicauca.edu.co; cpardo@unicauca.edu.co

[1] Facultad de ingeniería de sistemas, Universidad de San Buenaventura, la umbría pance, 760032, Cali, Colombia.

[2] Departamento de Sistemas, Facultad de Ingeniería Electrónica y Telecomunicaciones, Universidad del Cauca, Calle 5 N.º 4-70, 190003, Popayán, Cauca, Colombia.

Pages: 297-309

Resumen: Una de las actividades determinantes en la fase de análisis del proceso de pruebas funcionales, es la identificación de variables y valores a partir de la especificación de requisitos. Este proceso demanda tiempo, esfuerzo e impacta en el diseño de los casos de pruebas y su ejecución sino se realiza correctamente. Este trabajo presenta un modelo que apoya la identificación de variables y valores a partir del análisis de la especificación de requisitos (no estructurados) como insumo para la generación de casos de pruebas. El modelo utiliza algoritmos de procesamiento de lenguaje natural y tiene por objetivo, reducir el tiempo y esfuerzo en el análisis de requisitos. El modelo fue evaluado en la empresa GreenSQA por 38 profesionales del aseguramiento de calidad de software, el 95% de los participantes consideraron pertinentes las variables y valores identificadas por el modelo. Los resultados obtenidos permiten observar que el modelo ahorra un 90% de tiempo y esfuerzo frente a técnicas manuales utilizadas en la identificación de variables para el diseño de casos de pruebas.

Palabras-clave: Caso de prueba; Procesamiento de Lenguaje Natural (PLN); Variables; Valores de variables.

Anna: Model for requirements analysis and automatic identification of variables and values for the generation of functional test cases

Abstract: One of the determining activities in the analysis phase of the functional testing process is the identification of variables and values based on the specification of requirements. This process demands time, effort and impacts in the design of the test cases and their execution if not carried out correctly. This work presents a model that supports the identification of variables and values from the analysis of the specification of requirements (unstructured) as input for the generation of test cases. The model uses natural language processing algorithms and aims

to reduce time and effort in analyzing requirements. The model was accepted in the GreenSQA company by 38 software quality assurance professionals, 95% of the participants considered that variables and values identified by the model are relevant. Frequently selected results for test case design save 90% time and effort for the model compared to manual techniques used in identifying variables for test case design.

Keywords: Test case; Natural Language Processing (NLP); Variables; Variable values.

1. Introducción

Una característica importante para que el software sea desplegado y aceptado en el mercado, es que éste pase las pruebas de software necesarias. En ese sentido, para probar un sistema software, se generan casos de pruebas que se derivarán de todas las posibles combinaciones de variables con sus respectivos valores. Sin embargo, esto se convierte en una actividad ineficiente cuando se realiza de manera exhaustiva, principalmente, debido al tiempo y esfuerzo invertidos (Kuhn, Wallace, and Gallo 2004). Por lo anterior, las variables y sus valores son sometidas a diferentes técnicas de diseño de caja negra, que permiten aumentar la cobertura ocupando menos combinaciones de casos de pruebas (Cohen, D M Dalal, S R Fredman, M L Patton 1997).

Identificar las variables en los documentos de especificación de requerimientos es una tarea difícil, dado que el profesional quien realiza esta actividad, necesita tener un amplio conocimiento en el dominio del negocio para poder extraer de manera asertiva cada una de éstas. Además, su identificación depende del contexto que se trata en el requisito y su correcta interpretación, ya que en la mayoría de los casos las variables no se encuentran en el texto de especificación de requisitos (Lu, XiaoNan Jimbo 2019),(Kuhn, R Kacker, R Lei, Y Hunter 2009), (Zhu, Hanbing Zhang, Riquan Yu, Zhou Lian, Heng Liu 2019). Por otra parte, otro factor que afecta la identificación de las variables es el tiempo que se emplea para analizar la información, ya que en ocasiones la cantidad de requisitos es alta y esto demanda varias horas, incluso días, y generalmente, en la estimación y planeación no se contempla el esfuerzo requerido por los profesionales de pruebas para esta actividad.

Una alternativa para abordar el problema anterior consiste en aplicar técnicas de procesamiento de lenguaje natural (PLN) al texto de especificación de requisitos, esto, con el fin de identificar los comportamientos de un sistema de información (software), y así poder generar casos de pruebas automáticamente. Con el propósito de facilitar la identificación y extracción de los comportamientos que el software debería realizar, algunos autores como: (Garousi, Vahid Bauer, Sara Felderer 2018) y (Melo, Silvana M Carver and Souza, Paulo S L Souza 2019), han planteado soluciones basadas en la estructuración del texto de manera exhaustiva. Sin embargo, el costo en términos del tiempo y esfuerzo empleado en su aplicación es alto, esto debido a la gran cantidad de documentos de dominio específico que deben ser tratados a través del proceso de estructuración del texto.

Teniendo en cuenta lo anterior, y con el objetivo de disminuir el tiempo y esfuerzo empleado en la identificación de variables y valores para la generación de casos de

pruebas funcionales, en este artículo se presenta un modelo llamado Anna. Que aplica procesamiento de lenguaje natural para la identificación de variables y valores a partir del análisis de textos (documentos) de especificación de requisitos como alternativa para la generación automática de casos de pruebas funcionales. Así mismo, y como parte de las funcionalidades desarrolladas, se presenta un componente de desambiguación lingüística de variables y valores de variables, que permite desambiguar variables que pudieran tener la misma palabra, pero con distintos significados. Este componente se fundamentó en diferentes ecuaciones matemáticas, las cuales también son presentadas en este documento. Para comprobar la validez de las variables y valores identificados generados por Anna, se llevó a cabo una evaluación en la empresa GreenSQA como piloto de pruebas. En este piloto se utilizó la información de un proyecto real conformado por 38 ingenieros profesionales y expertos en aseguramiento de calidad de software. Los resultados obtenidos también se presentan en este documento.

El resto del documento está estructurado de la siguiente manera. La Sección 2 presenta un resumen de los trabajos relacionados. La Sección 3 presenta los componentes del modelo propuesto que permiten llevar a cabo la desambiguación de las variables polisémicas e identificación de las variables y valores. En la Sección 4, se describe la evaluación y resultado obtenidos en la empresa Green SQA. La Sección 5 presenta las conclusiones y trabajos futuros.

2. Trabajos relacionados

La especificación del requisito es una de las principales actividades que se realiza para la creación de casos de pruebas (Xu, S Chen, L Wang, C Rud 2016), dado que la definición de casos de pruebas se basa en la generación de un documento que contiene todos los requisitos que se deben cubrir. De acuerdo con el trabajo de (Garousi, Vahid Bauer, Sara Felderer 2018), donde se realiza una búsqueda sistemática enfocada en la utilización del procesamiento de lenguaje natural en las actividades de pruebas de software, presenta un listado de herramientas software capaces de generar casos de pruebas con un 77% de aceptación por parte de los ingenieros evaluadores a partir de los documentos de requisitos no estructurados que éstas reciben.

El trabajo de (Carvalho, Gustavo Barros, Flávia Lapschies, Florian Schulze, Uwe Peleska 2014), propone que a través de procesamiento de lenguaje natural controlado y aplicado a los requerimientos, se evita la ambigüedad que aparece cuando se redacta el requerimiento luego, se identifica cada una de las clases sintácticas (POS Taggin) para poder encontrar los árboles sintácticos en el texto que le ayudan a la herramienta a generar los casos de pruebas. Por otra parte, aclara que hay técnicas más estrictas que controlan el lenguaje natural, por ejemplo: las restricciones gramaticales que se presenta en trabajos como los propuestos en (Zhang, Man Yue, Tao Ali, Shaukat Zhang, Huihui Wu 2014), (Rane 2017) y (Ansari, Ahlam Shagufta, Mirza Baig Sadaf Fatima, Ansari Tehreem 2017), en donde el requerimiento debe venir con unas palabras claves que ayudan a la herramienta a identificar cada uno de los casos de prueba.

En (Kikuma, Kazuhiro, Yamada Takeshi Sato, Koki, Ueda 2019) se presenta una solución que descompone un caso de pruebas en partes, esto con el fin de buscar las variables e identificarlas en un documento o texto de especificación de requerimientos,

sin ningún tipo de restricciones. Este sistema realiza la extracción de información de texto relacionado con datos (como texto de tablas u objetos de power point), para luego aplicar técnicas de valores límites. Además, cuenta con un diccionario de casos de pruebas el cual almacena cada una de las palabras y patrones de comportamiento, lo que le permite extraerlos y formar el caso de prueba a partir de los datos encontrados y los datos contenidos en el diccionario, esté trabajo está formado por un clasificador bayesiano ingenuo (Naïve Bayes).

En (Dwarakanath and Sengupta 2012) se presenta una propuesta que no restringe el documento de requerimientos, y a partir de un analizador de enlaces gramaticales, identifica en el texto la estructura sintáctica entre pares de palabras, los cuales son almacenados en un diccionario para luego realizar un análisis en los corpus lingüísticos (TreeBank) para generar casos de pruebas. Esta herramienta es viable en documentos de requerimientos que se encuentren en el dominio de estructura sintáctica del software representada en un diccionario.

En (Landhäußer and Genaid 2012), se define una ontología que guarda instancias de código fuente a partir de breves descripciones de funcionalidades registradas en historias de usuario, las cuales se trasforman en funciones del sistema a través de código fuente. El código generado en las funciones es enlazado a la ontología para crear casos de pruebas funcionales por medio de análisis de código fuente usando la fórmula de frecuencia de términos por documento (tf/idf), mediante la cual se encuentran procesos similares y código fuente reutilizado en los sistemas a probar.

A partir del análisis anterior, este trabajo se centra en la identificación de variables y valores de variables en texto no estructurado. A diferencia de los trabajos anteriormente analizados, el modelo propuesto hace uso de procesamiento de lenguaje natural para identificar variables y valores a través de la identificación de adjetivos y sustantivos. Además, cuando una variable es representada por una palabra que puede tener varios significados (polisemia), el modelo propuesto utiliza un componente de desambiguación que identifica el contexto de la palabra a través de n-gramas. Por otra parte, como insumo para la generación de casos de pruebas de forma automática, el modelo propuesto, se basa en el conocimiento almacenado en la herramienta K-Tree, la cual contiene una amplia base de conocimiento basada en ontología de dominio general alimentada por los proyectos que se desarrollan por la empresa GreenSQA.

3. Modelo propuesto

El modelo propuesto está conformado por seis componentes principales, éstos son: (i) extracción de información de la herramienta K-Tree; (ii) indexación de variables y valores; (iii) normalización del texto; (iv) identificación de variables; (v) desambiguación de las variables polisémicas y (vi) envío e indexación del nuevo conocimiento. La Figura 1 presenta la arquitectura del modelo propuesto. Cada uno de estos componentes se describe a continuación.

1-Extracción de información del K-Tree: El K-Tree es una herramienta desarrollada por GreenSQA que contiene una ontología de dominio para diferentes sectores como Salud, Financiero, Educación, Servicios Públicos, entre otros. K-Tree es alimentada por cada

uno de los proyectos de pruebas de software que se llevan a cabo en GreenSQA. A partir de esta información contenida en K-Tree, el modelo propuesto extrae las variables y valores de las variables y las compara con las variables y valores detectadas en el texto (documento) de especificación de requerimientos analizado.

2-Indexación de variables y valores: Para cada una de las variables y valores extraídos del K-Tree, se crea una estructura de datos de tipo documento, los cuales son objetos en memoria que están compuestos por un título y un cuerpo. El nombre de la variable se guarda en el documento como el título y los valores extraídos de la variable se guardan en el cuerpo generando un documento. Estos documentos, son indexados usando la estructura de datos de índices invertidos (Białecki, Muri, and Ingersoll 2012), los cuales se utilizan con frecuencia en el proceso de recuperación de información cuando se tiene grandes volúmenes de datos a los que se desea acceder.

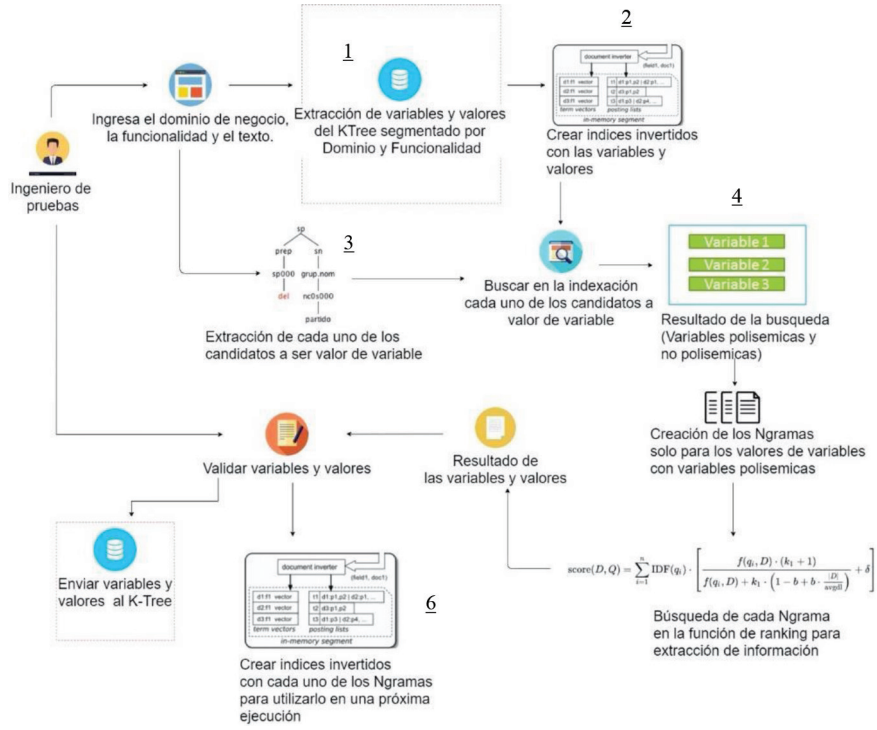


Figura 1 – Arquitectura del modelo propuesto

3-Normalización del texto: Una vez indexadas las variables y valores tomadas del K-Tree, el modelo propuesto procede a identificar los valores de las variables candidatos del texto de requerimientos que se quiere analizar. La estructura lingüística de un valor de la variable en el español se representa por un sustantivo o un sustantivo + adjetivo. La estrategia planteada para poder identificar una variable en el texto es encontrar los

posibles candidatos (palabras que pueden referenciar un valor) a ser valor de variables asociado a una variable. En este proceso se realiza la etiquetación gramatical de cada una de las palabras del texto de requerimientos y se extraen los sustantivos y adjetivos. El proceso de etiquetación gramatical utiliza redes de dependencias entre palabras y oraciones a través de una serie de ubicaciones log-lineales condicionales, conocido como entropía máxima o multiclase, la cual permite inferir de manera bidireccional en el texto (Toutanova, Kristina Klein and Manning, Christopher D Singer 2003), esto es posible a través de la función objetivo representada en la ecuación (1). Donde *Pa* (x_i) son los nodos con arcos al nodo x_i, es decir, un nodo es representado por un sustantivo, y los arcos por los adjetivos conectados a él. El puntaje de cada nodo será la probabilidad conjunta del nodo *P (x)*. Para identificar sustantivos y adjetivos, El modelo propuesto, fue entrenado utilizando dos corpus del español: el AnCora Spanish 3.0 el cual contiene 17.000 oraciones (disponible en https://bit.ly/36lmpQk) y el DEFT Spanish Treebank V2 (LDC2015E66) que contiene 5.000 oraciones (disponible en https://bit.ly/2Tr6drv).

$$score(x) = \prod_i P(x_i, Pa(x_i)) \qquad \text{Ecuación (1)}$$

4-Identificación de variables: Para cada uno de los candidatos identificados en la actividad anterior, se procede a buscarlo en la indexación realizada, como resultado se obtienen valores de variables que se esperaba estuvieran asociados a una única variable, como esto no siempre es así, puede suceder que un mismo valor esté vinculado a dos o más variables, a esto se le conoce como *variable polisémica* y para poder resolver esta ambigüedad, se utiliza el proceso de *desambiguación lingüística*.

5-Desambiguación de las variables polisémicas: Para poder realizar la desambiguación lingüística de las variables polisémicas, se crean *n-gramas* acumulativos hasta tamaño siete (Sidorov, 2013). Se inicia obteniendo cada uno de los sustantivos del texto y se pasa a formar un bigrama con la palabra que está a su lado derecho (la cual representa el posible valor) como se muestra en la Figura 2. Luego se forma otro bigrama utilizando el mismo sustantivo y la palabra (posible valor de variable) que está a su lado izquierdo como se muestra en la Figura 3.

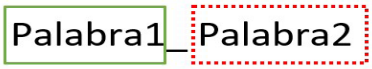


Figura 2 – Bigrama con la palabra de la derecha



Figura 3 – Bigrama con la palabra de la izquierda

A continuación, se forma un trigrama utilizando el mismo sustantivo y la palabra que está a su lado izquierdo y derecho como se muestra en la Figura 4. Este proceso se

repite con el mismo sustantivo hasta formar un N-Grama siete como se muestra en la Figura 5.



Figura 4 – Trigrama con el sustantivo y las palabras de los lados

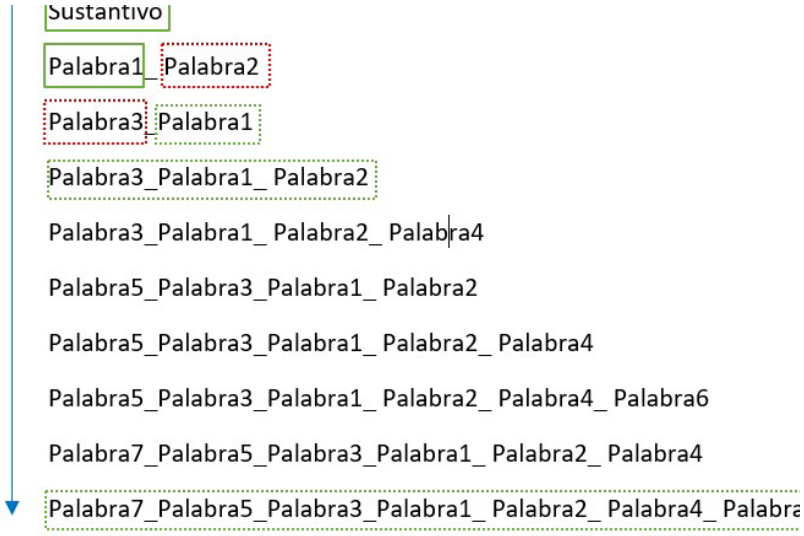


Figura 5 – N-Grama 7

Para los sustantivos que se encuentren en el texto y, que sea una posible variable polisémica, se selecciona desde el n-grama de tamaño 7 y se empieza a buscar utilizando una *función de ranking de recuperación de información* llamada *Okapi BM25* (Robertson, S E Walker, S Jones, K Sparck Hancock Beaulieu 1994) (representado en la ecuación 2) en los documentos indexados. Si el n-grama de tamaño 7 no arroja resultados, se irá buscando el siguiente n-grama de menor orden hasta llegar al n-grama de tamaño 1. El proceso de desambiguación inicia buscando del n-grama de mayor tamaño hasta el más pequeño, ya que se espera que el valor de la variable esté relacionado con el contexto de la variable. A mayor sea el n-grama, mayor es la semántica, y a menor sea el n-grama, menor será sintácticamente.

$$Score(D,Q) = \sum_{i=1}^{\{n\}} IDF(q_i) \cdot \left[\frac{f(q_i,D)\cdot(k_1+1)}{f(q_i,D)+k_1\cdot\left(1-b+b\cdot\frac{|D|}{avgdl}\right)} + \delta \right] \text{ Ecuación (2)}$$

A continuación, en la Tabla 1 se presenta cada uno de los componentes de la función Okapi BM25.

Id	Descripción
Q	Corresponde a la consulta o documentos dado para Anna.
D	Es el documento sujeto a la búsqueda. En una consulta se puede tener en cuenta más de un documento. Este documento se define en la actividad de indexación.
qi	*n-grama* con el que se realiza una búsqueda. En una búsqueda pueden incluirse más de un *n-grama*.
f (qi,D)	Indica cuantas veces aparece un n-grama *qi* en el los documento sujeto de búsqueda.
avgdl	presenta el valor de la longitud media de los documentos sujetos de búsqueda.
k1	Parámetro estándar con valores establecidos en conferencias TREC. Se recomienda que el valor 1.2.
b	Parámetro estándar con valores establecidos en conferencias TREC. Se recomienda que estos valores sean 0.75.
IDF	Inverse Document Frequency es el peso de los *n-gramas qi* que aparecen en la consulta *Q* (Jones, K. S. 1972) . *IDF* se calcula con ecuación 3.

Tabla 1 – Componentes de la función Okapi BM25

$$IDF(q_i) = log\ \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}$$ Ecuación (3)

Donde *N* es el número total de documentos, y *n*(*qi*) es el número de documentos que contienen el *n-grama* q_i.

Como resultado el modelo propuesto entrega un ranking con los n-gramas que mayormente se acerquen a cada variable y su valor en un determinado contexto. En este sentido, el ranking está ordenado de mayor a menor con los n-gramas candidatos a ser variables, para ser avaladas por un ingeniero de pruebas.

6- Envío e indexación del nuevo conocimiento: Una vez obtenidas las variables con sus valores, se presentan al usuario (profesional de pruebas) quien decidirá cuales son pertinentes para ser enviadas al K-Tree y así alimentar su base de conocimiento con nueva información. Para cada una de las variables con sus valores que fueran avaladas, se toman los n-gramas formados y se indexan, esto se realiza con el propósito de almacenar el contexto en el que se encontró el valor de la variable. La indexación es guardada bajo la estructura de carpetas de negocio y funcionalidad, para tener segmentada la información indexada.

Cuando se ha ejecutado previamente el modelo propuesto sobre un determinado dominio, las variables y valores identificadas que han sido avaladas por el usuario (Ingeniero de pruebas), se almacenan en la base de conocimiento (K-Tree). Una vez estas variables son almacenadas quedan disponibles para ser consultadas en futuras ejecuciones del modelo propuesto.

4. Evaluación y resultados

La evaluación de Anna se llevó acabo de dos maneras, las cuales son explicadas a continuación.

4.1. Evaluación por medio de la herramienta K-Tree

Anna utilizando el conocimiento del negocio "Banca" y la funcionalidad "Evaluación" de la herramienta K-Tree**,** luego se seleccionó un fragmento de texto de un requerimiento no ambiguo que fue enviado a Anna para que identifique las posibles variables y valores. Los resultados obtenidos se compararon frente al análisis de esta actividad que realizaron 5 ingenieros expertos del aseguramiento de calidad. En este proceso los expertos tenían identificadas 5 variables, Anna consiguió identificar cuatro de las cinco variables que se esperaban identificar, además de sus respectivos valores. La variable cinco no se identificó debido a que no estaba registrada en el KTree, por lo que se prosiguió a realizar el envío de ésta (quinta) variable al K-Tree a través de Anna y así indexar la variable con su valor. Una vez que se indexa la variable y se ha ejecutado Anna con la misma configuración y requerimiento, se obtienen todas las variables, es decir, el 100% de todas las variables identificadas por los expertos. Cuando se evalúan variables de dominios no registrados en el K-Tree, Anna identifica las posibles variables y valores. Éstas son presentadas a los expertos, quienes deciden con base en su experiencia si es pertinente o no cada una de las variables identificadas, si la variable y su valor es pertinente, ésta se indexa en el K-Tree.

4.2. Evaluación por medio de un cliente de GreenSQA

Para la segunda evaluación, se configuró Anna bajo el conocimiento de negocio "Banca" y la funcionalidad "análisis de riesgo", esto se realizó en el escenario de un cliente de GreenSQA, pero con la restricción de que el requerimiento del usuario contiene texto ambiguo. Anna consiguió identificar el 77% del total de las variables y valores de estas variables en 10 segundos. Esta actividad, se realizó con 38 diferentes ingenieros del aseguramiento de calidad de software, los cuales cuentan varios años de experiencia aplicada en diferentes dominios de negocios. En esta evaluación, se agruparon los ingenieros por el tiempo que demoraron en realizar la actividad de identificación de

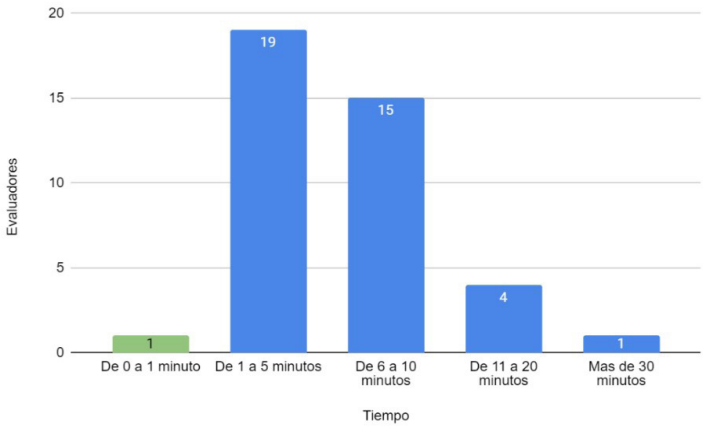


Figura 6 – Tiempo empleado en la identificación de variables y valores por los grupos y Anna

variables con sus valores, como resultado se obtuvieron 4 grupos. El grupo 1 estuvo conformado por 19 ingenieros que tardaron alrededor de 5 minutos, el grupo 2 estuvo conformado por 14 ingenieros que tardaron alrededor de 10 minutos, el grupo 3 lo conformó 4 ingenieros que tardaron hasta 20 minutos, finalmente, el grupo 4 lo conformó 1 ingeniero que tardó 30 minutos. En la Figura 6 se puede observar que Anna es 30 veces más rápido que el mejor grupo evaluador, es decir, el grupo 1. Permitiendo demostrar, el considerable ahorro de tiempo y esfuerzo cuando se utiliza Anna en la actividad de identificación de variables y valores de estas variables.

Seguido, se encontraron dos casos donde el modelo propuesto consiguió identificar variables, no detectadas por los ingenieros, las variables fueron "declara renta" y "cumple políticas". Además, se aplicó una encuesta, la cual estuvo conformada por 3 preguntas dicotómicas, es decir, de dos respuestas, SI o NO. La Tabla 2 presenta las preguntas realizadas y el número de personas asociadas a las respuestas. En general, se puede observar que en la Tabla 2, para la pregunta 1, el 23.7% de los ingenieros no consiguió identificar la variable *Tipo de vinculación*, en la segunda pregunta el 10.5% de los participantes no consiguió identificar la variable "*Puntaje acierta"*, en la tercera pregunta, el 28.9% de los ingenieros de la prueba no asociaron la variable "*Puntaje acierta"* con el valor >=1. Por último, una vez finalizada la actividad, para cada uno de los ingenieros se les realizó una última y cuarta pregunta donde se pedía calificar entre cero a diez, la pertinencia de la propuesta, en general, se pudo observar que el 68.3% de los participantes calificaron de 7 a 10 puntos, 26.4% dio una calificación de 5 a 6 puntos y el 5.3% dieron una calificación de 3 puntos. Esto quiere decir que Anna es relevante para apoyar la identificación de variables con sus valores en la actividad de análisis del proceso de pruebas funcionales de GreenSQA, los resultados están disponibles en: https://bit.ly/2WQbRFB).

5. Conclusiones y trabajos futuros

El modelo propuesto presentado en este documento permite encontrar variables y sus valores para diseñar casos de pruebas funcionales de software. En el proceso de evaluación del modelo propuesto en la empresa GreenSQA, se pudo identificar que las variables y valores encontrados son pertinentes, permitiendo tener un ahorro en tiempo y esfuerzo en la fase del análisis aproximadamente del 90%, esto se logra siempre y cuando la base de conocimiento del K-Tree esté bien alimentada o ya se tengan indexaciones del requerimiento a evaluar. A medida que los ingenieros sigan alimentando la indexación del modelo propuesto, van a aumentar los hallazgos en las variables y valores que éste recomienda. El modelo propuesto, fue diseñado con el fin de disminuir los tiempos y el esfuerzo en la etapa de análisis. El modelo propuesto permitió agilizar al proceso de diseño de casos de pruebas funcionales.

Id	Descripción	SI	NO
1	¿Usted identificó está variable? (Tipo de vinculación)?	29	9
2	¿Usted identificó esta variable? (Puntaje Acierta)?	34	4

Tabla 2 – Preguntas realizadas en la encuesta

3 ¿Usted identificó esta variable con este valor? (Puntaje acierta: 27 11 >=1)?

	Pertinencia	Calificación									
		1	2	3	4	5	6	7	8	9	10
4	Dados los resultados de las variables y valores de variables que Anna encontró	0	0	2	0	8	2	9	5	4	8

Para prevenir amenazas en la evaluación, la empresa GreenSqa, siempre cuenta con un arquitecto del conocimiento que se encarga de revisar las líneas base de conocimiento que contiene las variables y valores, que serán almacenadas en el K-Tree.

A partir de los resultados obtenidos durante la evaluación del modelo propuesto, se identificó una oportunidad de mejora, la cual está relacionado con el poder reconocer sinónimos de las variables en la base de conocimiento del K-Tree, esto, con el propósito de cubrir un mayor número de dominios de negocio.

Como trabajo a futuro, se espera llevar a cabo la implementación del modelo propuesto en diferentes servidores, como, por ejemplo, en cada uno de los servidores de los clientes de la empresa GreenSQA, esto, con el objetivo de que el modelo propuesto se pueda utilizar en múltiples fuentes de indexaciones para que la base de conocimiento siga creciendo. Asimismo, se pretende implementar la capacidad de recibir requerimientos por medio de voz y procesarlos para convertirlos en texto, y posteriormente, proceder a la identificación de variables y valores, esto debido a que muchos de los requerimientos se obtienen por medio de reuniones, lo que permitiría mejorar aún más los tiempos relacionados a la elicitación y análisis de requisitos. Además, se tiene pensado liberar el K-Tree en formato de tripletas RDF, y el código fuente de Anna, para que pueda ser utilizado en cualquier organización que realice pruebas de software.

6. Agradecimientos

Los autores de este artículo agradecen a la empresa GreenSQA por patrocinar este proyecto. Asimismo, los profesores Hugo Ordoñez y César Pardo agradecen la contribución de la Universidad del Cauca, donde laboran como profesores asociados.

Referencias

Ansari, M. B. Shagufta, A. Sadaf Fatima and S. Tehreem, “Constructing Test cases using Natural Language Processing,” 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2017, pp. 95-99, doi: 10.1109/AEEICB.2017.7972390.

Białecki, Andrzej, Robert Muri, and Grant Ingersoll. 2012. “Apache Lucene 4.” *Proceedings of the SIGIR 2012 Workshop on Open Source Information Retrieval* 17–24.

Carvalho, G., Barros, F., Lapschies, F., Schulze, U., & Peleska, J. (2014). Model-Based Testing from Controlled Natural Language Requirements. In C. Artho & P. C. Ölveczky (Eds.), *Communications in Computer and Information Science: Vol. 419 CCIS* (pp. 19–35). Springer International Publishing. https://doi.org/10.1007/978-3-319-05416-2_3.

Cohen, D. M., Dalal, S. R., Fredman, M. L., & Patton, G. C. (1997). The AETG system: an approach to testing based on combinatorial design. *IEEE Transactions on Software Engineering*, *23*(7), 437–444. https://doi.org/10.1109/32.605761

Dwarakanath, A., & Sengupta, S. (2012). Litmus: Generation of Test Cases from Functional Requirements in Natural Language. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 7337 LNCS* (pp. 58–69). https://doi.org/10.1007/978-3-642-31178-9_6

Garousi, Vahid Bauer, Sara Felderer, Michael. 2018. "NLP-Assisted Software Testing: A Systematic Review."

Kikuma, K., Yamada, T., Sato, K., & Ueda, K. (2019). Preparation Method in Automated Test Case Generation using Machine Learning. *Proceedings of the Tenth International Symposium on Information and Communication Technology - SoICT 2019*, 393–398. https://doi.org/10.1145/3368926.3369679

Kuhn, R., Kacker, R., Lei, Y., & Hunter, J. (2009). Combinatorial software testing. *Computer*, *42*(8), 94–96. https://doi.org/10.1109/MC.2009.253

Kuhn, D. R., Wallace, D. R., & Gallo, A. M. (2004). Software fault interactions and implications for software testing. *IEEE Transactions on Software Engineering*, *30*(6), 418–421. https://doi.org/10.1109/TSE.2004.24

Landhausser, M., & Genaid, A. (2012). Connecting User Stories and code for test development. *2012 Third International Workshop on Recommendation Systems for Software Engineering (RSSE)*, 33–37. https://doi.org/10.1109/RSSE.2012.6233406

Lu, X.-N., & Jimbo, M. (2019). Arrays for combinatorial interaction testing: a review on constructive approaches. *Japanese Journal of Statistics and Data Science*, *2*(2), 641–667. https://doi.org/10.1007/s42081-019-00056-w

Melo, S. M., Carver, J. C., Souza, P. S. L., & Souza, S. R. S. (2019). Empirical research on concurrent software testing: A systematic mapping study. *Information and Software Technology*, *105*(August 2018), 226–251. https://doi.org/10.1016/j.infsof.2018.08.017

Rane, Prerana Pradeepkumar. 2017. "Automatic Generation of Test Cases for Agile Using Natural Language Processing." 97.

Robertson, S E Walker, S Jones, K Sparck Hancock Beaulieu, M. M. 1994. "Okapi at TREC-3." *Proceedings of the Third Text REtrieval Conference*.

Sidorov, G. (2013). N-gramas sintácticos no-continuos. *Polibits*, *48*, 69–78. https://doi.org/10.17562/PB-48-9

Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL '03*, *1*, 173–180. https://doi.org/10.3115/1073445.1073478

Xu, S., Chen, L., Wang, C., & Rud, O. (2016). A comparative study on black-box testing with open source applications. *2016 IEEE/ACIS 17th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2016*, 527–532. https://doi.org/10.1109/SNPD.2016.7515953

Zhang, M., Yue, T., Ali, S., Zhang, H., & Wu, J. (2014). A Systematic Approach to Automatically Derive Test Cases from Use Cases Specified in Restricted Natural Languages. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8769, pp. 142–157). https://doi.org/10.1007/978-3-319-11743-0_10

Zhu, H., Zhang, R., Yu, Z., Lian, H., & Liu, Y. (2019). Estimation and testing for partially functional linear errors-in-variables models. *Journal of Multivariate Analysis*, *170*, 296–314. https://doi.org/10.1016/j.jmva.2018.11.005

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 30/09/2020
Aceitação/Acceptance: 13/11/2020

Comparación de funciones kernel para la predicción de la oferta energética fotovoltaica

Héctor Mora-Paz[1], Jaime A. Riascos[1], J.A Salazar-Castro[1], Germán Mora[1], Andrés Pantoja[2], Javier Revelo-Fuelagán[2], Laura Mancera-Valetts[3], Diego Peluffo-Ordoñez[4]

hector.mora@aunar.edu.co; jaime.riascos@aunar.edu.co; jose.salazar@aunar.edu.co; german.mora@aunar.edu.co; ad_pantoja@udenar.edu.co; javierrevelof@udenar.edu.co; laura.mancera@unir.net; dpeluffo@yachaytech.edu.ec.

[1]SEDMATEC, Corporación Universitaria Autónoma de Nariño, Colombia.

[2]GIIEE, Universidad de Nariño, Colombia.

[3]UNIR, España.

[4]SDAS, YachayTech. Ecuador.

Pages: 310-324

Resumen: Recientemente, en los campos del cambio climático y de la demanda de energía se ha volcado la atención al estudio y descubrimiento de patrones en energías renovables, como la fotovoltaica. Estos patrones pueden obtenerse extrapolando la radiación en función de las bandas del espectro electromagnético capturadas por satélites Landsat y MODIS de la NASA, donde los algoritmos de redes neuronales artificiales (ANN) y máquinas de soporte vectorial (SVM) han conseguido los mejores modelos. No obstante, la adquisición de los datos de entrenamiento desde esas fuentes es costosa y la exploración de funciones kernel para esta aplicación es escasa. Por consiguiente, en este estudio se realizaron ajustes en los anteriores aspectos, principalmente a través de: acoplamiento de nuevos kernels a ANN y SVM en la biblioteca scikit-learn, contribuyendo en la reutilización y robustez de estos algoritmos; e implementando un marco experimental para afinar hiperparámetros, generando así resultados comparables con el estado del arte.

Palabras-clave: Energía fotovoltaica; Función kernel; Imágenes satelitales; Maquinas de Soporte Vectorial; Redes Neuronales Artificiales.

Comparison of kernel functions for the prediction of the photovoltaic energy supply

Abstract: Recently, at the fields of climate change and energy demand have turned their attention to the study and discovery of patterns in renewable energies, such as the photovoltaic-type. Such patterns can be obtained by extrapolating radiation based on the electromagnetic spectrum bands captured by NASA's Landsat and MODIS satellites, where artificial neural network (ANN) and support vector machine (SVM) algorithms have produced the best models. Nonetheless,

the acquisition of training data from those sources is expensive, as well as it lacks the exploration of kernel functions for this application. Therefore, in this study, adjustments were made in the above aspects, mainly through: coupling of new kernels to ANN and SVM in the scikit-learn library, contributing to the reuse and robustness of these algorithms; and implementing an experimental framework to tune hyper-parameters, thus generating results comparable to those reported in the state of the art.

Keywords: Artificial Neural Networks; Kernel function; Photovoltaic energy; Satellite images; Support Vector Machines.

1. Introducción

Para entender la complejidad del universo, el ser humano ha abstraído los aspectos más relevantes de la realidad buscando los patrones inherentes en su esencia. Abstracción que ha trascendido a las maquinas (Machine Learning), esculpiendo la realidad mediante conjuntos de datos contextualizados a través de algoritmos para la obtención de modelos, donde las Redes Neuronales Artificiales (del inglés, Artificial Neuronal Networks - ANN), y las Máquinas de Soporte Vectorial, (del inglés Support Vector Machine - SVM) han venido ganando terreno debido a su capacidad de predicción en datos con relaciones no lineales (Sanchez Anzola Nicolas, 2015). Sin embargo, es posible mejorar los resultados de estos algoritmos con funciones kernel alternativas, las cuales permiten llevar el conjunto original de características a un espacio de Hilbert de dimensión mayor (Peluffo-Ordóñez et al., 2015). No obstante, generalmente se utilizan solo un par de funciones kernel; la gaussiana y polinómica, dejando configuraciones sin explorar (Pedregosa et al., 2011). Por otro lado, los patrones para oferta de energía fotovoltaica son de gran interés para los planes de energización a nivel mundial, debido a que permiten visualizar las mejores zonas geográficas para la ubicación de receptores de energía y de esa manera cubrir las necesidades de demanda y mitigar el cambio climático. En contraste a lo anterior, se proyecta una demanda de aproximadamente 376 Gwh para el año 2030, suponiendo una tendencia constante de 3.130 Kwh per cápita (Grupo Banco Mundial, 2019), con un crecimiento demográfico del 15% (Naciones Unidas, 2015); significando que, pese a los esfuerzos de aumentar la capacidad energética, se proyectaría alrededor de 674 millones de personas sin electricidad, cubriendo solo el 32% (Banco Mundial, 2019).

En ese contexto, en Colombia, desde la Unidad de Planeación Minero Energética (UPME) se han desarrollado esfuerzos mediante planes de energización sostenibles (PERS), donde se ha estudiado la demanda y la oferta de energía; alineados al compromiso de los países en la lucha contra el cambio climático expuestos en el acuerdo de Paris (Acciona, 2015). Específicamente, en el departamento de Nariño, el desabastecimiento de petróleo produce escases de combustible para los medios de transporte y zonas sin fluido eléctrico continuo (Pantoja, 2014). Desde el PERS Nariño nace el proyecto Análisis de Oportunidades Energéticas (ALTERNAR), donde se obtuvo un modelo de extrapolación de irradiancia en función del espectro electromagnético aterrizando en bandas capturadas en imágenes provenientes de los satélites Landsat y MODIS de la NASA (Cabrera et al., 2016).Allí, se utilizaron varios algoritmos de aprendizaje automático, donde ANN y SVM obtuvieron los mejores resultados, mediante un entrenamiento de partición simple tomando un tercio del total de datos para prueba, con hiperparámetros

afinados por defecto. No obstante, la adquisición de los datos de entrenamiento toma tiempos considerables, el afinamiento de hiperparámetros resulta parcializada para las subsecuentes comparaciones y, en general, no se han explorado kernels alternativos.

Por lo tanto, en esta investigación se refuerza los anteriores hallazgos mediante: mejoras en la adquisición de datos de entrenamiento a través del procesamiento distribuido; optimización de los algoritmos de ANN y SVM acoplando funciones kernel alternativas en la biblioteca scikit-learn; mejoras en las métricas y resultados del estado del arte (Cabrera et al., 2016) implementando un marco experimental utilizando tres diferentes normalizaciones de datos y afinamiento de hiperparámetros por algoritmos genéticos. De esta investigación se destaca que los kernel alternativos solo han mejorado los resultados para SVM y han empeorado para ANN. Se tiene la hipótesis de que los kernels funcionan mejor para SVM que en otros algoritmos de clasificación y predicción lineal, porque este algoritmo utiliza el criterio del margen máximo, abriendo la discusión para realizar nuevos experimentos teniendo en cuenta más algoritmos de aprendizaje automático que usan funciones kernel, realizando dicha evaluación con distintas bases de datos para clasificación y predicción.

2. Materiales y Métodos

En los experimentos del presente estudio se utilizaron principalmente los datos de las bandas del espectro electromagnético capturado por los satélites Landsat y MODIS de la NASA pre-procesados por la Universidad de Nariño (Cabrera, 2016) ver Tabla 1, *Dim* representa la dimensión, *BD* base de datos y *Tam* tamaño en disco en GB. Los scripts de exploración, limpieza y extracción se escribieron en SQL utilizando el sistema gestor de bases de datos PostgreSQL 9.4.

El desarrollo de los experimentos de comparación se escribieron en Python utilizando las bibliotecas: pandas para el preprocesamiento de datos, scikit-learn para extrapolar la radiación en función de las bandas de las imágenes satelitales con los algoritmos SVM y ANN; k vecinos cercanos (KNN) para interpolación, pykrige para interpolación con kriging (Monger, 2016) sklearn-deap para afinamiento por algoritmos genéticos y matplotlib para visualizar los resultados y mapas de radiación; utilizando una malla de puntos con proyección Mercator 3857 dentro del shape (shp) del departamento de Nariño.

Base de datos	Acrónimo	Registros	Dimensiones	Tam
Landsat	Lan	4.071.185	7	21
MODIS	MOD	373.670.476	7	79

Tabla 1 – Bases de datos para comparación de kernels

También, se estructuró un marco experimental sintetizado en la Figura 1, donde se visualiza, en primera instancia, la estructuración del conjunto de datos de entrenamiento desde PostgreSQL; posterior a ello, se selecciona un afinador de hiperparámetros, lista de normalizadores y algoritmos para encontrar los mejores modelos para luego ser evaluados usando la técnica de partición de datos de entrenamiento y prueba;

finalmente, se visualizan y rescatan los resultados para cada combinación, realizando las respectivas comparaciones. A continuación, se detallan los procesos más relevantes del marco experimental.

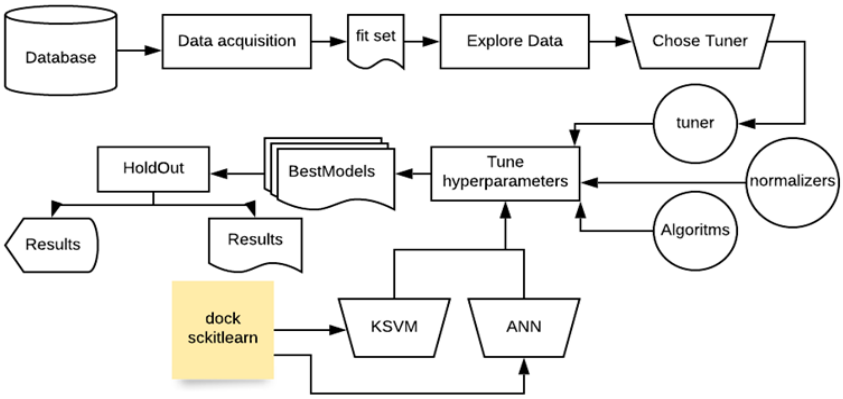


Figura 1 – Marco experimental de comparación.

2.1. Adquisición y exploración de datos

Para la adquisición de datos, se restauró en PostgreSQL las bases de datos Landsat y MODIS de Cabrera (2016), correspondientes a imágenes satelitales del departamento de Nariño para los años 2005 a 2014. Estas imágenes son capturadas realizando un barrido diario, resolución 500 metros por pixel para MODIS; y un barrido cada 16 días y 30 metros por pixel de resolución para Landsat. En la Figura 2 se muestran las tablas de reflectancia e irradiancia de esta base de datos.

La tabla "reflectance" almacena las conversiones del número digital DN de las bandas del espectro electromagnético infrarrojo, visible y ultravioleta acorde a lo estipulado por los sensores en Landsat y MODIS (ESRI, 2019), y la tabla "irradiance_grid_450" representa el mapeo sinusoidal de irradiación de una región tomados de VAISALA INC (3TIER) (Cabrera et al., 2016).

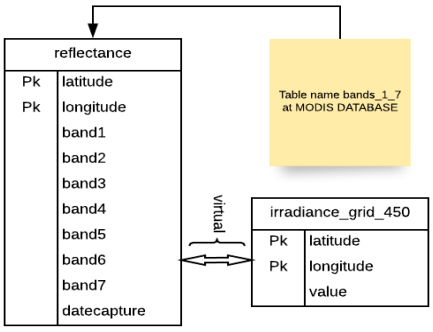


Figura 2 – Tablas de base de datos Landsat y MODIS.

Los datos de entrenamiento se obtienen desde el sistema gestor de base de datos, mediante las ecuaciones (1), (2) y (3) donde Y, π y ⋈ son los operadores del algebra relacional de agrupación, proyección y producto cartesiano natural, respectivamente; pk es la llave primaria, A la tabla "reflectance", con los elementos x_i representando a los atributos de A; así mismo, B representa la tabla "irradiance_grid_450" (ver Figura 2).

$$R_1 = \Upsilon\ pk, avg(x_1)\ x_1, \dots, avg(x_n)\ x_n\ (A) \tag{1}$$

$$R_2 = \pi * (B) \tag{2}$$

$$R_3 = R1 \bowtie R2 \tag{3}$$

Debido a que π y ⋈ son costosas computacionalmente es necesario realizar una distribución del procesamiento para estas operaciones, por lo tanto, la ecuación (1) se transforma como se indica en las ecuaciones (4), (5) y (6).

$$R_{k,m} = \Upsilon_k^m\ pk, \sum x_1\ x_1, \dots, \sum x_n\ x_n, count(*)\ c\ (A) \tag{4}$$

$$R_0 = \cup_{k=0}^{m} all(R_{i,j}) \mid \{i = k * m\ , j = t + i\ , t \leq n \rightarrow t = m, t > n \rightarrow t = n\} \tag{5}$$

$$R_1 = \Upsilon\ pk, \frac{\sum x_1}{\sum c}\ x_1, \dots, \frac{\sum x_n}{\sum c}\ x_n\ (R_0) \tag{6}$$

2.2. Afinamiento de hiperparámetros

Para realizar una comparación objetiva y justa de los resultados de las funciones kernel en los algoritmos ANN y SVM, es necesario aplicar un método de afinamiento de hiperparámetros común para cada algoritmo. En este estudio se utilizó búsqueda evolutiva por algoritmos genéticos, utilizando la biblioteca sklearn-deap por ofrecer el mejor compromiso en sus métricas de calidad y tiempo (Banerjee et al., s. f.,2017). Con el resultado de las búsquedas del afinador, se analizan los resultados de cada kernel agrupados por algoritmo y normalizador (Normalizer, Min-Max, Standard), escogiendo por cada grupo los mejores modelos. Estas búsquedas se obtuvieron mediante la biblioteca scikit-learn, acoplando las funciones kernel en los algoritmos SVM y ANN.

2.3. Acoplamiento de funciones kernel a SVM y ANN

Para realizar el acoplamiento de las funciones kernel, se tomaron las definiciones formales expresadas por Belanche (2016), ver Tabla 2. Estas funciones se encapsularon de forma estática en una clase denominada kernelF (ver Figura 3).

Para poder acoplar estas funciones se usó, las matrices de *Gram* a para SVM, y "*callables*" utilizando la aproximación de *Nystroem* para ANN. Posteriormente, se crean nuevas

clases que heredan los modelos SVM y ANN de la biblioteca *scikit-learn* (Buitinck et al., 2013), sobrescribiendo y agregando nuevas funciones kernel a su funcionamiento, como se puede ver en la Figura 3.

Kernel	
Gausiano (RBF)	**Triangular (tr)**
$e^{-\sum_{i=1}^{d}\gamma(x_i-x_i')^{\beta}}$ $\gamma > 0\,,\beta \in (0,2]$	$\begin{cases}\|x-x'\| \le a \to 1-\dfrac{\|x-x'\|}{a}\\ \|x-x'\| > a \to 0\end{cases}$ $a > 0$
Radial basic (Rb)	Rational quadratic (Rq)
$\left(\sum_{i=1}^{d} e^{-\gamma(x_i-x_i')^2}\right)^m$ $\gamma > 0\,, m \in \boldsymbol{N}$	$1-\dfrac{\|x-x'\|^2}{\|x-x'\|^2+a}$ $a > 0$
Canberra (Can)	Truncated (Tru)
$1-\frac{1}{d}\sum_{1}^{d}\gamma\frac{\|x_i-x_i'\|}{\|x_i\|+\|x_i'\|}$ $\gamma \in (0,1]$	$\frac{1}{d}\sum_{1}^{d}\max\left(0,\frac{\|x_i-x_i'\|}{\gamma}\right)$ $\gamma > 0$

Tabla 2 – Formulación matemática de las funciones kernel utilizadas.

Formalmente el truco kernel se introduce en SVM para clasificación en la clase SVC y se nutre de más funciones kernel con la clase *KSVC* mediante la matriz de Gram K como se muestra en las ecuaciones (7) y (8):

$$\boldsymbol{max: L(\alpha) = \sum_1^n \alpha_i - 1/2 \sum_{i=1}^n \alpha_i \alpha_j y_i y_j K\langle x_i, x_j\rangle} \quad (7)$$

$$s.a: \sum_1^n \alpha_i y_i = 0\,, 0 \le \alpha_i \le C, i = 1,2,\dots,n \quad (8)$$

Un proceso similar se sigue para el algoritmo de regresión provista con la clase *KSVR* cuya expresión formal se muestra en las ecuaciones (9) y (10**)**:

$$\boldsymbol{max: L(\alpha) = \sum_1^n(\alpha_i^- - \alpha_i^+)\, y_i - \epsilon \sum_1^n(\alpha_i^- + \alpha_i^+) - \frac{1}{2}\sum_{i,j=1}^n(\alpha_i^- - \alpha_i^+)(\alpha_j^- - \alpha_j^+)K\langle x_i, x_j\rangle} \quad (9)$$

$$\boldsymbol{s.a: \sum_1^n(\alpha_i^- - \alpha_i^+) = 0\,, 0 \le (\alpha_i^-, \alpha_i^+) \le C, i = 1,2,\dots,n} \quad (10)$$

Para ANN se utilizó la aproximación de *Nystroem* (SI et al., 2017) mediante una transformación de base sobre el vector de características. Formalmente:

$$Z_L = W_L . a_{L-1} + b_L \tag{11}$$

$$N_t = NyosTream(kf) \tag{12}$$

$$X = Nt * X \tag{13}$$

$$a_0 = X \tag{14}$$

$$a_L = f(Z_L) \tag{15}$$

$$\hat{y} = predict(X_{batch} * Nt^T) \tag{16}$$

Donde la ecuación (11) representa la suma ponderada, mientras que (12) la transformación de *NyOstrem* preparada con una función kernel kf (Buitinck et al., 2013). La ecuación (13) es la generación de la transformación, conjuntamente(14) es la activación inicial o capa de entrada y (15) la función de activación que se ha definido lineal por defecto, buscando que las funciones kernel sean las que cumplan el oficio de la transformación del espacio de características, en este caso un espacio de Hilbert de dimensión infinita. Cabe resaltar que es necesario sobrescribir el método *predict* del clasificador o regresor en KANNC y KANNR para realizar la transformación sobre cada *batch* del descenso del gradiente estocástico SGD *(16)*[1].

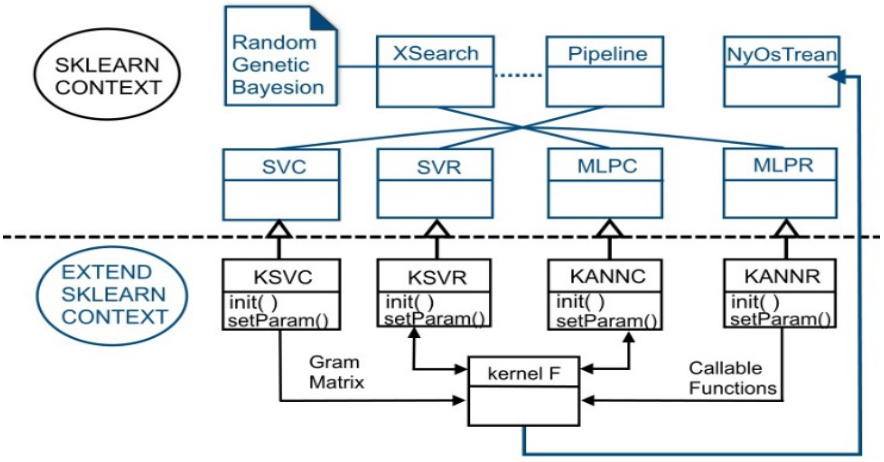


Figura 3 – Acoplamiento de funciones kernel a *scikit-learn*.

[1] Esta extensión puede instalarse como se muestra en https://github.com/magohector/sklearnkernels.

3. Resultados

Una vez obtenido un método objetivo alineado a una biblioteca de aprendizaje automático, se procedió a realizar los experimentos para evaluar el desempeño de los algoritmos en tres experimentos. El primero se centra en la inspección de la influencia de las funciones kernel, donde se corroboró la funcionalidad del acoplamiento y que los kernels alternativos mejoran el desempeño de los modelos dependiendo de los datos tal como lo enuncia Belanche (2016); en el segundo se pretende obtener una vista panorámica del desempeño de los modelos usando una búsqueda agrupada por algoritmo; finalmente, el tercero en una búsqueda agrupada por función kernel para afinar los resultados y realizar la comparación del compromiso de cada modelo en su coeficiente de determinación (r^2), desviación estándar (std) y tiempo (t).

Mediante un afinamiento de hiperparámetros por algoritmos evolutivos con 10 generaciones, tamaño de población igual a 50, probabilidad de mutación del gen 0.1, y tamaño de torneo de 3, para los hiperparámetros de regularización C, kernel (k), gamma (γ), termino independiente (a) y configuraciones para validaciones cruzadas hechas con 5 particiones tomando el 20% de los datos de prueba, se obtuvieron los resultados de la Tabla 3, Tabla 4 y Tabla 5.

DB	Algoritm	Normalizer	k	r2	Std	γ	a	C
Lan	KSVR	**Min_Max**	tr	0,9101	0.0186	1000.0	-	10000
			rq	0,9313	**0.0150**	-	1.0	1000
		Standard	rq	0,7482	0.0268	-	0.1	10000
			rbf	0,8953	**0.0145**	0.001	-	10000
	KANNR	Min_Max	rq	0,6150	0.1611	-	0.001	10
			hyp	0,7019	**0.0606**	1000.0	-	100
		Standard	**hyp**	0,8438	**0.0269**	0.001	-	1000
MOD	**KSVR**	**Min_Max**	**rq**	**0,8956**	**0.0173**	-	0.1	1000
			can	0,8074	0.0267	0.1	-	10000
		Normalizer	**tr**	**0,7573**	**0.0599**	0.1	-	10000
		Standard	**rq**	**0,8950**	**0.0097**	-	1.0	1000
			rbf	0,8599	0.0171	1.0	-	100
	KANNR	Min_Max	tru	**0,6485**	**0.0389**	1.0	-	1000
		Standard	**hyp**	**0,7492**	**0.0501**	1000.0	-	10

Tabla 3 – Afinamiento de hiperparámetros por algoritmo.

En la Tabla 3 se han descartado los resultados con R^2 menores 0.6. Análoga a la configuración de afinamiento anterior, los resultados de los experimentos de la Tabla 4 y Tabla 5 se han obtenido afinador a cada función kernel de forma independiente, descartando los resultados con R^2 menor a 0.7.

Model	Normalizer	Kernel	R^2	Std	t	γ	A	C
KANNR	Min_Max	rbf	0,7013	0,1039	10,6	1.0	-	10000.0
	Normalizer	can	0,7279	0,0449	7,5	10.0	-	10.0
		hyp	0,8455	0,0261	5,8	**1.0**	**-**	**100.0**
	Standard	can	0,7221	0,0372	4,8	100.0	-	1000.0
		rbf	0,7170	0,1014	11,7	0.01	-	10.0
KSVR	Min_Max	**rquadratic**	**0,9313**	**0,0150**	**2,3**	-	**1.0**	**1000.0**
		rbf	0,9130	0,0257	0,0	10.0	-	10.0
		radial_basic	0,9045	0,0127	3,9	0.1	-	1000.0
		tru	0,8838	0,0169	4,2	100.0	-	10000.0
		can	0,8761	0,0186	4,4	1.0	-	100.0
		hyperbolic	0,8563	0,0167	1,1	0.01	-	10000.0
		triangle	0,7904	0,0306	1,9	1000.0	-	1000.0
	Normalizer	can	**0,8358**	0,0265	4,5	0.01	-	10000.0
		tru	0,7856	0,0491	3,6	0.01	-	10000.0
	Standard	triangle	0,9302	0,0137	2,1	**100.0**	**-**	**1000.0**
		rbf	0,9079	0,0062	0,5	0.01	-	10000.0
		radial_basic	0,8993	0,0062	23,5	0.01	-	10000.0
		can	0,8473	0,0117	4,3	10.0	-	1000.0
		hyperbolic	0,8305	0,0194	1,1	0.01	-	100.0
		rquadratic	0,7482	0,0268	2,3	-	0.1	10000.0

Tabla 4 – Afinamiento por función kernel, base de datos Landsat.

Model	Normalizer	Kernel	R^2	Std	t	g	A	C
KSVR	Min_Max	rbf	0,9026	0,0082	0,1	**10.0**	**-**	**100.0**
		rquadratic	0,8956	0,0173	2,9	-	0.1	1000.0
		triangle	0,8455	0,0264	2,4	10.0	-	100.0
		radial_basic	0,8442	0,0371	3,3	1.0	-	10.0
		Tru	0,8187	0,0214	4,0	100.0	-	10000.0
		can	0,8074	0,0268	4,7	1.0	-	1000.0
		hyperbolic	0,7779	0,0332	1,3	0.01	-	10000.0
	Normalizer	tru	0,8464	0,0122	4,1	**0.001**	**-**	**10000.0**
		can	0,7720	0,0369	5,4	10.0	-	1000.0
	Standard	rquadratic	0,8950	0,0097	2,6	**-**	**1.0**	**1000.0**
		radial_basic	0,8826	0,0283	4,2	0.001	-	10000.0
		rbf	0,8599	0,0171	0,1	1.0	-	100.0
		triangle	0,8152	0,0300	2,0	10.0	-	10.0
		tru	0,8113	0,0104	4,0	1.0	-	100.0
		can	0,7859	0,0246	4,5	0.01	-	10000.0
KANNR	Min_Max	radial_basic	0,7658	0,0603	8,0	10.0	-	10000.0
		hyperbolic	0,7547	0,0398	7,1	0.001	-	1000.0
		radial_basic	0,7375	0,0936	8,4	0.1	-	10000.0

Tabla 5 – Afinamiento por función kernel, base de datos MODIS.

La Figura 4 muestra aquellos modelos cuyo rango de confianza (μ±σ) sea superior a 0.9 para Landsat y 0.87 para MODIS, de esta manera se puede visualizar los modelos con mejor coeficiente de determinación (R^2), la mediana y la variabilidad de los resultados entre modelos en su índice intercuartil (IQR).

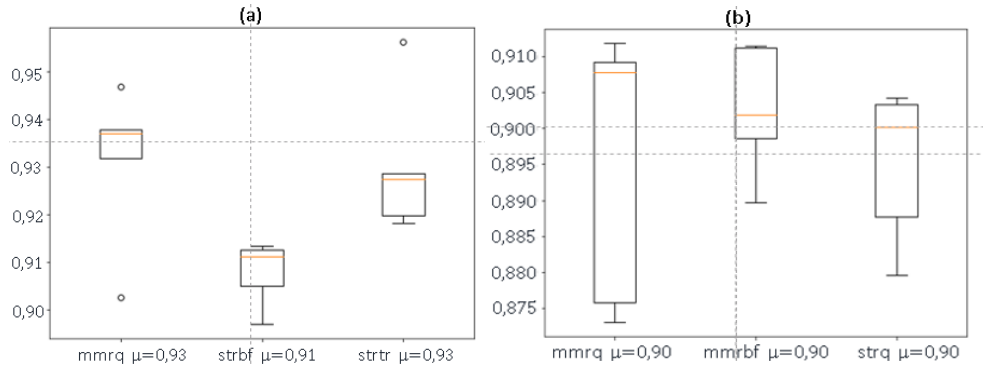


Figura 4 – Caja de bigotes base de datos Landsat(a) y MODIS(b).

Con los modelos visualizados en la Figura 4 se realizó 1000 entrenamientos con partición de datos aleatoria, usando un tercio de datos para prueba. Con estos últimos resultados se obtuvieron las métricas de suma de error absoluto (SAE), error absoluto medio (MAE), raíz del error cuadrático medio (RMSE) y coeficiente de determinación (R^2). Para realizar la comparación se ha sumado la media y una desviación estándar para el caso de SAE, MAE y RMSE y restado la media y una desviación estándar para R^2; haciendo que la variabilidad en los datos este en el límite del rango de confianza para cada métrica. Luego se agregaron los mejores resultados mostrados por Cabrera (2016) a saber: multi layer perceptron ensemble (MLPE), multi layer perceptron (MLP), y (KSVM) y finalmente se escalaron las métricas utilizando normalizador Min-Max. Las Figura 5 y Figura 6 muestran las comparaciones realizadas con las bases de datos Landsat y MODIS.

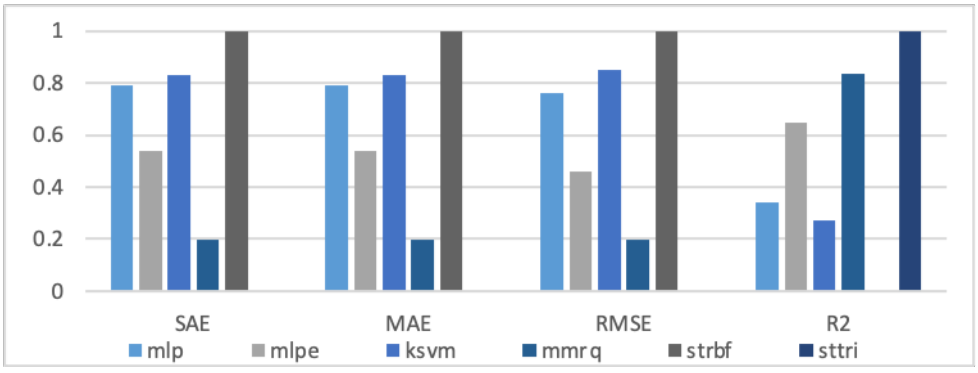


Figura 5 – Comparación funciones kernel base de datos Landsat.

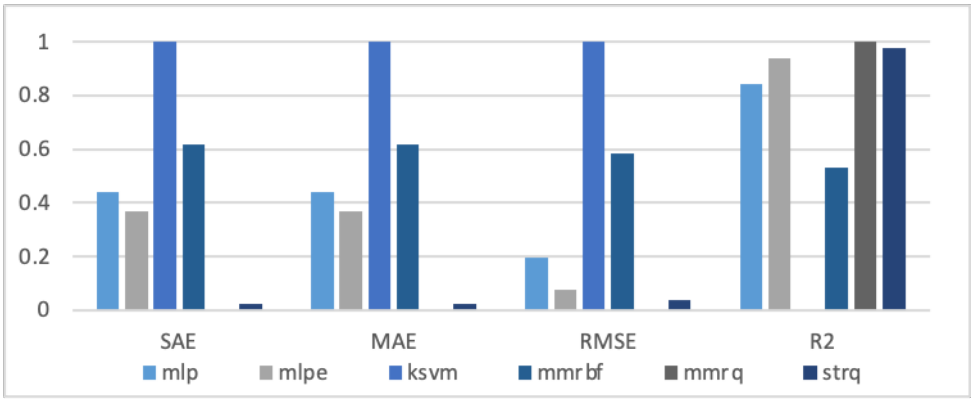


Figura 6 – Comparación funciones kernel base de datos MODIS.

En las Figura 5 y Figura 6 mmrq corresponde a la combinación de una normalización Min-Max con el kernel Rational Quadratic; strbf corresponde a la combinación de una normalización Standard con el kernel RBF; sttri corresponde a la combinación de una normalización Standard con kernel Triangular; mmrbf corresponde a una normalización Min-Max con kernel gausiano (RBF); y finalmente, strrq a la normalización Standard con kernel Triangular. Al interpolar los datos de los puntos predichos del modelo con mejor compromiso en sus métricas de calidad, se obtuvieron los mapas para la región de Nariño (Colombia), los cuales se observan en la Tabla 3. Figura 7 para los datos del satélite Landsat y Figura 8 para los datos satelitales MODIS. Para la escala de color se ha utilizado la misma escala colores utilizados por los mapas de Cabrera (2016), donde el R^2 de la mejor interpolación del experimento ha sido aproximadamente 0.89.

Para los mapas b y c de la Figura 7 y Figura 8, se interpolaron los datos de predicción mediante el algoritmo kriging ordinario con variograma esférico, meseta 410, 450 pasos, sin efecto pepita, con un R^2 de aproximadamente 0.95 (ver mapas b), el mapa de la derecha se interpoló con el algoritmo KNN utilizando 4 vecinos (k=4). Con un R^2 de aproximadamente 0.93 (ver mapas c). [2]

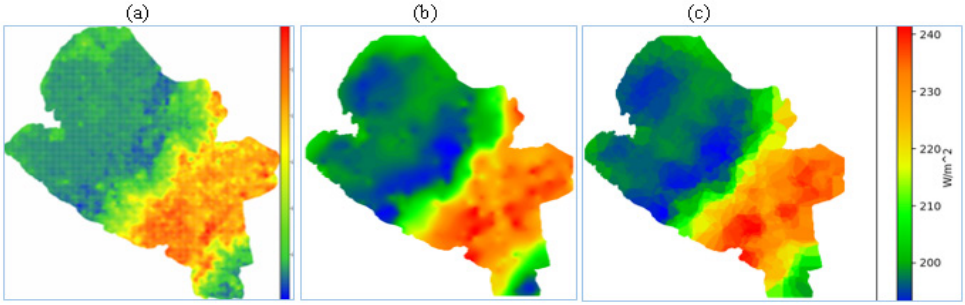


Figura 7 – Mapas Landsat, Cabrera (2016) (a), Kriging (b) y KNN (c).

[2] Todos los scripts de este estudio pueden detallarse en https://github.com/magohector/fkernel.

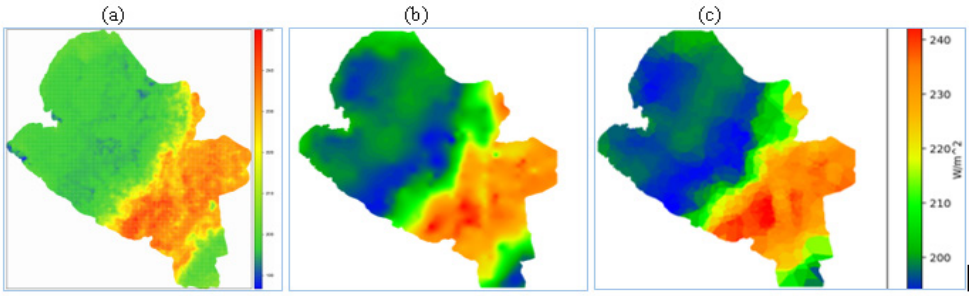


Figura 8 – Mapas MODIS, Cabrera (2016) (a), Kriging (b) y KNN (c).

4. Discusión y análisis de los resultados

De acuerdo a lo enunciado por Belanche (2016) y Yu et al (2016), las funciones kernel como la Canberra, Truncated y Triangular proporcionan mejores resultados que la función RBF para cierto tipo de datos, esto se puede apreciar en los elementos resaltados en la Tabla 4 y 5.

Los experimentos de afinamiento de hiperparámetros muestran que, considerando una búsqueda agrupada por kernel, mejoran las métricas de calidad de los modelos de aprendizaje automático y evitan el sobreajuste. En contraste, el experimento agrupado por función kernel obtiene mejores indicadores de calidad que en el experimento de afinamiento agrupado por algoritmo. En el experimento de afinamiento por algoritmo (Tabla 3) se observa, en primera instancia, que los modelos obtenidos por KANNR y KSVR obtienen mejores resultados para la base de datos Landsat; en segunda instancia, se visualiza que los modelos provistos por KSVR en general funcionan mejor que KANNR; y finalmente, los normalizadores tipo Normalizer han dado resultados de R^2 inferiores a 0.5 para la mayoría de las configuraciones afinadas en las dos bases de datos. Las anteriores afirmaciones son corroboradas por los resultados enmarcados en la Tabla 4 y Tabla 5. Igualmente, se puede apreciar configuraciones de hiperparámetros adicionales, por ejemplo, para la base de datos MODIS es probable que la mejor ruta posible de configuración sea la combinación del algoritmo KSVR, con el normalizador Min-Max y kernel RBF. Lo anterior se deduce a partir de los resultados de R^2 y el tiempo de entrenamiento (t); además, la Figura 4 muestra que la variabilidad de los resultados de esta configuración es menor a sus dos competidores más cercanos.

Por otra parte, en la Tabla 5 pueden distinguirse que para la base de datos Landsat existen dos rutas de configuración, ambas partiendo de modelos obtenidos por KSVR. La primera incluye un normalizador Min-Max con kernel Rational Quadratic, mientras que la segunda un normalizador Standard con kernel Triangle. El tiempo de entrenamiento es similar para las dos configuraciones, pero al observar la Figura 4 se aprecia que el kernel Rational Quadratic tiene menor variabilidad.

Al comparar los resultados obtenidos con los de Cabrera (2016), se observa en la Figura 5 que, para los datos Landsat, los modelos obtenidos por KSVR con normalizador

Standard, kernel Triangle y normalizador Min–Max, kernel Rational Quadratic presentan las mejores métricas, ratificando con la partición de datos para validación que la mejor combinación está dada por un normalizador Standar con kernel Triangle como la mejor alternativa. Por otro lado, para los datos MODIS, en la Figura 6 se visualiza que el algoritmo KSVR normalización Min-Max o Standard con kernel Rational Quadratic obtienen el mejor compromiso en ese orden, aunque la brecha para este conjunto de datos está muy cercana al mejor resultado de Cabrera (2016). No obstante, KANNR ha empeorado los resultados en comparación a los de Cabrera (2016). Finalmente, los mapas resultantes siguen un patrón relativamente similar al obtenido por Cabrera (2016), donde la interpolación por kriging (Monger, 2016) parece ofrecer una mejor visualización como lo evidencia Monger (2016). Por otro lado, los mapas obtenidos con los modelos de predicción mejorados además de obtener un R^2 más alto, detectan zonas con picos de radiación por encima de 235Kw/m² y pueden consignar, de esta manera, menos incertidumbre a la hora de tomar decisiones para la instalación de receptores en zonas geográficas determinadas. Igualmente, se logra descartar algunas zonas donde la radiación está muy por debajo de los 200 $\frac{Kw}{m^2}$.

5. Conclusiones y trabajos futuros

El marco experimental planteado permitió confirmar postulados y corroborar la tesis de que las funciones kernel en estudio se acondicionan mejor frente a los conjuntos de datos Landsat y MODIS y afinar la configuración de hiperparámetros ecuánimemente. Los resultados del análisis reflejan que SVM con los normalizadores Standard y Min-Max y los kernel Triangular y Rational Quadratic son los más indicados para realizar regresión sobre los datos Landsat y MODIS. Consecuentemente como trabajo futuro, se podría estudiar la posibilidad de mejorar el desempeño del afinamiento de hiperparámetros con búsquedas bayesianas y, asimismo, evaluar la posibilidad de hacer uso de funciones de activación no lineales y funciones kernel para evaluar el compromiso entre métricas de calidad y coste computacional inspeccionando bases de datos de uso común, tanto para problemas de regresión como clasificación.

La visualización de los resultados permitió evaluar justa y objetivamente a cada algoritmo, los mapas finales corroboran que los patrones obtenidos son similares a (Cabrera et al., 2016), con obtención de picos de radiación más altos que sirvan para ubicar receptores de energía fotovoltaica. Igualmente, el acoplamiento de las funciones kernel a la biblioteca scikit-learn permite interactuar de forma familiar con los algoritmos desarrollados, significa un aporte a la comunidad del software libre, y a estudios académicos de punto de partida para construir modelos acoplados a esta biblioteca usada a nivel mundial.

Finalmente, es importante considerar la exploración del compromiso de las SVM y las funciones kernel para escenarios de predicción de series de tiempo, consolidar un aplicativo de información geográfica interactivo para acceder a los mapas finales, como también nutrir a la biblioteca desarrollada acoplada a scikit-learn con la extensión de algoritmos de clasificación y regresión lineal con el truco kernel siempre que sea viable técnicamente.

Referencias

Acciona. (2015). *Compromisos de los países más contaminantes contra el cambio climático* [Sostenibilidad]. https://www.sostenibilidad.com/cambio-climatico/compromisos-paises-contaminantes-cambio-climatico/

Badii, M.H., A. Guillen, & J.L. Abreu. (2016). *Energías Renovables y Conservación de Energía (Renewable Energies and Energy Conservation) http://www.spentamexico.org/v11-n1/A12.11(1)141-155.pdf*

Belanche. (2016). *Developments in kernel design*. 369-378. https://upcommons.upc.edu/handle/2117/23278

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). API design for machine learning software: Experiences from the scikit-learn project. http://arxiv.org/abs/1309.0238

Cabrera, Champutiz, Calderon, & Pantoja. (2016). *Landsat and MODIS satellite image processing for solar irradiance estimation in the department of Narino-Colombia*. 1-6. https://doi.org/10.1109/STSIVA.2016.7743306

ESRI. (2019). *Imágenes y Teledetección*. https://www.arcgis.com/apps/Cascade/index.html?appid=5072b8d56cef4f7bb5d24e5d840461da

GIRÓN, Alicia. (s. f.). *Objetivos del Desarrollo Sostenible y la Agenda 2030: Frente a las Políticas Públicas y los Cambios de Gobierno en América Latina*. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0301-70362016000300003

Grupo Banco Mundial. (2019). *Consumo de energía eléctrica (kWh per cápita)*. https://datos.bancomundial.org/indicador/EG.USE.ELEC.KH.PC?end=2018&start=1975&view=chart

Monger, S. H., Morgan, E. R., Dyreson, A. R., & Acker, T. L. (2016). Applying the kriging method to predicting irradiance variability at a potential PV power plant. *Renewable Energy*, *86*, 602-610.

Naciones Unidas. (2015, diciembre 14). *Población*. https://www.un.org

Ortiz, & Parra. (2019). *De los hidrocarburos a las energías renovables en Colombia |Cultura Latinoamericana*. https://editorial.ucatolica.edu.co/index.php/ RevClat/article/view/2743

Pantoja, F. (2014). *PERSN*. PERS. pers.udenar.edu.co

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peluffo-Ordóñez, Lee, & Verleysen. (2015). *Generalized kernel framework for unsupervised spectral methods of dimensionality reduction—IEEE Conference Publication*.

Revelo, J., & Peluffo, D. (s. f.). *Educación y Formación Cultural en Fuentes de Energía Alternativa para el Departamento de Nariño*. 8.

Sanchez Anzola Nicolas. (2015). *Vista de Máquinas de soporte vectorial y redes neuronales artificiales en la predicción del movimiento USD* https://revistas.uexternado.edu.co/index.php/odeon/article/view/4414/5004

SI, HSIEH, & Cho-Jui. (2017). *Memory efficient kernel approximation*. https://dl.acm.org/doi/abs/10.5555/3122009.3122029

Yu, B., Wang, Y. T., Yao, J. B., & Wang, J. Y. (2016). A comparison of the performance of ANN and SVM for the prediction of traffic accident duration. *Neural Network World*, *26*(3), 271-287. https://doi.org/10.14311/NNW.2016.26.015

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 31/08/2020
Aceitação/Acceptance: 30/10/2020

Mapeando la evolución investigativa de la web semántica: análisis y visualización cienciométrica

Santiago Guerrero Narváez, Miguel Ángel Niño Zambrano,
Gustavo Adolfo Ramírez González

santoguerrero@unicauca.edu.co; manzamb@unicauca.edu.co; gramirez@unicauca.edu.co

Universidad del Cauca, Cl. 5 #4-70, 190003, Popayán, Cauca, Colombia.

Pages: 325-336

Resumen: La Web semántica ha tomado gran importancia para el desarrollo múltiples industrias, tanto a nivel comercial como científico, generando la necesidad de varios estudios del campo. Este documento analiza la Web Semántica desde 1995 hasta el 2020Q1 (primer cuarto del año). El principal aporte de este estudio es la identificación de patrones en las publicaciones, incluyendo, a los más importantes autores e instituciones, categorías de alta frecuencia y palabras clave, revistas y contribuciones de autor, temas de investigación más importantes y evolución del campo basado en análisis grupal. La metodología utilizada fue una aproximación cienciométrica basada en un análisis bibliométrico. Registros desde WOS y Scopus han sido reunidos usando dos herramientas bibliométricas: CiteSpace y SciMAT.

Palabras clave: Web; Semántica; Scopus; CiteSpace; SciMAT.

Mapping the evolution of Semantic Web research: scientometric analysis and visualization

Abstract: The last decades the research of semantic web has taken great importance for the development of multiple industries at a commercial level and from the scientific point of view, this led to the need for various studies on this field. This paper analyzes the field of semantic web from 1995 to 2020Q1 (first quarter of the year) identifying the patterns from the publications, including the most important authors and institutions; high-frequency categories and keywords; journal and author contributions; most important research topics and evolution of the field based on cluster analysis. A scientometric approach based on bibliometric analysis is employed. Records from WOS and Scopus have been gathered using two bibliometric tools CiteSpace and SciMAT.

Keywords: Semantic; Web; Scopus; CiteSpace; SciMAT.

1. Introducción

La Web Semántica se la puede entender como la evolución de la web tradicional, en donde se busca, que el contenido y los metadatos de cada recurso web, puedan ser identificados

con mayor precisión. Este campo intenta responder al problema de como identificar la gran cantidad de datos que son subidos y actualizados en la web cada segundo.

El campo de la investigación de la Web Semántica está evolucionando como cualquier otro campo interdisciplinario de estudio científico. Su relevancia está basada en las necesidades sociales actuales y su impacto en muchos otros campos. Para analizar esta evolución es necesario implementar una revisión cienciométrica (Broadus, 1987). Este artículo describe los métodos de visualización de información y técnicas para mapear los comandos de conocimiento adoptados para ilustrar el estado de la Web Semántica. Todos los artículos usados en este estudio han sido publicados en la colección principal de Web of Science (WOS) y Scopus (fuentes de principal relevancia) la metodología de búsqueda en estas herramientas fue aportada por los estudios de (Bar-Ilan, 2007) y (Gavel, 2008). Las herramientas de soporte usadas son CiteSpace (Chen, 2006) y SciMAT (Cobo, 2012).

Podemos encontrar otros artículos que realizan un análisis sobre la literatura encontrado sobre Web Semántica como en (Jensen, 2017) y (Ismail, 2016), sin embargo, estos no utilizan las herramientas de CiteSpace y SciMAT de forma conjunta, limitando una visión más amplia sobre los resultados encontrados en la literatura científica.

El presente artículo está organizado de la siguiente manera: la sección 2 describe la metodología que incluye la descripción de la base de datos para el análisis de visualización, el parámetro designado para configurar CiteSpace y SciMAT y el tipo de análisis a desarrollar de acuerdo a cada una de las herramientas. La sección 3 presenta los resultados del análisis con SiteSpace y SCiMAT. Finalmente, en la Sección 4, se presentan las conclusiones.

2. Metodología

2.1. Conjunto de datos para análisis de visualización

Decidimos adoptar la definición generalizada de "semantic web", que incluye varios tipos diferentes de trabajos acerca de este término, evitando varios trabajos de la web que no tengan enfoque semántico; esto se hizo tanto para Scopus como para WOS. Obtuvimos 21.039 registros para fuentes de Scopus y el período de búsqueda se realizó entre 1996 a 2020Q1 y para WOS se encontraron 12.891 documentos relacionados con artículos de revistas y actas de conferencias. Los criterios de exclusión se enfocaron en filtrar los documentos que no cumplieran con la definición completa (ejemplo, que tengan web o semántica de forma separada) y que no se encuentren en las áreas afines a la informática.

La Figura 1 proporciona una visión a largo plazo de los resultados por tiempo, incluidos los resultados de 1995 al primer trimestre de 2020 para Scopus y de 1996 a 2017 para la colección principal de WOS. Utilizando esta visión por tiempo podemos ver la evolución en cuanto a número de publicaciones.

2.2. Diseño de parámetros

En CiteSpace hay diferentes configuraciones disponibles para varios niveles de confiabilidad de los resultados. La división del tiempo es uno de los mecanismos más relevantes utilizados para generar la visualización de la red. Para este estudio, el valor

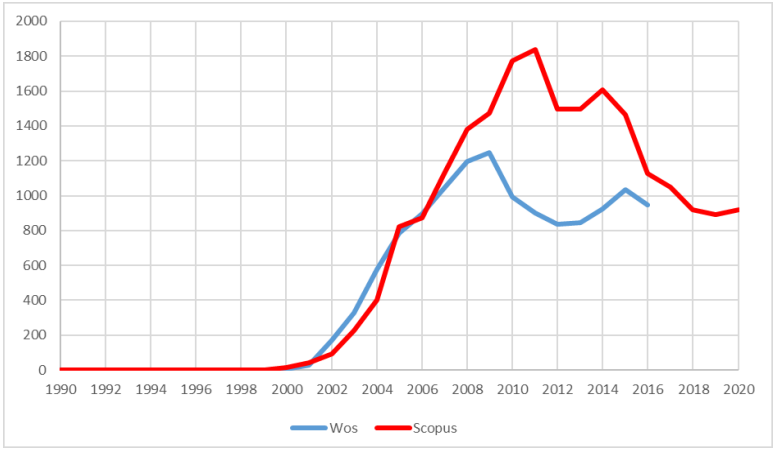


Figura 1 – Distribución de documentos a lo largo del tiempo para Scopus y WOS. Eje X: años, Eje Y: cantidad de documentos.

de escala de tiempo fue establecido en 1. En este caso, esta escala significa que todo el intervalo de tiempo desde 1995 hasta 2017 es dividido en segmentos de 1 año para el procesamiento.

Para la herramienta SciMAT utilizamos el registro de Scopus, para este registro se aplicó un proceso de duplicación (seleccionar las palabras que pueden referirse al mismo concepto y agruparlas en un solo término) a la unidad de análisis seleccionada (palabras clave), para agrupar aquellas palabras que representan el mismo concepto. Para desarrollar nuestro estudio, todo el período (1995–2020Q1) fue dividido en 5 períodos consecutivos, 1995-2001, 2002-2006, 2007-2011, 2012-2017 y 2018-2020Q1. Se encontraron 21.039 registros, 2000 son registros de 2018-2020Q1, 8.971 son registros de 2012 a 2017, 7.595 son registros de 2007 a 2011, 2.418 son registros de 2002 a 2006 y 55 son registros de 1995 a 2001.

2.3. Análisis de diseño

CiteSpace se utiliza para aplicar tres tipos de técnicas bibliométricas (como se desarrolló en (Song, 2016). El primero es el análisis de coautores, que toma los nombres de los autores, los países de afiliación y las afiliaciones institucionales como unidades de análisis y examina las coincidencias de autores, las coincidencias de países y las coincidencias de instituciones. El segundo es el análisis de co-palabras, basada en palabras clave; y por último, el análisis de co-citaciones que identifica a los autores y a las revistas citadas conjuntamente, basándose en referencias, artículos, autores y revistas.

SciMAT (Cobo, 2011), brinda dos funcionalidades para para explorar un campo de investigación: análisis de rendimiento y mapeo científico (Van Raan, 2004). El primero se centra en el impacto basado en citas de la producción científica, y el segundo se centra en la búsqueda de la estructura conceptual de la producción científica. (Cobo, 2011)

define un enfoque que combina ambos. El análisis de co-palabras se usa en un marco longitudinal, que permite analizar y rastrear la evolución de un campo de investigación a lo largo de períodos de tiempo consecutivos de acuerdo con (Garfield, 1994).

Siguiendo el enfoque propuesto por (Cobo, 2011) tres pasos fueron establecidos para analizar el campo de la investigación de la Web Semántica: detección de grupos de investigación, trazado de diagramas estratégicos y trazado de áreas temáticas. Para la detección de grupos de investigación, la herramienta SciMAT utiliza un índice bibliométrico normalizado equivalente a palabras clave (Callon, 1983) y (Callon, 1991). Esto es seguido por una agrupación de palabras clave a tópicos/temas usados en el algoritmo de centros simples (Coulter, 1998). El trazado de diagramas estratégicos se logra basándose en los valores de centralidad de clúster y rango de densidad (Callon, 1991). La centralidad mide la intensidad de la interacción de un grupo con otros grupos, entre más fuerte este vínculo es más relevante al campo de investigación. La densidad mide la intensidad del enlace interno dentro del grupo. Entonces, con estas dos medidas, los grupos pueden clasificarse en cuatro grupos a medida que se implementa la herramienta: 1) temas motores; 2) temas básicos y transversales; 3) temas emergentes y 4) temas altamente desarrollados y aislados. El tercer paso es trazar áreas temáticas. Se analiza la evolución de los temas de investigación para detectar las principales áreas de evolución general del campo de investigación y se construye su mapa de evolución basado en interrelaciones.

3. Resultados

3.1. Análisis con CiteSpace

La productividad de los autores en un campo se puede representar aplicando análisis de coautoría. La red de coautoría resultante se compone de 706 nodos y 674 enlaces, que contiene todos los autores principales del campo. Cada nodo representa un autor, y los

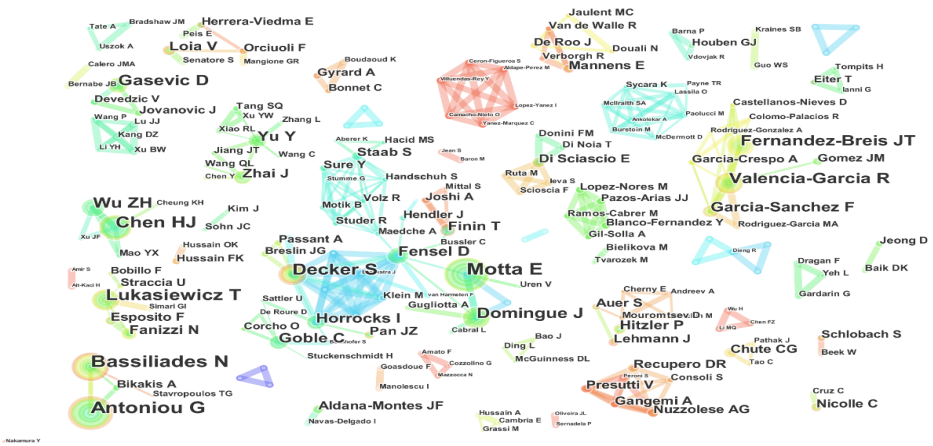


Figura 2 – Red de coautoría, con 706 nodos y 674 enlaces

enlaces entre los autores representan la coautoría de los artículos. Los diferentes nodos representan el número de publicaciones proporcionalmente. Está claro que Enrico Motta (Motta E), profesor de la Open University en el Reino Unido, es el autor más citado en el campo de la Web Semántica, seguido de Antoniou G., Bassiliades N. y Decker S. El ancho de las líneas y círculos indica la proporción de las relaciones cooperativas y el número de publicaciones. Los diferentes colores representan diferentes fases y corresponden a diferentes años desde 1996 hasta 2017.

Para describir las características de la distribución de publicaciones relacionadas con la Web Semántica, generamos una red basada en las instituciones y países de los autores. Las conexiones entre los países que realizan investigaciones en el campo de la Web Semántica se muestran en la Figura 3. A partir de esta figura, podemos concluir que las principales contribuciones, en términos de la producción total de la publicación, provienen principalmente de autores de China, Estados Unidos y Alemania. En cuanto a la distribución de publicaciones, 1799 artículos de China, seguidos de 1759 de EE. UU., 1134 de Alemania, 1101 de Inglaterra, 927 de Italia, 924 de España, 692 de Francia, 451 de Grecia, 442 de Canadá, 371 de Australia, 370 de Corea del Sur, 357 de India y 354 de Austria. Los otros países tienen menos de 352 documentos. Los investigadores en los Países Bajos tienen las relaciones de trabajo más consistentes con investigadores en otros países. La identificación de los nodos con la mayor centralidad entre medias indica que países como Francia (0.21), Alemania (0.19) y EE. UU. (0.16) ocupan posiciones clave en los caminos relevantes en el desarrollo de la investigación de la Web Semántica.

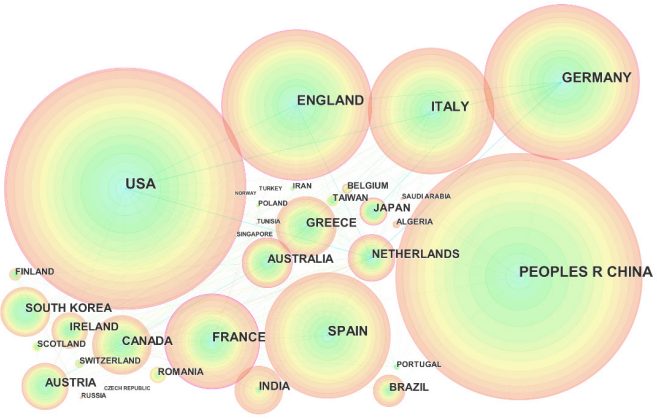


Figura 3 – Red de país, con 80 nodos y 403 enlaces

La Figura 4 presenta las instituciones con mayor productividad. Podemos encontrar que la Universidad Abierta del Reino Unido, la Universidad de Southampton, la Universidad de Manchester, la Academia de Ciencias de China, son las que más productividad tienen en el campo de la Web Semántica.

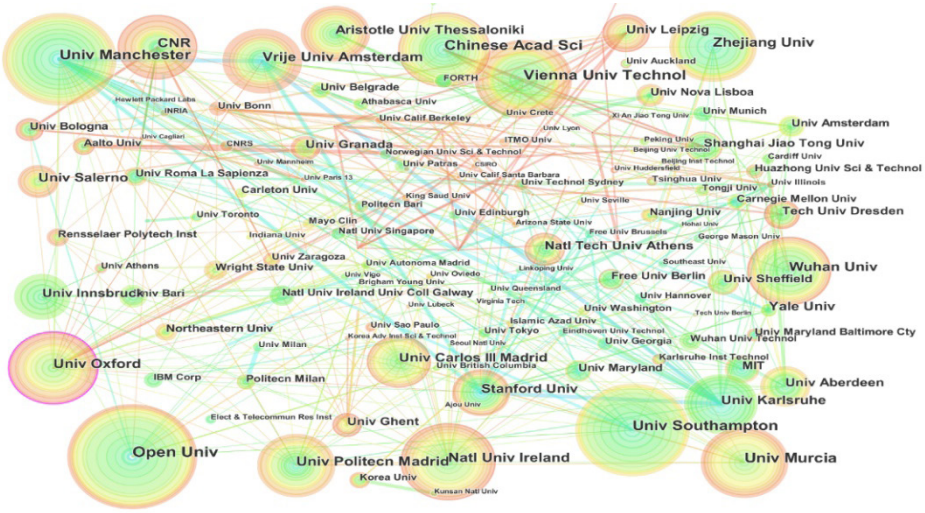


Figura 4 – Red de instituciones, con 690 nodos y 893 enlaces.

Para analizar las tendencias emergentes, construimos una red de categorías de temas concurrentes en la investigación de la Web Semántica, que incluye 132 nodos y 516 enlaces, como se muestra en la Figura 5. Entre todos los nodos, los que tienen una alta centralidad entre medias, como lo indican los anillos morados, se observan las categorías: "Engineering" (0.32) y "Mathematics" (0.2). Esta centralidad indica que son los principales puntos de inflexión intelectual que vinculan los estudios realizados en diferentes fases y puede haber tenido influencias significativas en el desarrollo de la investigación de la Web Semántica.

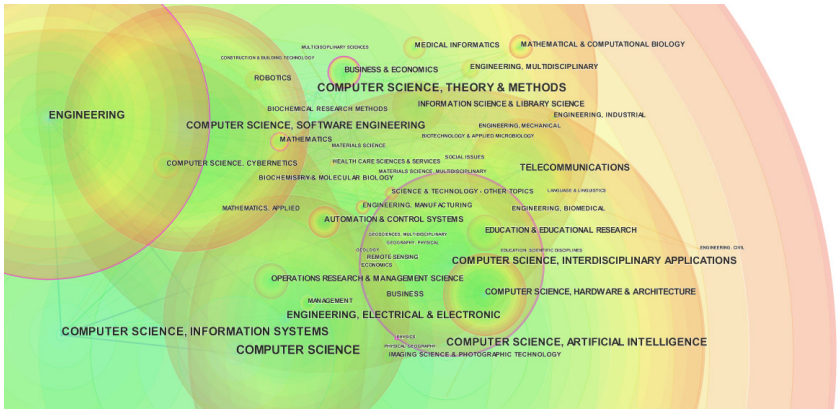


Figura 5 – Red de categorías de materias concurrentes en WOS con 132 nodos y 516 enlaces

Para analizar las tendencias emergentes, construimos una red de categorías de temas concurrentes en la investigación de la Web Semántica, que incluye 292 nodos y 1139 enlaces. Las palabras clave de un artículo proporcionan información de su contenido y el desarrollo del campo de investigación. La red de palabras clave concurrentes, con 292 nodos y 1139 enlaces, se muestra en la Figura 6. Cada nodo de palabras clave está compuesto por varios anillos, y los anillos y enlaces están representados en un conjunto de colores correspondientes a los años de ocurrencia. Para facilitar la lectura, nos centramos solo en la parte más poblada de los clústeres.

De todas las palabras que se muestran en la Figura 6, las palabras clave más relevantes por conteo son "Ontology" con recuentos de citas de 2836, seguida por "System" (621) y "Web" (618). Además, este gráfico puede revisarse por medida de centralidad. Según Song (2016), la centralidad representa los principales puntos de inflexión intelectual y actúa como puentes que unen las revistas en diferentes fases, el elemento mejor clasificado por centralidad es "Ontology" con 0.37, seguido de "Hypertext" (0.33), "Metadata" (0.25).

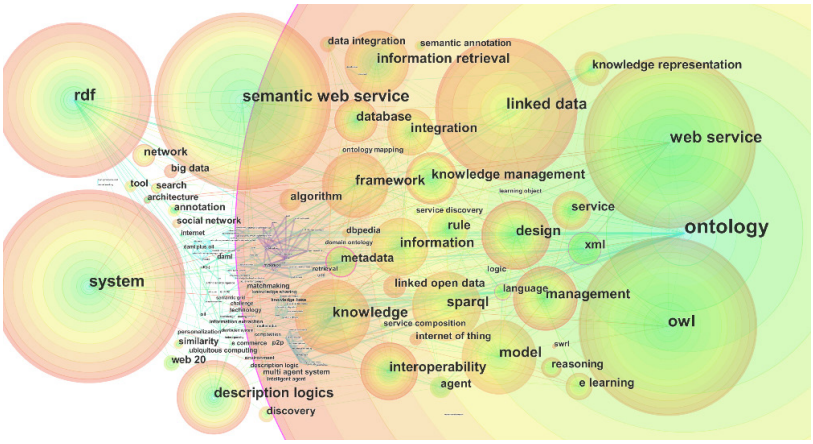


Figura 6 – Red de palabras clave concurrentes, con 292 nodos y 1139 enlaces

Con base en las estadísticas de WOS, las 10 principales revistas en investigación de la Web Semántica se enumeran en la Tabla 1. La revista con la mayor cantidad de publicaciones registradas para esta fuente es "Conference Notes in Computer Science", que ha publicado 11.390 artículos sobre investigación en Web Semántica. Es importante remarcar que estos son registros en WOS y no son especialmente todos los registros de cada revista. Para detectar las revistas citadas más importantes de la investigación en Web Semántica y evaluar las influencias y los patrones de citas de los estudios centrales en este dominio, se generó un mapa de citas de la revista, como se muestra en la Figura 7. El tamaño de cada nodo en el mapa representa la frecuencia de citas de la revista correspondiente (Chen 2014).

SOURCE JOURNAL	COUNTS	%
LECTURE NOTES IN COMPUTER SCIENCE	11390	21.96
JOURNAL OF WEB SEMANTICS	3163	6.1
LECTURE NOTES OF ARTIFICIAL INTELLIGENCE	2994	5.77
SCIENTIFIC AMERICAN MAGAZINE (SCI AM)	2206	4.25
COMMUNICATION OF THE ACM	2039	3.93
ARTIFICIAL INTELLIGEENCE	1743	3.36
IEEE INTELLIGENT SYSTEMS AND THEIR APPLICATIONS	1716	3.31
IEEE INTERNET COMPUTING	1692	3.26
IEEE INTELLIGENT SYSTEMS	1582	3.05
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING	1463	2.82

Tabla 1 – Las 10 principales publicaciones de revistas fuente en WOS 51859

La Figura 7 contiene 310 nodos y 1015 enlaces, derivados de 51.859 publicaciones relevantes. En orden cronológico, las revistas con más citas dentro del alcance del conjunto de datos incluyen las Lecture Notes in Computer Science (con 11390 citas). En términos de frecuencia de co-citación, las revistas más influyentes son: Lecture Notes in Computer Science y Journal of Web Semantics. En la figura 7, los nodos con una alta centralidad intermedia indican que las revistas como Communication of the ACM (0.65), ACM Transactions on Information Systems (0.22), IEEE Intell Syst App (0.21) representan los principales puntos de inflexión intelectual como puentes que unen las revistas. Estos hallazgos indican fluctuaciones significativas en el impacto de estas revistas.

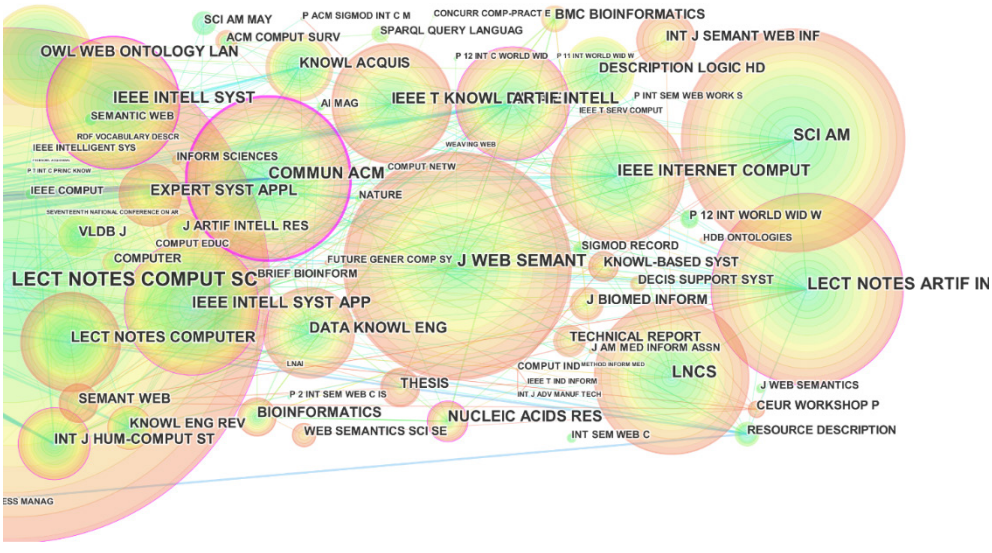


Figura 7 – Red de co-citas de revista, con 310 nodos y 1015 enlaces.

3.2. Análisis con SciMAT

Para analizar los temas más destacados del campo Web Semántica, se dividieron los documentos en 5 periodos diferentes, el criterio de separación fue intentar homogeneizar el número de documentos por periodo. El análisis muestra diversos términos centrales por periodo (los más relevantes para cada periodo), cada término contiene valores de densidad y centralidad, los cuales ubican al término en diferentes áreas de desarrollo de la investigación. Las áreas posibles son:

- Temas motores (con alta densidad y alta centralidad). Altamente desarrollados e importantes para la construcción del campo científico y con una alta adopción (poco asilados).
- Temas transversales (con baja densidad y alta centralidad). Importantes para el desarrollo del campo científico, estables y una alta adopción (poco asilados), pero con poco desarrollo.
- Temas emergentes o en declive (con baja densidad y baja centralidad). Poco desarrollados con una evolución al alza o en retroceso.
- Temas periféricos (con alta densidad y baja centralidad). Altamente desarrollados internamente, pero que se encuentran aislados, tienen un papel marginal en el desarrollo del campo científico

El periodo 2005-2001 fue donde empieza el desarrollo de la investigación en Web Semántica. Para este periodo podemos encontrar como tema principal “Internet”, densidad (13.24), centralidad (40.55). Como indican los datos arrojados por SciMAT, este concepto tiene una alta centralidad y alta densidad. Esto indica que es un tema motor y está altamente desarrollado y compartido por los documentos científicos.

Durante el período 2002-2006, se aumentó el número de temas. Para este periodo podemos encontrar como temas principales para el análisis los siguientes:

- “Owl”, con valores: densidad (0.43), centralidad (4.71). Estos valores lo ponen como un tema emergente, que, aunque no tiene un alto desarrollo, está en crecimiento y adopción.
- “Context”, con valores: densidad (0.36), centralidad (4.27). Estos valores lo ponen como un tema emergente, que, aunque no tiene un alto desarrollo, está en crecimiento y adopción.
- “Framework” con valores: densidad (1.57), centralidad (3.4). Estos valores lo ponen como un tema básico y transversal, el cual no ha tenido mucho desarrollo, pero es considerado como importante y está medianamente adoptado.
- “Application-software” con valores: densidad (2.91) y centralidad (7.08). Estos valores lo ponen como un tema motor.

Durante el período 2007-2011 se aumentó el número de temas, los temas motores fueron: “Vocabulary” Y “Twitter”. Los temas periféricos fueron “Dynamic-description-logic” Y “Triple-store”. Los temas transversales fueron “Text” Y “QoS”. Los temas emergentes fueron: “Information-retriveal” Y “Jena”. Los temas con mayor número de documentos fueron: “Collaborative-systems” Y “Information-retrieval”.

Para este periodo podemos encontrar como temas principales para el análisis los siguientes:

- "Information-retrieval" con valores: densidad (0.06), centralidad (3.14). Estos valores lo ponen como un tema básico y transversal, que tiene poco desarrollo, es considerado como importante y está medianamente adoptado.
- "Jena" con valores: densidad (0.26), centralidad (3.24). Estos valores lo ponen como un tema básico y transversal, que tiene poco desarrollo, es considerado como importante y está medianamente adoptado.
- "Semantic-discovery" con valores: densidad (1.01), centralidad (3.5). Estos valores lo ponen como un tema motor, que tiene un mediano desarrollo y una mediana adopción.

Durante el período 2012-2017 se aumentó la cantidad de temas, los temas motores fueron "SWS", "Middleware" y "Tag-recomendation". Los temas periféricos fueron "Hbase", "Query-expansion" y "Metrics". Los temas emergentes fueron "RDF", "Service-composition" y "Classification" Los temas transversales fueron: "Entity-resolution", "Linked-data" y "Ambient-intelligence". Los temas con más documentos asociados o con más recuento fueron: "Ontologies" con 880 documentos, "Service-composition" con 298 documentos "Linked-data" con 259 documentos.

Durante el período 2018-2020Q1 se redujo el número de temas. El tema motor fue "Interoperability". El tema periférico "Linked-Data". El tema emergente fue "Agents". El tema transversal fue "SPARQL". El tema con más recuento de documento fue "SPARQL" (109).

Para este periodo podemos encontrar como temas principales para el análisis los siguientes:

- "SPARQL" con valores: densidad (0.33), centralidad (2.29). Estos valores lo ponen como un tema transversal, el cual tiene poco desarrollo, pero tiene una mediana adopción e importancia.
- "Embedding" con valores: densidad (0,48), centralidad (1,52). Estos valores lo ponen como un tema emergente, que, aunque no tiene un alto desarrollo, está en crecimiento y adopción.
- "Ontology" con valores: densidad (0.43), centralidad (2.23). Estos valores lo ponen como un tema en declive, tiene un bajo desarrollo, y como en periodos anteriores estaba altamente posicionado, la adopción de este tema está en bajada.
- "Interoperability" con valores: densidad (0.53), centralidad (2.06). Estos valores lo ponen como un tema emergente, que, aunque no tiene un alto desarrollo, está en crecimiento y adopción.

4. Conclusiones

La investigación en Web Semántica está bien definida en las fuentes seleccionadas (WOS y Scopus), esto significa que ambos tienen registros similares y su evolución en ambas fuentes es similar. Los resultados para el análisis de palabras clave en ambas herramientas son similares.

Los resultados para SciMAT son más explícitos según períodos y podemos organizar de una mejor forma los resultados, además de filtrar, corregir y unir, términos con significados similares. Para CiteSpace tenemos una mayor riqueza visual y podemos hacer un análisis más global, debido a los gráficos con alto grado de información y versatilidad que entrega la herramienta.

Las principales contribuciones provienen de autores de China, Estados Unidos y Alemania. El autor más productivo en el campo de la investigación es Enrico Motta (Motta E). La Universidad Abierta del Reino Unido, es la institución de mayor productividad. La categoría principal es "Informatics", pero la categoría de "engineering" puede representar el principal punto de inflexión intelectual que vinculan los estudios realizados en diferentes fases. La palabra clave más relevante por conteo es "Ontology", pero "Ontology", "Hypertext" y "Metadata" son las palabras clave con más centralidad que representan principales puntos de inflexión intelectual y actúan como puentes que unen las revistas.

Aunque en los últimos años la investigación en esta área ha bajado considerablemente, no se ha perdido el propósito de ahondar en herramientas donde las computadoras puedan tener mayor conciencia "awareness", esto se puede ver en este documento cuando aparecen términos nuevos como "IoT", "Natural Language Processing" o "Big Data" relacionados a los términos que antes fueron utilizados en Web Semántica.

Referencias

Bar-Ilan, J. (2007). Which h-index? A comparison of WoS, Scopus and Google Scholar. Scientometrics, 74(2), 257-271. https://doi.org/10.1007/s11192-008-0216-y

Broadus, R. (1987). Toward a definition of "bibliometrics". Scientometrics, 12(5-6), 373-379. https://doi.org/10.1007/BF02016680

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. Scientometrics, 22(1), 155-205. https://doi.org/10.1007/BF02019280

Callon, M., Courtial, J.P., Turner, W.A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. Information (International Social Science Council), 22(2), 191-235. https://doi.org/10.1177/053901883022002003

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. Proceedings of the National Academy of Sciences, 101(suppl 1), 5303-5310. https://doi.org/10.1073/pnas.0307513100

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for information Science and Technology, 57(3), 359-377. https://doi.org/10.1002/asi.20317

Chen, C. (2006). Information visualization: Beyond the horizon. Springer Science & Business Media.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. Journal of the American Society for Information Science and Technology, 62(7), 1382-1402. https://doi.org/10.1002/asi.21525

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. Journal of Informetrics, 5(1), 146-166. https://doi.org/10.1016/j.joi.2010.10.002

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. Journal of the American Society for Information Science and Technology, 63(8), 1609-1630. https://doi.org/10.1002/asi.22688

Coulter, N., Monarch, I., & Konda, S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. Journal of the Association for Information Science and Technology, 49(13), 1206-1223. https://doi.org/10.1002/(SICI)1097-4571(1998)49:13<1206::AID-ASI7>3.0.CO;2-F

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. IEEE transactions on intelligent transportation systems, 12(2), 596-614. https://doi.org/10.1109/TITS.2010.2092770

Gavel, Y., & Iselid, L. (2008). Web of Science and Scopus: a journal title overlap study. Online information review, 32(1), 8-21. https://doi.org/10.1108/14684520810865958

Garfield, E. (1994). Scientography: Mapping the tracks of science. Current Contents: Social & Behavioural Sciences, 7(45), 5-10.

Jiang, Z., Streit, D. A., & El-Gindy, M. (2001). Heavy vehicle ride comfort: literature survey. International Journal of Heavy Vehicle Systems, 8(3-4), 258-284. https://doi.org/10.1504/IJHVS.2001.001163

Song, J., Zhang, H., & Dong, W. (2016). A review of emerging trends in global PPP research: analysis and visualization. Scientometrics, 107(3), 1111-1147. https://doi.org/10.1007/s11192-016-1918-1

Van Raan, A. F. (2004). Measuring science. In Handbook of quantitative science and technology research (pp. 19-50). Springer Netherlands.

Jensen, J (2017). A systematic literature review of the use of Semantic Web technologies in formal education. British Journal of Educational Technology, 50(2), 505-517. https://doi.org/10.1111/bjet.12570

S, Ismail., T, Shaikh. (2016). A Literature Review on Semantic Web - Understanding the Pioneers' Perspective. The Sixth International Conference on Computer Science, Engineering and Applications, 6(11), 15-28. https://doi.org/10.5121/csit.2016.61102

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 20/08/2020
Aceitação/Acceptance: 16/10/2020

Trading Algorítmico para la Predicción de Series de Tiempo Financieras: Una Revisión Sistemática

Jesús-Rodrigo Navia-Rodríguez, Carlos-Alberto Cobos-Lozada, Martha-Eliana Mendoza-Becerra

jesusnavia@unicauca.edu.co; ccobos@unicauca.edu.co; mmendoza@unicauca.edu.co

Grupo investigación en Tecnologías de la Información (GTI), Universidad del Cauca, Popayán, 190003, Cauca, Colombia.

Pages: 337-357

Resumen: El mercado de activos financieros es un tema de gran interés para investigadores de diversas áreas debido a los desafíos que presenta, entre ellos, la hipótesis de mercados eficientes donde se cuestiona la posibilidad de predecir precios futuros con el análisis de precios históricos. En ciencias de la computación, matemáticas y áreas relacionadas, una alternativa para trabajar este tema consiste en usar o proponer técnicas para el análisis de series de tiempo de activos financieros que permitan obtener estrategias de negociación cada vez más eficientes. Este artículo expone técnicas, tendencias y datos utilizadas para el trading algorítmico de activos financieros, conforme se ha reportado en 43 artículos de Web of Science. Las técnicas que reportan los mejores resultados utilizan aprendizaje automático, metaheurísticas, redes neuronales y lógica difusa, pero a la fecha todavía es difícil definir cuáles de ellas son las mejores y en qué condiciones pueden ser la mejor opción.

Palabras-clave: Trading algorítmico, análisis de series temporales, predicción, aprendizaje de máquina, metaheurísticas, redes neuronales artificiales, lógica difusa, "portafolio".

Algorithmic Trading for Financial Time Series Prediction: A Systematic Review

Abstract: The financial asset market is a topic of great interest to researchers in various areas due to the challenges presented by, among others, the efficient market hypothesis, which questions the possibility of predicting future prices by means of an analysis of historical prices. In computer science, mathematics and related areas, an alternative in working on this topic is to use or propose techniques for a time series analysis of financial assets that makes it possible to attain increasingly efficient trading strategies. This article exhibits techniques, trends, and data used in the algorithmic trading of financial assets, as reported in 43 Web of Science articles. The techniques that report the best results use machine learning, metaheuristics, neural networks, and fuzzy logic. But it remains difficult to define which of these are most effective and under what conditions they may be considered as the best option.

Keywords: Algorithmic trading, time series analysis, prediction, machine learning, metaheuristics, artificial neural networks, fuzzy logic, "portfolio".

1. Introducción

Trading es la palabra en inglés usada como la acción de comprar y vender activos financieros (divisas, índices, acciones, bonos, entre otros) en un mercado bursátil. Es realizado por traders, personas que compran y venden dichos activos financieros. Los mercados bursátiles han avanzado tecnológicamente a tal punto, que han logrado automatizar la toma de decisiones para la compra o venta de estos activos financieros, mediante el trading automático o también llamado trading algorítmico (Víctor Manuel Pérez Santos 2014). El trading algorítmico tiene ventajas respecto al trading manual realizado por las personas, ya que no se afecta por las emociones al momento de tomar las decisiones. Los algoritmos funcionan como inyectores de liquidez, reducen los spreads (en el mercado de valores, un spread es la diferencia del valor de compra con el de venta) y los costos de las transacciones, disminuyendo así el riesgo. Los avances en el trading algoritmo le dan ventajas competitivas a las empresas que lo realizan, tal es el caso de la empresa Renaissance Technologies, la cual obtuvo utilidades por un valor de $2,5 billones de dólares en un año usando esta tecnología. Esta tecnología es hoy día un tema muy importante de investigación y la transferencia y adopción rápida de los mejores algoritmos es clave para los inversores (Vasquez and Muñoz 2015).

En el mercado bursátil es muy importante predecir los valores de los activos financieros en el tiempo (predicción de series de tiempo financieras), por esta razón, diversos investigadores de las ciencias de la computación y otros campos han contribuido aplicando técnicas basadas en aprendizaje de máquina, metaheurísticas, redes neuronales artificiales, lógica difusa, entre otras (Nair and Mohandas 2015).

El presente documento muestra una revisión de 43 artículos primarios reportados en Web of Science relacionados con el trading algorítmico en la predicción de activos financieros. Para el análisis de estos documentos se realiza una agrupación con base en la técnica que en cada artículo reporta los mejores resultados, siendo estos: aprendizaje de máquina, metaheurísticas, redes neuronales y lógica difusa. Para cada artículo se presenta además de una corta descripción, los datos, las fuentes de esos datos, los algoritmos con los que se hace la comparación de la técnica que se propone y las métricas usadas en la comparación.

El resto de este documento está organizado así: en la Sección 2 se describe el protocolo de investigación llevado a cabo para realizar el estado del arte. La Sección 3 muestra los resultados obtenidos de dicha revisión y la discusión de las observaciones principales de estos resultados, así como las limitaciones e implicaciones de la revisión en el campo de investigación. Y, finalmente, en la Sección 4 se presentan las conclusiones y el trabajo futuro que espera realizar el grupo de investigación en el área.

2. Protocolo de Investigación

El presente estado del arte se ha llevado a cabo siguiendo las guías provistas en (Kitchenham 2004), realizando tres etapas principales: Planificación, Ejecución y Discusión.

En la **Etapa de Planificación** se llevaron a cabo las siguientes actividades: 1) Definición de la pregunta de investigación, 2) Definición de la estrategia de búsqueda, 3)

Establecimiento de los criterios de selección de los estudios primarios, 4) Establecimiento de los criterios de evaluación de calidad, 5) Definición de la estrategia de extracción de datos, y 6) Selección de los métodos de síntesis.

Preguntas de Investigación: Esta investigación tiene como objetivo principal identificar las técnicas, métodos, modelos o similar, utilizadas en el trading algorítmico para la predicción de series de tiempo financieras. Para ello se han establecido las preguntas de investigación que se presentan en la **Tabla 1**. Con estas preguntas se recopila y categoriza la información existente, con el fin de realizar una revisión sistemática del tema, identificando los caminos prometedores y las mejores propuestas existentes a la fecha.

Para realizar la búsqueda de los artículos se utilizó la base de datos de Web of Science. La **Tabla 2** muestra la cadena de búsqueda final, la cual se obtuvo de un proceso iterativo de mejora con el análisis de los resultados reportados por cada versión de esta. La metodología utilizada para la construcción de la cadena de búsqueda se basó en PICOT (población-intervención-comparación-logro-tiempo) (Munn et al. 2018), estructurando los términos por población, intervención, comparación, resultados y tiempo (PICOT en inglés). La cadena se presenta dividida según PICOT y se unen con el operador AND. La búsqueda se realizó sobre el título, el resumen y las palabras claves de los artículos en la base de datos.

Criterios de Selección de los Estudios Primarios: Con el objetivo de definir cuáles de los resultados entregados por el motor de búsqueda se incluyen en la revisión, se realizaron y aplicaron unos criterios de aceptación basado en la información en el título, resumen y palabras clave. Los criterios de inclusión fueron los siguientes: 1) Artículos escritos en español o inglés, 2) Estudios que muestren la toma de decisiones en mercados financieros y 3) Propuestas de trading algorítmico.

Luego, se eliminaron los documentos que cumplieron con alguno de los siguientes criterios de exclusión: 1) El estudio utiliza portafolios de inversión, 2) El estudio utiliza técnicas de análisis de sentimientos u otras fuentes de datos relacionadas con noticias y similares, y 3) El estudio aplica procedimientos de trading algorítmico de alta frecuencia. Luego de aplicar los criterios de selección y dejar los documentos base para el estudio, se procedió a su evaluación.

Criterios de Evaluación de Calidad: como es común en las revisiones sistemáticas (Kitchenham 2004) se aplicaron unos criterios para evaluar la calidad de los estudios primarios recolectados. Esta evaluación es necesaria para comprender mejor las posibles comparaciones entre los estudios y para guiar la interpretación de los hallazgos. Para esto se desarrolló una prueba con cada una de las preguntas de investigación presentadas en la **Tabla 2** y un sistema de puntuación de tres valores (0, 0.5 , 1). En general, el valor de cero (0) significa que el artículo no aporta nada en relación con la pregunta de investigación o no cumple con lo que se pregunta; el valor de cero punto cinco (0.5) significa que el articulo realiza un aporte parcial a resolver la pregunta de investigación o cumple parcialmente con lo que pregunta; y finalmente el valor de uno (1) significa que hace un aporte muy completo a resolver la pregunta de investigación o cumple totalmente con lo que se pregunta. Como son 7 preguntas, los trabajos primarios tendrán un valor mínimo de 0 y máximo de 7.

Preguntas de Investigación	Motivación
Q1. ¿Cómo se ha ampliado el conocimiento o la comprensión del tema con la propuesta?	Busca determinar si un modelo de trading algorítmico es nuevo o combina de una forma novedosa elementos existentes que permiten identificar hacia donde se desarrolla la investigación en el tema.
Q2. ¿Qué tan claro y consistente es el documento ?	Identificar si la solución presentada es clara, fácil de comprender y muestra consistencia en cada uno de los componentes propuestos.
Q3. ¿Qué nivel de correlación se encuentra entre los resultados experimentales, la interpretación de estos y las conclusiones, es decir, qué nivel de soporte tiene cada conclusión?	Determinar si las conclusiones tienen su soporte en los resultados del experimento y que la interpretación de estos resultados no está sesgada.
Q4. ¿Qué tan bien apoyada es la evaluación de la propuesta y que tan creíbles son los resultados?	Determinar las métricas de evaluación de mayor uso en el trading algorítmico e identificar qué tan creíbles son los resultados.
Q5. ¿Cuán explicita y repetible es la implementación de la propuesta de acuerdo con sus objetivos y requisitos?	Identificar las herramientas de implementación en los modelos de trading algorítmico y definir si la descripción es lo suficientemente detallada para facilitar la repetición de la propuesta y experimentos.
Q6. ¿cuán detallado es el proceso de recopilación de datos, sus características y disponibilidad?	Identificar los activos financieros analizados con sus respectivas fuentes y características de análisis.
Q7. ¿cuán completo y bien documentado ha sido el proceso de investigación ?	Determinar si la propuesta ha sido bien realizada desde el punto de vista metodológico.

Tabla 1 - Preguntas de investigación

Estructura	Términos Alternativos y Sinónimos	Descripción
Población	(“financial” AND “trading” AND “stock”)	Define la temática general que se está buscando: financiera, negociación y stock.
Población	NOT (“high frequency” AND portfolio)	Se excluyen las temáticas: alta frecuencia y portafolios.
Intervención	(“time series” OR “alpha” OR “buy” OR “sell” OR “no trade”)	Se define qué actividad abordara: Series de tiempo, tipos de negociación, compara, venta y comercio
Tiempo	PUBYEAR > 2014 AND DOCTYPE (“ar” OR “ip”)	Se delimita el año y tipo de documento: mayores a 2014, artículo o artículo “in press”.

Tabla 2 - Cadena de búsqueda

Extracción de Datos: Para la clasificación de los documentos se utilizaron los datos extraídos del texto completo de estos junto con las respuestas encontradas a cada una de las preguntas de investigación previamente presentadas. Las preguntas de investigación, como sus posibles respuestas se muestran en la **Tabla 3**.

Preguntas	Respuestas
Q1	**Si**, presenta una nueva propuesta que transforma o modifica los modelos existentes de trading algorítmico por lo que genera nuevo conocimiento. **Parcialmente**, presenta modelos de trading ya existentes con un aporte al nuevo conocimiento de forma limitada. **No**, utiliza modelos o datos de trading diferentes a los usados en otros reportes.
Q2	La información del articulo **Si** es relevante en los modelos de trading algorítmico. La información del artículo es **Parcialmente** relevante en los modelos de trading algorítmico. La información del artículo **No** es relevante en los modelos de trading algorítmico.
Q3	Las conclusiones del modelo propuesto **Si** tiene un soporte cuantitativo. Las conclusiones del modelo propuesto exponen **Parcialmente** el soporte cuantitativo. Las conclusiones del modelo **No** tienen un soporte cuantitativo.
Q4	**Si** presenta las métricas de evaluación en el trading algorítmico para identificar qué tan creíbles son los resultados. **Parcialmente** presenta las métricas de evaluación en el trading algorítmico ... **No** presenta las métricas de evaluación en el trading algorítmico ...
Q5	**Si** presentan las herramientas de implementación de los modelos de trading algorítmico lo suficientemente detallada para facilitar la repetición de la propuesta y experimentos. **Parcialmente** presentan las herramientas de implementación ... **No** presentan las herramientas de implementación ...
Q6	**Si** presenta los activos financieros analizados con sus respectivas fuentes y características de análisis. **Parcialmente** presenta los activos financieros ... **No** presenta los activos financieros ...
Q7	**Si** tiene estructura consistente para que transmita el conocimiento del modelo de trading algorítmico investigado. **Parcialmente** tiene una estructura consistente ... **No** tiene la estructura consistente ...

Tabla 3 – Clasificación de evaluación en los documentos

Métodos de Síntesis: En la presente revisión se aplicó una síntesis de datos cuantitativos y cualitativos basada en una tabla que resumen los estudios seleccionados ordenaos según el año, enfoque, técnicas, referencia, criterio de calidad y datos con su respectiva fuente del activo financiero investigado. Además, un párrafo que resume cada artículo organizado por los tipos de técnicas usadas.

Calendario de la revisión sistemática: La presente revisión sistemático inició la recolección de datos en agosto de 2019 y finalizó en septiembre del mismo año.

3. Resultados obtenidos (Ejecución y Discusión)

Una vez definido el protocolo de la revisión de literatura, se procedió con la etapa de ejecución, esta misma fue guiada mediante las siguientes fases principales: 1) Se aplicó la cadena de búsqueda (**Tabla 2**) en la base de datos científica de Web of Science y se obtuvieron 206 documentos diferentes, 2) Con estos documentos se procedió a aplicar los criterios de selección (inclusión/exclusión) sobre el título, resumen y palabras claves de cada uno de ellos, una vez finalizada esta segunda fase, se obtuvo un total de 43

estudios relevantes. 3) Durante la tercera fase, se llevó a cabo la aplicación de los criterios de evaluación de calidad sobre cada uno de los 43 estudios relevantes, pero en este caso leyendo el documento completo. Las siguientes subsecciones presentan los resultados de la evaluación de los estudios de acuerdo con los criterios de evaluación.

Los 43 estudios relevantes se organizaron de acuerdo con la técnica utilizada, a saber: aprendizaje de máquina (Machine Learning, ML) (ver **Tabla 4**), metaheurísticos (MetaHeuristics, MH) (ver **Tabla 5**), redes neuronales artificiales (Artificial Neural Networks, ANN) (ver **Tabla 6**) y lógica difusa (Fuzzy Logic, FL) (ver **Tabla 7**). Es preciso comentar que los estudios relacionados con redes neuronales artificiales se separan de los estudios de aprendizaje de máquina, dada la evolución e importancia que esta área de investigación ha tenido en los últimos años en el trading algorítmico. A continuación, se describen cada uno.

3.1. Contribuciones basadas en aprendizaje de máquina

En 2015 se utilizó biclustering y clasificación con los k vecinos más cercanos (Biclustering and k-Nearest Neighbor, BIC-KNN) para encontrar patrones en series financieras de índices y acciones de empresas, obtenidas de Yahoo! Finance. Los algoritmos se implementaron en MATLAB obteniendo ganancias mayores al 39%. Este es el primer reporte del uso de biclustering para obtener patrones con datos financieros (Q. Huang et al. 2015).

En 2016 se utilizó K-means con descomposición por valores singulares (Singular Value Decomposition, SVD) en una propuesta denominada K-SVD, para realizar predicciones en series de tiempo financieras con datos de Yahoo! Finance del mercado de valores de EE.UU y de esta forma se obtuvieron utilidades mayores a las que ofrece el mercado (Rosas-Romero, Díaz-Torres, and Etcheverry 2016). También en 2016 se planteó que el uso de la hipótesis de mercado eficiente (Efficient Market Hypothesis, EMH) no permite predecir los retornos de sus activos. Para lograr demostrar esto, se evaluó el uso de máquinas de soporte vectorial vs redes neuronales artificiales (SVM vs ANN) encontrando que el modelo SVM obtuvo el mejor desempeño (Hsu et al. 2016).

En 2017 propusieron un modelo que utiliza una máquina de aprendizaje extrema sin supervisión (UnSupervised-Extreme Learning Machine, US-ELM) y la Regresión por Soporte Vectorial (Support Vector Regression, SVR), para predecir el índice de futuros de materia primas ENERGY con datos históricos obtenidos de http://www.mcxindia.com. Los autores afirman que la precisión de la predicción de la propuesta es mejor que los modelos de referencia (Das and Padhy 2017). También en 2017 se propuso un método para predecir las tendencias alcistas o bajista de series de tiempo financieras con el algoritmo MKK (modelo construido de los k valores alcistas o bajistas del histórico conocido) en activos financieros, utilizando los datos extraídos de Yahoo! Finance. La propuesta comparó el valor observado y encontró patrones de tendencia, donde en promedio las tasas de predicción del algoritmo MKK fue de 89.8%, y utilizando el indicador técnico bursátil MACD (Moving Average Convergence Divergence) fue tan sólo del 33.4% (Kim 2017). Este mismo año se utilizaron mapas auto organizados (Self-Organizing Maps, SOM) para generar opciones de compra o venta de activos financieros, de datos previamente pre procesados de Yahoo! Finance utilizando un

filtro Hodrick y Prescott. Como resultado se obtuvieron ganancias mayores respecto a propuestas similares (Nair et al. 2017). Siguiendo con el 2017, se realizó una propuesta para pronosticar precios europeos de emisiones de carbono con SVM, K-NN y árboles de decisión (Decision Trees, DT), usando Matlab R2015 con Statistic y Machine Learning Toolbox v10.0. Como resultado de la investigación se logró una tasa de precisión para todos los clasificadores superior al 48% (Hong, Jung, and Park 2017). Terminando el 2017, se presentó un método para predecir series de tiempo de los índices bursátiles S&P500, FTSE 100 y Nikkei 225, utilizando el modelo SVR desarrollado en Matlab con ayuda de LibSVM versión 2.91. El método inicializa los parámetros obtenidos de Yahoo! Finance optimizados por enjambre de partículas (Particle Swarm Optimization, PSO), luego entrena el SVR con los parámetros generados para realizar la predicción con modelos estadísticos guiados por error (MSE, MAPE, DS) y modelos guiados por ganancias (Naıve, Filter, DMA), obteniendo como mejor resultado a SVR_{MSE} según la prueba de Friedman. (Hu et al. 2017).

En 2018, proponen predecir índices del mercado financiero con un modelo hibrido que utiliza multi-categoría ponderada (WM) usando máquinas de soporte vectorial (SVM), programación genética (GEP), Random Forest, (RF) e indicadores técnicos para formar la propuesta llamada RF-WMGEPSVM. Utiliza como entrada los datos históricos de los índices recogidos de Quandl.com. Como conclusión se observó que los modelos híbridos funcionan mejor que sus versiones que no lo son (Thakur and Kumar 2018). Este mismo año (2018), proponen diferentes árboles de decisión: arboles de clasificación (CT), arboles de regresión (RT), árboles de clasificación fija (FCT) y árboles de regresión fija (FRT), para predecir el índice bursátil S&P500 utilizando datos obtenidos de Thomson-Reuters y Eikon. Utilizan procesos de Markov binarios que ayudan a ser flexible la captura de patrones arbitrarios y como resultado el modelo FCT tiene mejor relación el indicador técnico bursátil Sharpe (que mide la relación del riesgo con la rentabilidad), igual a 0,96 con un rendimiento anual compuesto del 7,87%, y un nivel de confianza en las estrategias mayor al 95%, que fue el mejor resultado obtenido por las cuatro técnicas compradas (Fiévet and Sornette 2018). También en 2018, se planteó un método que combina el análisis de componentes principales (Principal Component Analysis, PCA) y máquinas de soporte vectorial ponderado (Weighted Support Vector Machines, WSVM) para predecir puntos de negociación en acciones, utilizando indicadores técnicos bursátiles como variables de entrada según los datos históricos de las acciones obtenidos en Yahoo! Finance. Como resultado, la propuesta PCA-WSVM obtuvo una precisión máxima del 57% y una rentabilidad del 222,80%, resultados superiores a los obtenidos con PCA y WSVM por separado (Chen and Hao 2018).

En 2019 utilizan la transformada wavelet discreta para crear una nueva base de datos extendida mediante filtros pasa baja y pasa alta. Así estos nuevos datos se utilizan por una regresión por soporte vectorial (SVR), para predecir el precio del oro. La combinación de estas técnicas utiliza los datos mensuales desde mayo de 1972 a diciembre de 2016 disponibles en https://fred.stlouisfed.org y para generar la serie de tiempo original con un resultado que aumenta la riqueza de dos tercios en el periodo de predicción (Risse 2019). También en 2019 utilizan el modelo de clasificación Random Forest con un filtro de similitud difusa y entropía (FSAE) para predecir el índice bursátil S&P500 basado en cuatro configuraciones de FSAE con datos obtenidos de Yahoo! Finance. La mejor

configuración del filtro obtuvo una presión del 44.78% seguida por las configuraciones restantes con 44,38%, 43,06% y 42,91%, además el modelo Random Forest obtiene una precisión del 44,72% superando a los modelos KNN, Naive Bayes y árbol de decisión (Lohrmann and Luukka 2019). En el mismo año, 2019, proponen un bosque aleatorio adaptativo de conceptos recurrentes (Recurring Concepts Adaptive Random Forest, RCARF), para predecir el índice bursátil S&P500 con 10 indicadores técnicos bursátiles (Simple n-second moving average, Weighted n-second moving Average, Momentum, Stochastic K%, Stochastic D%, RSI, MACD, LarryWilliam's R%, A/D y CCI), calculados con ayuda de la biblioteca para análisis técnico TA-Lib con datos de quantquote. La implementación RCARF y los modelo de comparación RCD, DWM, AHOEFT y ARF se realizaron en el framework MOA con una evaluación Interleaved-Test-Then-Train, obteniendo como resultado un error máximo en precisión de 34,7791% para RCARF, superando a los demás (Suárez-Cetrulo, Cervantes, and Quintana 2019). Finalmente, en 2019, utilizan el modelo hibrido de representación lineal por partes con máquina de vectores de soporte ponderado (PLR-WSVM), apoyado en el indicador técnico bursátil RSI para determinar si los puntos de inflexión en la serie tiempo son de compra o venta en 20 activos financieros extraídos de english.sse.com.cn. El modelo propuesto obtuvo un 43,74% de rentabilidad superando a los demás modelos comparados (PLR– WSVM, IPLR–WSVM , PLR–ANN, BHS) (Tang, Dong, and Shi 2019).

Año	Referencia	Técnica	Técnicas	Datos | fuente	Criterio de Calidad							Puntaje
					C1	C2	C3	C4	C5	C6	C7	
2015	(Q. Huang et al. 2015)	ML	Biclustering, k-nn	S&P500 | finance.yahoo.com	1	1	1	1	1	0,5	0,5	6,0
2016	(Rosas-Romero et al. 2016)	ML	k-means, SVD	AA, AMAT | finance.yahoo.com	1	1	0,5	1	0,5	0,5	1	5,5
2016	(Hsu et al. 2016)	ML	SVM, ANN	S&P 500, SP/TSX, Nikkei 225, KOSPI 200, Hang Seng, traits Times Index, Shanghai SE, FTSE Bursa Malaysia KLCI, MAI security, JKSE, CAC 40, FTSE 100, FTSE MIB, DAX, BUX, FTSE/JSE top 40, ISE-100, Swiss Market Index, IBEX 35, AEX, BEL20, PSI-20, OMX ALLSHARE Stockholm, OSE All Share Index, OMX Copenhagen, OMXH25, ATX, prague stock exchange index, OMX Vilnius Index, OMX Tallinn Index, OMX Riga Index, INDEXDJX, NASDAQ-100, Brazilian Bovespa Futures | tickdata.com	0,5	1	1	1	0,5	0,5	1	5,5
2017	(Das and Padhy 2017)	ML	US-ELM, SVR	MCX| mcxindia.com	1	1	1	1	0,5	0,5	1	6,0
2017	(Kim 2017)	ML	MKK, MKK – prior MKK–MACD	S&P 500, Amazon, Apple, Bank of America, Citi Group, Coca Cola, Disney, Ford, GE, Microsoft, Nike | finance.yahoo.com	1	1	1	1	0,5	0,5	1	6,0

Año	Referencia	Técnica	Técnicas	Datos | fuente	Criterio de Calidad							Puntaje
					C1	C2	C3	C4	C5	C6	C7	
2017	(Nair et al. 2017)	ML	SOM Clustering	ACC, GAIL, ICICI, NTPC, TE, BoA, BA, GE, M3, PEP, BT, CCL, GSK, RDS, SBRY, DTEX, EMBR, ITSA4, PETR, TFA | finance.yahoo.com	0,5	1	1	0,5	1	1	0,5	5,5
2017	(Hong et al. 2017)	ML	SVM, k-nn, DT	Brent crude, GN, FTSE100, DAX30, CAC40, S&P500 | Investing.com, finance.yahoo.com, bloomberg.com	0,5	1	1	1	0,5	0,5	0,5	5,0
2017	(Hu et al. 2017)	ML	SVR, PSO	S&P 500, FTSE 100, Nikkei 225 | finance.yahoo.com	0,5	1	1	1	1	1	1	6,5
2018	(Thakur and Kumar 2018)	ML, MH	WM, SVM, GEP,RS, BUY-HOLD, RF-PSVM, RF-BMPSVM, RF-LSTSVM, RF-WMGEPSVM	NASDAQ, DOWJONES, S&P 500, NIFTY 50 y NIFTY BANK | quandl.com	1	1	1	0,5	0,5	1	0,5	5,5
2018	(Fiévet and Sornette 2018)	ML	B&H, FCT, FRT, AR, RTMSE, CTGini	S&P 500 | thomsonreuters.com	0,5	1	0,5	0,5	0,5	0,5	0,5	4,0
2018	(Chen and Hao 2018)	ML	PCA-WSVM, WSVM, PCA-ANN, BHS	002617.SZ, 002314.SZ, 002149.SZ | finance.yahoo.com	0,5	1	1	1	0,5	0,5	1	5,5
2019	(Risse 2019)	ML	OLS, OLS EW, LASSO, BOOST, FOREST, SVR, DWT-SVR, Buy & hold	Gold, cpi, twex, tbod, tms, sp500,gsci, crb, IMF-IFS foreign exchange reserves Adjusted realized volatility of gold | fred.stlouisfed.org	0,5	0,5	1	1	1	0,5	0,5	5,0
2019	(Lohrmann and Luukka 2019)	ML	RF	S&P500 | finance.yahoo.com	0,5	0,5	1	1	0,5	1	0,5	5,0
2019	(Suárez-Cetrulo et al. 2019)	ML	RCARF, ARF, RCD, DWM, AHOEFT	SPDR, S&P 500 | quantquote.com	1	1	0,5	0,5	0,5	0,5	1	5,0
2019	(Tang et al. 2019)	ML	PLR– WSVM, IPLR–WSVM , PLR–ANN, BHS	600736, 600197, 600211, 600694, 600351, 600488, 600054, 600019, 600058, 600682, 600687, 600066, 600881, 600228, 600697, 600107, 600053, 600051, 600163, 600167 | english.sse.com.cn	0,5	1	1	0,5	0,5	0,5	0,5	4,5

Tabla 4 – Resultados de la evaluación de los estudios basados en aprendizaje de maquina

3.2. Contribuciones basadas en metaheurísticas

En 2015, proponen una estrategia universal de negociación utilizando espacios de Hilbert con kernel reproducible (RKHSs) para predecir datos históricos (de los activos financieros: SIBN, GAZP, LKOH, MTSI, ROSN, SBER) en la bolsa de valores de Rusia y la bolsa de valores electrónica BATS (Bats Global Markets) de US, obtenidos de www.finam.ru. La técnica fue afinada con los algoritmos de Foster-Vohra y Kakade–Foster. La propuesta obtuvo una rentabilidad considerables ante otras técnicas de baja complejidad

sin mencionarlas (V'yugin and Trunov 2015). Este mismo año 2015, se identifican estudios en computación inteligente aplicados a la construcción de operaciones comerciales con pares de activos financieros y como propuesta de investigación utilizan un enfoque general para más de dos acciones como grupo de inversión, empleando algoritmos genéticos (Genetic Algorithms, GA) que optimizan la media móvil, Bandas de Brolliger y un coeficiente de ponderación del stock. La anterior propuesta usa datos de la bolsa de Taiwan obteniendo una precisión en las predicciones superiores al 0,7 con las mayores capitalizaciones del mercado, superando al modelo de regresión lineal (C. F. Huang et al. 2015).

En 2016, utilizan máquinas de soporte vectorial evolutivo (ESVM) para predecir la dirección de los índices bursátiles ASE20, Dax30, Nikkei225 y S&P 500 extrayendo históricos de bloomberg.com donde se comparan con estrategias comerciales Buy&Hold, Naive Bayes, el indicador técnico MACD, modelo auto regresivo ARMA y un perceptrón multicapa (MLP). Como resultado el modelo hibrido ESVM obtuvo un 70,29% de rentabilidad anual excluida del costo, seguida de la estrategia MPL que fue de 42,19% (Karathanasopoulos et al. 2016). En el mismo año (2016), se propone un modelo hibrido basado en series de tiempo difusas de múltiples órdenes y un análisis técnico con un algoritmo genético, todo desarrollado en C++ para predecir series de tiempo de índices bursátiles (TAIEX, Dow Jones, NASDAQ, HSI, S&P500) y el par de divisas NTD/USD (dólar taiwanés/dólar) con datos históricos extraídos de twse.com.tw/zh/, cbc.gov.tw. La propuesta es superior en la tasa de precisión direccional entre un 1% al 3% respecto a las series de tiempo difusas con el algoritmo genético (FTSGA) y con un algoritmo de colonia de hormigas con regresión automática (ACO-AR) (Ye et al. 2016).

En el año 2017, se estudió la fluctuación dinámica del movimiento (la correlación entre volatilidades pasadas y rendimientos futuros) de precios en 20 índices bursátiles (MERV, S&P500, AXJO, BFX, BVSP, GSPTSE, IPSA, SCI, FTSE, FCHI, DAX, HIS, BSESN, JKSE, N225, KOSPI, KLSE, MXX, NZ50, TWII) con datos de Yahoo! Finance para ser procesados en MatlabR2012a y obtener una estrategia de efecto impulsado por la fluctuación (FDE). La propuesta FDE obtuvo un retorno acumulado de un 70% mientras tanto las propuestas basadas en indicadores técnicos bursátiles (RSI, MACD, Momentum) y la estrategia B&H obtuvieron como máximo un 40% (Li et al. 2017). Continuando con el año 2017, proponen un modelo PAA-MS-IDPSO-V (aproximación agregada por partes usando optimización auto adaptativa por Multi enjambre mejorada de partículas con validación) para obtener patrones en las series de tiempo del índice bursátil S&P100 extraídas de Yahoo! Finance. La propuesta PAA-MS-IDPSO-V obtiene mejores resultados que las variaciones de la propuesta PAA-IDPSO, PAA-IDPSO-V y PAA-MS-IDPSO en los retornos de inversión según la prueba no paramétrica de Friedman y la prueba post-hoc de Nemenyi con un nivel de confianza del 95% (Brasileiro, Souza, and Oliveira 2017). Finalizando el 2017 proponen una herramienta en java para el pronóstico y la negociación de activos con programación genética usando los datos extraídos de la bolsa de valores de Londres, Nueva York y Frankfurt. La propuesta supera en precisión de pronóstico y utilidad a los modelos basados en una caminata aleatoria (Random Walks o Random Search, RS), una comparación poco informativa de la calidad de la propuesta (Karathanasopoulos 2017).

En el 2018, proponen negociar futuros del índice KOSPI200 con un sistema de comercio de coincidencias de patrones (PMTS) que utiliza el algoritmo envolvente de tiempo dinámico (DTW) en las series de tiempo extraídas de data.koscom.co.kr. La propuesta que obtuvo la mejor relación del indicador técnico bursátil Sharpe con 0,94 según los parámetros de costo de deslizamiento de 0.02 pt, una relación de stop-loss del 0.5%, un período de entrenamiento de 18 meses, un período de prueba de 3 meses, 20 criterios de filtrado ea y 65% de frecuencia U/D (Kim et al. 2018). En ese mismo año, 2018, se utilizó programación evolutiva para predecir series de tiempo en activos financieros (FTSE MIB, PSI-20, ISEQ 20, FTSE/Athens Large Cap, IBEX 35) obtenidos de borsaitaliana. it, bolsamadrid.es, bolsadelisboa.com, ise.ie y athexgroup.gr. Para el algoritmo evolutivo se estima y ajusta los parámetros con modelos STAR (ESTAR, LSTAR, TSTAR, GBELL-STAR) y la métrica Diebold-Mariano. Como resultado quien obtuvo la mejor relación del indicador técnico bursátil sharpe ratio fue LSTAR con un valor de 0.72410 dado en el índice PSI-20 superando a los demás modelos (Avdoulas, Bekiros, and Boubaker 2018). Continuando con el 2018, proponen un sistema que utiliza programación genética, optimización multiobjetivo, análisis técnico y una sección donde determina cuando comprar o vender el activo financiero, para series temporales de BBAS3, BOVA11, CMIG4, EMBR3, GBR4 y VALE5 obtenidas de bb.com.br, bmfbovespa.com.br, cemig. com.br, embraer.com, gerdau.com y vale.com. La propuesta obtuvo hasta un 77,61% en retornos en comparación a la segunda mejor estrategia B&H con un 46,57% en retornos, seguidas de las estrategias TA-10 y TA-36 (Pimenta et al. 2018). Siguiendo en 2018, utilizan el algoritmo de biclustering para encontrar patrones de negociación con los indicadores técnicos bursátiles (SMA, RSI, %R, ROC, CCI, EMV, UO, BRAR, KJD y MTM) de activos financieros (Apple, BA, CAT, JNJ, XOM, VZ y S&P500) obtenidos de la bolsa de valores de EE. UU, posteriormente aplican el algoritmo Naïve Bayes para determinan la tendencia de las series de tiempo y por último se optimiza la precisión de la tendencia con el algoritmo Adaboost. La propuesta fue desarrollada en Qt Creator obteniendo una tasa de retorno futuro del 51,08% como máximo, seguido del modelo TS-Fuzzy con un 45,16% y los demás algoritmos comparados obtienen porcentajes inferiores PLR-SVR, PLR-BPN, BIC-K-NN,BIC-NV-AD (Huang et al. 2018). Además, en 2018, se utilizó programación evolutiva para generar estrategias de negociación cambiando la ventana de tiempo, pero conservando la población o inicializando la población de los genotipos, para el par divisa EUR/USD (euro/dólar) del mercado Forex. La propuesta fue implementada en Java SE6 y ser comparada con la estrategia B&H (Drezewski, Dziuban, and Pajak 2018). Para terminar el 2018, se proponen el algoritmo de campeonato de liga (LCA) con la estructura de red equipada con backward Q-learning (BQL), para extraer y guardar reglas de negociación óptimas para 20 activos financieros (MKBT, BDAN, ABDI, KSIM, MARK, ARDK, KSHJ, ROOI, BARZ, IKCO, INFO, MAPN, NSTH, STEH, NAFT, PARK, IRNZ, OIMC, TSHE, BTEJ) de la bolsa de valores de Teherán (tse.ir). La propuesta obtuvo una ganancia máxima en un activo financiero de 460,04% seguida de la estrategia B&H con un 441,28% y terminado en pérdidas en todos los activos con la programación de redes genéticas con acumulación de reglas (GBP-RA) (Alimoradi and Husseinzadeh Kashan 2018).

Año	Referencia	Técnica	Técnicas	Datos | fuente	Criterio de Calidad							Puntaje
					C1	C2	C3	C4	C5	C6	C7	
2015	(V'yugin and Trunov 2015)	MH	RKHS	SIBN, GAZP, LKOH, MTSI, ROSN, SBER | finam.ru	1	0,5	0,5	0,5	0,5	1	1	5,0
2015	(C. F. Huang et al. 2015)	MH	GA	TPE: 2317, 2454, 1301, 1303, 1326, 2412, 2882, 2308 y 2008 | twse.com.tw/zh/	1	1	1	1	0,5	0,5	0,5	5,5
2016	(Karathanasopoulos et al. 2016)	MH	NAIVE, MACD, ARMA, MLP, ESVM	ASE20, DAX30, NIKKEI225, FTSE100, S&P500 | bloomberg.com	0,5	1	1	0,5	1	1	1	6,0
2016	(Ye et al. 2016)	MH	FTSGA, ACO-AR	TAIEX, Dow Jones, NASDAQ, HSI, SP500, NTD / USD | twse.com.tw/zh/, cbc.gov.tw	1	1	1	0,5	0,5	0,5	0,5	5,0
2017	(Li et al. 2017)	MH	FDE, RSI, MACD	MERV, S&P500, AXJO, BFX, BVSP, GSPTSE, IPSA, SCI, FTSE, FCHI, DAX, HIS, BSESN, JKSE, N225, KOSPI, KLSE, MXX, NZ50, TWII| finance.yahoo.com	1	1	0,5	0,5	0,5	1	0,5	5,0
2017	(Brasileiro et al. 2017)	MH	PAA-IDPSO, PAA-IDPSO-V, PAA-MS-IDPSO, PAA-MS-IDPSO-V	S&P100 | finance.yahoo.com	0,5	1	1	1	0,5	1	1	6,0
2017	(Karathanasopoulos 2017)	MH	GEP, RS, RW, MACD, ARMA, MLP, RNN, HONN, GP, GEP Trader Tool	FTSE100, DAX30, S&P500 | londonstockexchange.com, nyse.com , deutsche-boerse.com	1	1	0,5	0,5	0,5	0,5	0,5	4,5
2018	(Kim et al. 2018)	MH	WTS	KOSPI 200| data.koscom.co.kr	0,5	1	1	0,5	0,5	0,5	1	5,0
2018	(Avdoulas et al. 2018)	MH	ESTAR, LSTAR, TSTAR, GBELL-STAR	FTSE MIB, PSI-20, ISEQ 20, FTSE/Athens Large Cap, IBEX 35 | borsaitaliana.it, bolsamadrid.es, bolsadelisboa.com, ise.ie, athexgroup.gr	0,5	0,5	0,5	0,5	0	0	0,5	2,5
2018	(Pimenta et al. 2018)	MH	GA-TA-B&H, TA-10, TA-36	BBAS3, BOVA11, CMIG4, EMBR3, GBR4, VALE5, | bb.com.br, bmfbovespa.com.br, cemig.com.br, embraer.com, gerdau.com y vale.com	1	1	0,5	0,5	0,5	0	0,5	4,0

Año	Referencia	Técnica	Técnicas	Datos | fuente	Criterio de Calidad							Puntaje
					C1	C2	C3	C4	C5	C6	C7	
2018	(Huang et al. 2018)	MH	TS fuzzy, PLR-SVR, PLR-BPN, BIC-K-NN,BIC-NV-AD	Apple, BA, CAT, JNJ, XOM, VZ, S&P500 | bolsa de valores de EE.UU	1	1	0,5	1	0,5	0,5	1	5,5
2018	(Drezewski et al. 2018)	MH	AE, Buy & Hold	Forex | No reporta	0,5	1	1	1	0,5	0,5	1	5,5
2018	(Alimoradi and Husseinzadeh Kashan 2018)	MH	LCA-NBQL, GNP-RA, Buy & Hold	MKBT, BDAN, ABDI, KSIM, MARK, ARDK, KSHJ, ROOI, BARZ, IKCO, INFO, MAPN, NSTH, STEH, NAFT, PARK, IRNZ, OIMC, TSHE, BTEJ | tse.ir	1	1	0,5	0,5	1	0,5	1	5,5

Tabla 5 – Resultados de la evaluación de los estudios basados en metaheurísticas

3.3. Contribuciones basadas en redes neuronales artificiales

En 2015 se presentó un análisis de 113 artículos que datan de 1933 a 2013, donde se identificaron las técnicas más comúnmente utilizadas en el ámbito financiero y la inteligencia artificial (IA) para el pronóstico de los mercados. Como resultado del análisis, se observó un creciente usó de las técnicas de IA en los últimos años. Con base en esta revisión, se propuso una nueva técnica que usa filtrado Hodrick-Presscott, una transformada discreta de Fourier y un Perceptron multicapa (Multi-Layer Perceptron, MLP) con una capa oculta entrenada con el algoritmo de Levenberg-Marquardt. En los resultados de la comparación, se definió una propuesta combinada entre MLP y árboles de decisión que supera la predicción de las series temporales basado en exactitud a MLP con ANFIS usando los datos disponibles en Yahoo! Finance para los índices bursátiles S&P500, FTSE-100 y BSE-Sensex (Nair and Mohandas 2015).

En 2016, se usó PSO para definir los pesos iniciales de una red MLP implementada en lenguaje C, para pronosticar el movimiento de los precios de índices financieros tomando los datos históricos de Yahoo! Finance e indicadores técnicos bursátiles. La propuesta obtuvo el 154.80% de rendimiento promedio. Los autores observaron que una mayor precisión de predicción no garantiza un mayor rendimiento de los retornos (Chiang et al. 2016). También en 2016, utilizan un modelo hibrido de PCA con transformadas de wavelet y redes neuronales artificiales (WPCA-NN), para predecir los precios de los índices bursátiles NIKKEI-225, SiMSCI, KOSPI-200 y TAIEX de acuerdo con los indicadores técnicos bursátiles RSI, MACD, MACDsig, MACDHis, Stochastics (fast %K, slow %K y %D), y Ultimate Oscillator (UO) extraídos de Bloomberg. El modelo WPCA-NN con red neuronal auto regresiva no lineal con entradas exógenas (NARX-NN) entrenada con el algoritmo Levenberg-Marquardt obtiene mayor rentabilidad que los algoritmos W-NARX-NN o NARX-NN. Todos los algoritmos se evaluaron en las mismas condiciones y fueron implementados en Maltab (M'Ng and Mehralizadeh 2016).

En 2017, se estudió la predicción de la tendencia del índice bursátil KOSPI200 con una red MLP entrenado con el algoritmo back-propagation, un modelo SVM y una red neuronal con función de base radial (RBF). Estos modelos se implementaron con Matlab utilizando los datos históricos del índice obtenidos del portal web https://global.krx.co.kr/contents/GLB con 10 indicadores financieros adicionales. Este trabajo logró la evaluación y comparación de los modelos MLP, SVM y RBF en un framework significativamente realista que obtuvo una precisión de las predicciones del 75.74%, 71.52% y 62.23%, respectivamente (Pyo et al. 2017).

En 2018 propusieron replicar una estrategia de trading utilizando máquinas de memoria a corto plazo (LSTM). Donde implementan una CNN para mapear datos sin procesar y eventualmente incrustar correlaciones parciales que después ingresan estos datos con indicadores técnicos a la LSTM. Esta técnica fue desarrollada en Python con la API Keras y Tensorflow. Se proporcionó un conjunto de datos de 30 empresas que conforman el índice bursátil Dow Jones. La red implementada fue entrenada 3.000 épocas en el caso de elección aleatoria (RC), y 15.000 épocas para el Crossover y Media móvil Convergencia/Divergencia MACD y también una tasa de abandono de 0.25 para un posible sobreajuste. Los resultados indicaron que este enfoque es viable, pero se observaron que hubo un deterioro en el rendimiento al aumentar datos de entrada como los índices técnicos. Además, esta arquitectura de LSTM al dedicarle subconjuntos de indicadores combinado con capas adicionales aumenta el rendimiento de la red (Troiano, Villa, and Loia 2018). Ese mismo año (2018), propone una novedosa técnica para el trading algorítmico basado en redes neuronales convolucionales profundas (Deep convolutional neural networks, DCNN). El método propuesto, inicia convirtiendo los datos de series temporales unidimensionales a imágenes bidimensionales de 15x15 que fueron etiquetados como vender, comprar y retener, según los valles o colinas de las series de tiempo. La DCNN utiliza nueve capas para su entrenamiento incluyendo convolucionales, maxpooling, drop-out y una capa densa (completamente conectada) al final. Los datos usados para el entrenamiento de la DCNN fueron tomados del índice bursátil Dow-30 y fondos de inversión cotizados (ETFs) disponibles en Yahoo! Finance, los cuales usan 15 indicadores técnicos bursátiles. Como resultado de este trabajo, los valores de precisión total para el índice bursátil Dow30 fueron de 0,58 y para ETF de 0,62. Por último, la evaluación financiera de la propuesta fue comparada contra los métodos "Buy&Hold", (BaH), RSI (14 días, 70–30), SMA (50 días), LSTM y MLP, obteniendo mejores resultados en los dos datasets, que incrementan desde un 1,5% hasta un 9% de acuerdo al método con el que se compara (Sezer and Ozbayoglu 2018).

En 2019 proponen un modelo denominado red neuronal de filtros múltiples (Multi-Filters Neural Network, MFNN) para predecir movimientos del índice bursátil CSI 300. Toma como entrada 120 indicadores técnicos bursátiles, los procesa a través de un módulo de neuronas convolucionales y recurrentes para obtener un mapa de características. La propuesta obtuvo una precisión de predicción promedio del 49,53%, superando a otras 7 técnicas (RNN, CNN, LSTM, SVM, Logistic Regression, Random Forest y Linear Regression) más y en la prueba de rangos con signos de Wilcoxon supera a los demás en el rendimiento. Como resultado la técnica más estable debido al riesgo y rentabilidad fue una red neuronal recurrente (Recurrent Neural Network, RNN) con un indicador

bursátil Sharpe de 6.42 y con una ganancia del 67.43%, y MFNN obtuvo el segundo mejor resultado (Long, Lu, and Cui 2019). Ese mismo año (2019), proponen un algoritmo en Python de memoria de corto plazo (Long-Short Term Memory, LSTM) que utiliza Keras y TensorFlow para predecir el precio de cierre de acciones en un determinado tiempo, con datos obtenidos de Yahoo! Finance desde octubre del año 2000 a octubre de 2018. La primera red pronostica la tendencia del stock con una capa de entrada que utiliza datos del cierre de acciones, una capa oculta de 400 neuronas y con una neurona en la salida, la segunda red realiza la regresión de series temporales para predecir el precio de cierre apoyado con un modelo estadístico de Markov. La precisión del método propuesto tiene un rango del 50% al 99% (Rundo et al. 2019). También en 2019, identificaron la existencia de pocos estudios para pronosticar retornos diarios del mercado financiero con los índices bursátiles S&P500, SPDR y ETF, utilizando redes neuronales profundas (Deep Neural Network, DNN) con backpropagation. El modelo propuesto combina la DNN con el algoritmo de análisis de componentes principales que toma hasta 60 parámetros financieros obtenidos de Yahoo! Finance, investing.com y fred.stlouisfed.org, alcanzando una precisión de pronóstico del 60,4%, respecto al 57,5% de una red neuronal feed-forward tradicional (ANN) y una relación ANN/DNN del Sharpe ratio de 0,0778 (Zhong and Enke 2019). Además, en 2019, se propone predecir la cantidad de acciones a negociar combinando aprendizaje por refuerzo con tres modelos de DNN continuos (regresores DNN) que luego se usan para tomar la decisión de comprar, vender o retener activos financieros (S&P500, KOSPI, HSI y EuroStoxx50) que utiliza datos de Thomsonreuters y Yahoo! Finance. Como resultado de la investigación el modelo aumento el beneficio comercial total desde cuatro veces en el índice bursátil S&P500 hasta doce veces en el índice bursátil HSI con respecto a sistemas de comercio fijo y también mejoran el modelo a la adaptabilidad del mercado real (Jeong and Kim 2018). Finalmente, en 2019, se propone un modelo hibrido con la red neuronal basada en máquina de descomposición y factorización empírica (EMD2FNN) para pronosticar lo precios de cierre en los activos financieros (SSEC, NASDAQ y S&P 500) con datos extraídos de Yahoo! Finance. Los modelos EMD2FNN simple y EMD2FNN múltiple obtuvieron un rendimiento promedio anual del 13% y el 25%, respectivamente (Zhou et al. 2019).

Año	Referencia	Técnica	Técnicas	Datos | fuente	Criterio de Calidad							Puntaje
					C1	C2	C3	C4	C5	C6	C7	
2015	(Nair and Mohandas 2015)	ANN	MLP, Fourier, Hodrick-Presscott	BSE-Sensex, FTSE-100, S&P 500 | finance.yahoo.com	1	1	1	1	1	1	1	7,0
2016	(Chiang et al. 2016)	ANN	PSO, RNA	IXIC, SPY, QQQ, MDY, IWM, FXI, EFA, EEM, HSI, TWII, BSESN, XLF, XLY, XLE, XLV, XLK, SMH, OIH, FTSE, ATX, FCHI, AEX | finance.yahoo.com	1	1	1	0,5	0,5	1	1	6,0
2016	(M'Ng and Mehralizadeh 2016)	MH ANN	WPCA-NN, WNN, NN, B&H	Hang Seng, NIKKEI 225, MSCI, KOSPI 200, TAIEX | bloomberg.com	1	1	1	1	0,5	0,5	1	6,0
2017	(Pyo et al. 2017)	MH, ML ANN	ANN, RBF, SVM	Kospi200 | global.krx.co.kr/contents/GLB	0,5	1	1	1	1	1	1	6,5

Año	Referencia	Técnica	Técnicas	Datos | fuente	Criterio de Calidad							Puntaje
					C1	C2	C3	C4	C5	C6	C7	
2018	(Troiano et al. 2018)	ANN	LSTM, CNN	AAPL, AXP, BA, CAT, CSCO, CVX, DD, DIS, GE, GS, HD, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, NKE, PFE, PG, TRV, UNH, UTX, V, VZ, WMT, XOM | No reporta	1	1	1	0,5	0,5	0	0,5	4,5
2018	(Sezer and Ozbayoglu 2018)	ANN	DCNN, CNNTAr, BaH, RSI, SMA, LSTM, MLP	Dow-30, ETFs: XLF, XLU, QQQ, SPY, XLP, EWZ, EWH, XLY, XLE | finance.yahoo.com	1	1	1	1	1	0,5	1	6,5
2019	(Long et al. 2019)	ML ANN	MFNN, RNN, CNN, LSTM, SVM, Logistic Regression, Random Forest, Linear Regression	CSI300 | no reporta	0,5	1	1	0,5	0,5	0,5	1	5,0
2019	(Rundo et al. 2019)	ML ANN	LSTM Markov,	ENEL.MI, UCG.MI, BMPS.MI, CVAL.MI, ISP.MI | finance.yahoo.com	1	0,5	1	1	0,5	1	0,5	5,5
2019	(Zhong and Enke 2019)	ANN	ANN(MLP), DNN, PCA	SPDR, S&P500, ETF | finance.yahoo.com, investing.com, fred.stlouisfed.org	1	1	1	1	0,5	0,5	1	6,0
2019	(Jeong and Kim 2018)	ANN	DNN, DQN	S&P500, KOSPI, HSI, EuroStoxx50 | thomsonreuters.com, finance.yahoo.com	1	1	0,5	0,5	0,5	0,5	1	5,0
2019	(Zhou et al. 2019)	ANN	EMD2FNN, FNN, NN(MLP)	SSEC, NASDAQ, S&P 500 | finance.yahoo.com	1	0,5	0,5	0,5	0,5	0,5	1	4,5

Tabla 6 – Resultados de la evaluación de los estudios basados en redes neuronales artificiales

3.4. Contribuciones basadas en lógica difusa

En 2015, se recuperó información de profesionales expertos en análisis técnico con más de 15 años de experiencia, mediante cuestionarios para el sistema de inferencia difuso tipo Mandani, obteniendo como resultado que indicadores utilizar, y con la información se extraen reglas de negociación con el enfoque de conjunto aproximado basado en el dominio de la consistencia variable (Variable Consistency Dominance-based Rough Set Approach, VC-DRSA). El VC-DRSA utiliza el software jMAF y alcanzó un 76% de precisión en clasificación, superando al DRSA que también utiliza jMAF y al análisis discriminatorio que emplea una herramienta estadística de Matlab con los datos extraídos de sysjust.com.tw (Shen and Tzeng 2015).

En 2016 se implementó un sistema experto para el mercado Forex con un enfoque de razonamiento probatorio basado en reglas (Rule-Based Evidential Reasoning, RBER) para tomar decisiones de comprar o vender, basado en análisis de gráficos de velas mediante indicadores técnicos adaptados con lógica difusa. Esta implementación se adaptó a la plataforma MetaTrader4 y la evaluación mostró una rentabilidad mínima del 45.77% (Dymova, Sevastjanov, and Kaczmarek 2016). También en 2016, implementaron un sistema para predecir puntos de inflexión utilizando datos históricos de la bolsa de valores de Estados Unidos con el modelo difuso Takagi-Sugeno-Kan (TSK) a partir de

indicadores técnicos y señales de negociación del modelo SVR. El resultado de esta investigación indica que la propuesta supera a otros modelos como PLR-SVR y PLR-BPN hasta en un 6% en la tasa de beneficio (Chang, Wu, and Lin 2016).

En 2018 se implementó un modelo de red Petri difusa de alto nivel (High-Level Fuzzy Petri Net, HLFPN) en C# con ayuda de máquinas de soporte vectorial que utilizan LibSVM (Library for Support Vector Machines) para tomar decisiones de compra o venta de los activos financieros TWSE, Wei Chuan, Formosa Plastic, China Steel, UMC, TSMC, Chunghwa Telecom, Grand Formosa Regent Taipei y Mega Bank Public Co. Ltd obtenidas de twse.com.tw/en. Como resultado de la investigación obtuvieron una predicción total del 64,05% y con el retorno anual promedio del 4,80% con la propuesta HLFPN siendo mejor que SVM, Neural Fuzzy, Technical Indicator Strategies y Trend Following Method, (Shen et al. 2018).

Año	Referencia	Técnica	Técnicas	Datos | fuente	Criterio de Calidad							Puntaje
					C1	C2	C3	C4	C5	C6	C7	
2015	(Shen and Tzeng 2015)	FL	FIS-Mandini, VC-DRSA	TAIEX | sysjust.com.tw	1	1	1	1	0,5	0,5	0,5	5,5
2016	(Dymova et al. 2016)	FL	RBER	Forex | metatrader4.com	1	0,5	0,5	0,5	0,5	0,5	0,5	4,0
2016	(Chang et al. 2016)	FL	TSK, SVR	Bolsa US | No reporta	0,5	1	1	0,5	0,5	0,5	1	5,0
2018	(Shen et al. 2018)	FL	HLFPN, SVM, Neural Fuzzy, Technical Indicator Strategies, Trend Following Method, Trend Following Method, Buy & Hold	TWSE, Wei Chuan, Formosa Plastic, China Steel, UMC, TSMC, Chunghwa Telecom, Grand Formosa Regent Taipei, and Mega Bank Public Co. Ltd | twse.com.tw/en	1	1	1	1	1	0,5	1	6,5

Tabla 7 – Resultados de las evaluaciones basados en lógica difusa

4. Conclusiones recomendaciones y trabajo futuro

En este documento se identifican las técnicas, métodos, modelos o similares, utilizados en el trading algorítmico para la predicción de series de tiempo financieras. Estos trabajos centran sus contribuciones principalmente en aprendizaje de máquina, metaheurísticas, redes neuronales artificiales y lógica difusa, observándose para los últimos años, una mayor actividad investigativa en el aprendizaje de máquina seguido de las redes neuronales artificiales. Las investigaciones utilizan principalmente datos de Yahoo! Finance debido a su fácil acceso respecto a otras fuentes expuestas tomando en cuento la ventana de tiempo de investigación que fue desde 2015 a 2019. Una de las técnicas más innovadoras en redes neuronales fue la de convertir series temporales en imágenes para ser procesadas por redes neuronales convolucionales obteniendo hasta un 62% de precisión de pronóstico. En el enfoque metaheurístico se encuentran propuestas que sobrepasan el 70% de precisión en pronóstico bajo condiciones similares, pero no iguales. Debido a esto es difícil definir cuáles son las mejores propuestas y en que contextos obtienen esos mejores resultados, ya que entre ellas no se comparan, razón por la cual se necesita un framework de evaluación y comparación de algoritmos de trading algoritmo de activos financieros.

Agradecimientos

A la Universidad del Cauca por financiar parcialmente la presente investigación.

Referencias

Alimoradi, Muhammad Reza and Ali Husseinzadeh Kashan. 2018. "A League Championship Algorithm Equipped with Network Structure and Backward Q-Learning for Extracting Stock Trading Rules." *Applied Soft Computing Journal* 68:478–93.

Avdoulas, Christos, Stelios Bekiros, and Sabri Boubaker. 2018. "Evolutionary-Based Return Forecasting with Nonlinear STAR Models: Evidence from the Eurozone Peripheral Stock Markets." *Annals of Operations Research* 262(2):307–33.

Brasileiro, Rodrigo C., Victor L. F. Souza, and Adriano L. I. Oliveira. 2017. "Automatic Trading Method Based on Piecewise Aggregate Approximation and Multi-Swarm of Improved Self-Adaptive Particle Swarm Optimization with Validation." *Decision Support Systems* 104:79–91.

Chang, Pei Chann, Jheng Long Wu, and Jyun Jie Lin. 2016. "A Takagi-Sugeno Fuzzy Model Combined with a Support Vector Regression for Stock Trading Forecasting." *Applied Soft Computing Journal* 38:831–42.

Chen, Yingjun and Yijie Hao. 2018. "Integrating Principle Component Analysis and Weighted Support Vector Machine for Stock Trading Signals Prediction." *Neurocomputing* 321:381–402.

Chiang, Wen Chyuan, David Enke, Tong Wu, and Renzhong Wang. 2016. "An Adaptive Stock Index Trading Decision Support System." *Expert Systems with Applications* 59:195–207.

Das, Shom Prasad and Sudarsan Padhy. 2017. "Unsupervised Extreme Learning Machine and Support Vector Regression Hybrid Model for Predicting Energy Commodity Futures Index." *Memetic Computing* 9(4):333–46.

Drezewski, Rafal, Grzegorz Dziuban, and Karol Pajak. 2018. "The Bio-Inspired Optimization of Trading Strategies and Its Impact on the Efficient Market Hypothesis and Sustainable Development Strategies." *Sustainability (Switzerland)* 10(5).

Dymova, Ludmila, Pavel Sevastjanov, and Krzysztof Kaczmarek. 2016. "A Forex Trading Expert System Based on a New Approach to the Rule-Base Evidential Reasoning." *Expert Systems with Applications* 51:1–13.

Fiévet, L. and D. Sornette. 2018. "Decision Trees Unearth Return Sign Predictability in the S&P 500." *Quantitative Finance* 18(11):1797–1814.

Hong, Ki Hoon, Hojin Jung, and Minjae Park. 2017. "Predicting European Carbon Emission Price Movements." *Carbon Management* 8(1):33–44.

Hsu, Ming Wei, Stefan Lessmann, Ming Chien Sung, Tiejun Ma, and Johnnie E. V. Johnson. 2016. "Bridging the Divide in Financial Market Forecasting: Machine Learners vs. Financial Economists." *Expert Systems with Applications* 61:215–34.

Hu, Zhongyi, Yukun Bao, Raymond Chiong, and Tao Xiong. 2017. "Profit Guided or Statistical Error Guided? A Study of Stock Index Forecasting Using Support Vector Regression." *Journal of Systems Science and Complexity* 30(6):1425–42.

Huang, Chien Feng, Chi Jen Hsu, Chi Chung Chen, Bao Rong Chang, and Chen An Li. 2015. "An Intelligent Model for Pairs Trading Using Genetic Algorithms." *Computational Intelligence and Neuroscience* 2015.

Huang, Qinghua, Zhoufan Kong, Yanshan Li, Jie Yang, and Xuelong Li. 2018. "Discovery of Trading Points Based on Bayesian Modeling of Trading Rules." *World Wide Web* 21(6):1473–90.

Huang, Qinghua, Ting Wang, Dacheng Tao, and Xuelong Li. 2015. "Biclustering Learning of Trading Rules." *IEEE Transactions on Cybernetics* 45(10):2287–98.

Jeong, Gyeeun and Ha Young Kim. 2018. "Improving Financial Trading Decisions Using Deep Q-Learning: Predicting the Number of Shares, Action Strategies, and Transfer Learning." *Expert Systems With Applications* 117:125–38.

Karathanasopoulos, Andreas. 2017. "Modelling and Trading the London, New York and Frankfurt Stock Exchanges with a New Gene Expression Programming Trader Tool." *Intelligent Systems in Accounting, Finance and Management* 24(1):3–11.

Karathanasopoulos, Andreas, Konstantinos Athanasios Theofilatos, Georgios Sermpinis, Christian Dunis, Sovan Mitra, and Charalampos Stasinakis. 2016. "Stock Market Prediction Using Evolutionary Support Vector Machines: An Application to the ASE20 Index." *European Journal of Finance* 22(12):1145–63.

Kim, Heejean. 2017. "A Prediction Methodology for the Change of the Values of Financial Products." 51(3):197–210.

Kim, Sang Hyuk, Hee Soo Lee, Han Jun Ko, Seung Hwan Jeong, Hyun Woo Byun, and Kyong Joo Oh. 2018. "Pattern Matching Trading System Based on the Dynamic Time Warping Algorithm." *Sustainability (Switzerland)* 10(12):1–18.

Kitchenham, Barbara. 2004. *Procedures for Performing Systematic Reviews*. Keele.

Li, Yan, Bo Zheng, Ting Ting Chen, and Xiong Fei Jiang. 2017. "Fluctuation-Driven Price Dynamics and Investment Strategies." *PLoS ONE* 12(12):1–15.

Lohrmann, Christoph and Pasi Luukka. 2019. "Classification of Intraday S&P500 Returns with a Random Forest." *International Journal of Forecasting* 35(1): 390–407.

Long, Wen, Zhichen Lu, and Lingxiao Cui. 2019. "Deep Learning-Based Feature Engineering for Stock Price Movement Prediction." *Knowledge-Based Systems* 164:163–73.

M'Ng, Jacinta Chan Phooi and Mohammadali Mehralizadeh. 2016. "Forecasting East Asian Indices Futures via a Novel Hybrid of Wavelet-PCA Denoising and Artificial Neural Network Models." *PLoS ONE* 11(6):1–29.

Munn, Zachary, Cindy Stern, Edoardo Aromataris, Craig Lockwood, and Zoe Jordan. 2018. "What Kind of Systematic Review Should I Conduct? A Proposed Typology and Guidance for Systematic Reviewers in the Medical and Health Sciences." *BMC Medical Research Methodology* 18(1):5.

Nair, Binoy B., P. K. Saravan. Kumar, N. R. Sakthivel, and U. Vipin. 2017. "Clustering Stock Price Time Series Data to Generate Stock Trading Recommendations: An Empirical Study." *Expert Systems with Applications* 70:20–36.

Nair, Binoy B. and V. P. Mohandas. 2015. "Artificial Intelligence Applications in Financial Forecasting-a Survey and Some Empirical Results." *Intelligent Decision Technologies* 9(2):99–140.

Pimenta, Alexandre, Ciniro A. L. Nametala, Frederico G. Guimarães, and Eduardo G. Carrano. 2018. "An Automated Investing Method for Stock Market Based on Multiobjective Genetic Programming." *Computational Economics* 52(1):125–44.

Pyo, Sujin, Jaewook Lee, Mincheol Cha, and Huisu Jang. 2017. "Predictability of Machine Learning Techniques to Forecast the Trends of Market Index Prices: Hypothesis Testing for the Korean Stock Markets." *PLoS ONE* 12(11):1–17.

Risse, Marian. 2019. "Combining Wavelet Decomposition with Machine Learning to Forecast Gold Returns." *International Journal of Forecasting* 35(2):601–15.

Rosas-Romero, Roberto, Alejandro Díaz-Torres, and Gibran Etcheverry. 2016. "Forecasting of Stock Return Prices with Sparse Representation of Financial Time Series over Redundant Dictionaries." *Expert Systems with Applications* 57:37–48.

Rundo, Francesco, Francesca Trenta, Agatino Luigi Di Stallo, and Sebastiano Battiato. 2019. "Advanced Markov-Based Machine Learning Framework for Making Adaptive Trading System." *Computation* 7(1).

Sezer, Omer Berat and Ahmet Murat Ozbayoglu. 2018. "Algorithmic Financial Trading with Deep Convolutional Neural Networks: Time Series to Image Conversion Approach." *Applied Soft Computing* 70:525–38.

Shen, Kao Yi and Gwo Hshiung Tzeng. 2015. "Fuzzy Inference-Enhanced VC-DRSA Model for Technical Analysis: Investment Decision Aid." *International Journal of Fuzzy Systems* 17(3):375–89.

Shen, Rong Kuan, Cheng Ying Yang, Victor R. L. Shen, Wei Chen Li, and Tzer Shyong Chen. 2018. "A Stock Market Prediction System Based on High-Level Fuzzy Petri Nets." *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems* 26(5):771–808.

Suárez-Cetrulo, Andrés L., Alejandro Cervantes, and David Quintana. 2019. "Incremental Market Behavior Classification in Presence of Recurring Concepts." *Entropy* 21(1):1–18.

Tang, Huimin, Peiwu Dong, and Yong Shi. 2019. "A New Approach of Integrating Piecewise Linear Representation and Weighted Support Vector Machine for Forecasting Stock Turning Points." *Applied Soft Computing Journal* 78:685–96.

Thakur, Manoj and Deepak Kumar. 2018. "A Hybrid Financial Trading Support System Using Multi-Category Classifiers and Random Forest." *Applied Soft Computing Journal* 67:337–49.

Troiano, Luigi, Elena Mejuto Villa, and Vincenzo Loia. 2018. "Replicating a Trading Strategy by Means of LSTM for Financial Industry Applications." *IEEE Transactions on Industrial Informatics* 14(7):3226–34.

V'yugin, V. V. and V. G. Trunov. 2015. "Adaptive Universal Trading Strategy." *Journal of Communications Technology and Electronics* 60(6):658–72.

Vasquez, Gustavo and Jorge Muñoz. 2015. "SANTI: Robot de Trading Con Pronósticos Basados En Inteligencia Artificial, Una Evolución Para El Mercado de Valores."

Víctor Manuel Pérez Santos. 2014. "RoboTrader: Sistema Algorítmico de Trading En Los Mercados Financieros." Universidad de La Laguna, San Cristóbal de La Laguna.

Ye, Furong, Liming Zhang, Defu Zhang, Hamido Fujita, and Zhiguo Gong. 2016. "A Novel Forecasting Method Based on Multi-Order Fuzzy Time Series and Technical Analysis." *Information Sciences* 367–368:41–57.

Zhong, Xiao and David Enke. 2019. "Predicting the Daily Return Direction of the Stock Market Using Hybrid Machine Learning Algorithms." *Financial Innovation* 5(1).

Zhou, Feng, Hao min Zhou, Zhihua Yang, and Lihua Yang. 2019. "EMD2FNN: A Strategy Combining Empirical Mode Decomposition and Factorization Machine Based Neural Network for Stock Market Trend Prediction." *Expert Systems with Applications* 115:136–51.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 03/09/2020
Aceitação/Acceptance: 08/11/2020

Pautas del modelo de usabilidad PACMAD para la construcción de un aplicativo móvil de gestión de actividades académicas

Celinda Isabel Cueto De la hoz, Carlos Javier Alcalá Orozco,
Gabriel Elías Chanchí Golondrino, Martin Emilio Monroy Ríos

**ccuetod@unicartagena.edu.co; calcalao@unicartagena.edu.co;
gchanchig@unicartagena.edu.co; mmonroyr@unicartagena.edu.co**

Universidad de Cartagena, Avenida del Consulado, calle 30 No. 39 B - 192, 130001, Cartagena de Indias D.T. y C., Colombia.

Pages: 358-372

Resumen: La usabilidad como atributo de calidad de software, ha permitido el desarrollo de aplicaciones interactivas y competitivas, que se adecuan a las necesidades de los usuarios finales. El modelo PACMAD (People At the Centre of Mobile Application Development) reúne atributos significativos de diferentes modelos de usabilidad aplicados a dispositivos móviles. Teniendo en cuenta que no se ha evidenciado la existencia de pautas asociadas a los atributos de este modelo, en este artículo se propone un conjunto de pautas de usabilidad para cada uno de los atributos presentados en este modelo. Como resultado se definieron 60 pautas de usabilidad y se desarrolló un aplicativo móvil para la gestión de actividades académicas de los estudiantes de la Universidad de Cartagena. Se concluye que las pautas definidas y el aplicativo desarrollado, sirven de referencia para la construcción de aplicaciones móviles centradas en el usuario.

Palabras-clave: Aplicativo móvil, gestión de actividades, modelo de usabilidad PACMAD, pautas de usabilidad.

PACMAD model usability guidelines for the construction of a mobile application for managing academic activities

Abstract: The usability as a criterion of software quality has allowed the development of interactive and competitive applications, which are adapted to the needs of the end users. The PACMAD (People at the Center of Mobile Application Development) model integrates significant attributes of different usability models applied to mobile devices. Considering that the existence of guidelines associated with the attributes of this model has not been evidenced, in this paper we propose a set of usability guidelines for each of the attributes presented in this model. As a result, 60 usability guidelines were defined and a mobile application for the management of academic activities of the students of the University of Cartagena was developed. It is concluded that the defined guidelines and the developed

application serve as a reference for the construction of mobile applications focused on the user.

Keywords: Mobile application, management of activities, PACMAD usability model, usability guidelines.

1. Introducción

De acuerdo a Nielsen (2012) la usabilidad es entendida como un atributo de la calidad del software, que mide la facilidad de la interacción entre el usuario y la interfaz. Según la ISO 9241-11, la usabilidad es el grado con el que un producto puede ser utilizado por usuarios específicos para conseguir objetivos concretos con efectividad, eficiencia y satisfacción en un determinado contexto de uso (González-Sánchez, Montero-Simarro, & Gutierrez-Vela, 2012). La medición de la usabilidad es un proceso complejo en el campo de las aplicaciones móviles, porque requiere de técnicas y métodos diseñados de manera específica para cada contexto, que tienen en cuenta aspectos como: el contexto de uso, las características de los dispositivos móviles, entre otros (Ramírez & Otálora, 2016).

En este sentido, para el caso de las aplicaciones móviles se ha definido el modelo de usabilidad PACMAD (*People At the Centre of Mobile Application Development*), el cual reúne atributos significativos de diferentes modelos de usabilidad para crear uno más completo, que además fue diseñado para para abordar las limitaciones de los modelos de usabilidad existentes cuando se aplica a dispositivos móviles (Harrison, Flood, & Duce, 2013). El modelo PACMAD identifica tres factores que pueden influir sobre la usabilidad general de una aplicación móvil: usuario, tareas y contexto de uso. Dentro de sus atributos, en comparación con los definidos por la ISO 9241-11 y de los de Nielsen, se incluye el atributo de "Carga Cognitiva" (Chanchí, Gómez, & Campo, 2020). Este factor se refiere a la cantidad de acciones realizadas por el usuario, para usar una aplicación. En estudios de usabilidad tradicionales, es común que el usuario este realizando sólo una tarea y por lo tanto puede concentrarse completamente en esa tarea. Sin embargo, en el contexto móvil el usuario puede estar realizando varias actividades al mismo tiempo. Los tres atributos más comunes, eficacia, eficiencia y satisfacción, corresponden a los atributos identificados por la norma de ISO para la usabilidad (Chanchí, Gómez, & Campo, 2020).

Aunque el modelo PACMAD se ha definido para el uso en contextos de aplicaciones móviles, no se ha evidenciado la existencia de un conjunto de pautas puntuales asociadas a cada atributo del modelo, que puedan ser usadas para guiar el proceso de diseño y evaluación de aplicaciones móviles desde la perspectiva de la usabilidad (Saleh & Ismail, 2015). Por lo anterior, en este artículo se propone un conjunto de pautas de usabilidad para los diferentes atributos del modelo PACMAD, las cuales fueron obtenidas a partir de la revisión y estudio teórico, usando técnicas de análisis de contenido, de diferentes heurísticas tales como las de Nielsen, Schneiderman y Prierotti, porque han sido utilizadas como referentes por múltiples estudios (Abascal, y otros, 2002) (Sánchez, 2011).

Con el fin de evaluar las pautas propuestas se diseñó, implementó y evaluó el aplicativo móvil Kronos, que sirve para la gestión de actividades académicas de los estudiantes de la Universidad de Cartagena, porque permite organizar y programar sus actividades,

mejorando el uso del tiempo y en consecuencia el rendimiento académico. Como aporte novedoso, Kronos aplicó el tablero Kanban para la gestión de actividades de los usuarios, empleado en el proceso de desarrollo para organizar las diferentes actividades y clasificarlas en tres posibles estados: pendientes, en curso y realizadas (Castellano, 2019) (Ahmad, Markkula, & Oivo, 2013) (Sendil & Panneerselvam, 2007). Incorporando de esta forma al aplicativo desarrollado las buenas prácticas de la metodología Kanban al aplicarlas en la gestión del tiempo.

Las pautas propuestas y el aplicativo móvil desarrollado sirven de referencia para el diseño, construcción y evaluación de aplicaciones móviles para diferentes contextos, desde la perspectiva de la usabilidad. El resto del artículo está organizado de la siguiente forma: en la sección 2 se presentan las diferentes fases metodológicas empleadas en esta investigación; en la sección 3 se describen un conjunto de conceptos que se tuvieron en cuenta para el desarrollo de la presente investigación; en la sección 4 se describen las pautas de usabilidad propuestas y asociadas a cada uno de los atributos del modelo PACMAD; en la sección 5 se presenta el diseño de alto nivel y la implementación del aplicativo móvil propuesto; en la sección 6 se presenta la evaluación del aplicativo móvil propuesto; finalmente en la sección 7 se muestran las conclusiones y trabajos futuros derivados de la presente investigación.

2. Metodología

La investigación se realizó en 3 fases: 1) identificación de los atributos de usabilidad del modelo PACMAD, 2) definición de pautas de usabilidad por atributo, 3) Evaluación de las pautas definidas. Estas fases corresponden a una adaptación de la metodología *Design Thinking*, porque aplica el diseño centrado en el usuario (Brown, 2008) a partir de la generación de prototipos en las fases de diseño e implementación (Serrano & Blazquez, 2013; Muller-Roterberg, 2018). En la Figura 1 se representa cada una de las fases de la metodología aplicada.

Fase 1 – Identificación de atributos del modelo PACMAD: En esta fase se realizó una revisión y estudio teórico que condujo a la identificación del alcance de los siete atributos planteados en el modelo PACMAD: eficiencia, efectividad, facilidad de aprendizaje, capacidad de ser recordado, satisfacción, errores, carga cognitiva.

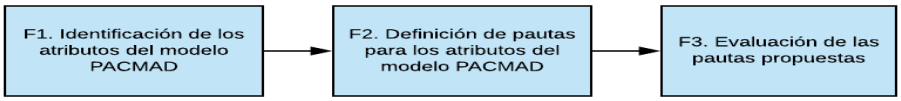


Figura 1 – Metodología aplicada

Fase 2 – Definición de pautas para los atributos del modelo PACMAD: Una vez comprendido el alcance de los diferentes atributos del modelo de PACMAD, se procedió a utilizar técnicas de análisis de contenido para generar un conjunto de pautas aplicables al diseño de aplicativos móviles.

Fase 3 – Evaluación de las pautas propuestas: Se realizó en tres momentos y por equipos diferentes para asegurar la validez de las pautas. El primer momento consistió en diseñar el aplicativo Kronos, cumpliendo con un listado de pautas que se entregó al equipo de trabajo en calidad de restricción de diseño. El segundo momento lo realizó el equipo de desarrollo, a quien también se le dio instrucción de cumplir con las mismas pautas. El último momento lo llevó a cabo el equipo de prueba, y consistió en utilizar el listado de pautas para verificar si realmente se aplicaron en el diseño y desarrollo del sistema.

3. Marco Conceptual

En esta sección se presentan los conceptos claves que te tuvieron en cuenta para el desarrollo de la investigación. Dentro de estos conceptos se encuentran: usabilidad y modelo PACMAD.

3.1. Usabilidad

La usabilidad es definida como el grado en el que un producto puede ser usado por determinados usuarios para lograr sus propósitos con eficacia, eficiencia y satisfacción en un contexto de uso específico (ISO, 2018). De igual manera, puede ser entendida como la disciplina que estudia la forma de diseñar aplicaciones interactivas, cómodas e intuitivas, para que los usuarios puedan interactuar con ellas de la manera más fácil (Chanchí, Gómez, & Campo, 2020).Así mismo, de acuerdo a la ISO 9126-1 la usabilidad es definida como la capacidad del software de ser entendido, aprendido, y usado en forma fácil y atractiva (Alarcón-Aldana, Díaz, & Callejas-Cuervo, 2014) (González-Sánchez, Montero-Simarro, & Gutierrez-Vela, 2012).

3.2. Modelo PACMAD

El modelo PACMAD fue diseñado con el fin de tratar las limitaciones de los modelos de usabilidad existentes en el contexto de dispositivos móviles. PACMAD utiliza atributos significativos de los modelos de usabilidad de Nielsen, Tognazzini y Pierotti (Saleh & Ismail, 2015) entre otros para crear un modelo más completo, además de identificar tres factores que puedan afectar la usabilidad general de una aplicación móvil: usuario, tareas y contexto de uso. En la tabla 1 se presenta la descripción de los diferentes atributos propuestos por este modelo.

Atributo de usabilidad	Descripción
Eficiencia	El aplicativo móvil permite realizar las diferentes tareas en el tiempo adecuado, una vez han aprendido el diseño del sistema.
Efectividad	El aplicativo móvil permite al usuario desarrollar las diferentes tareas de manera adecuada y cumpliendo las expectativas de uso.
Facilidad de aprendizaje	El aplicativo móvil permite realizar de manera intuitiva las tareas básicas al usuario la primera vez que interactúa con el sistema.
Capacidad de ser recordado	El aplicativo móvil permite a los usuarios recordar rápidamente las principales funciones después de un periodo de no uso.

Atributo de usabilidad	Descripción
Satisfacción	El aplicativo móvil debe resultar agradable y familiar al usuario, para lo cual se deben seguir convenciones de aplicaciones del mismo contexto y cumplir con estándares de diseño ampliamente aceptados.
Errores	El aplicativo móvil debe diseñarse de modo de que se puedan prevenir los posibles errores que ocurran y permitir al usuario recuperarse de los mismos , lo cual sugiere una solución adecuada.
Carga cognitiva	El aplicativo móvil debe evitar que el usuario se vea obligado a recordar información entre las diferentes vistas de este.

4. Pautas de usabilidad propuestas

Aplicando la metodología previamente explicada, para cada atributo de usabilidad del modelo PACMAD se definió un conjunto de pautas, según se observa en la Tabla 2, donde se presenta a nivel de resumen el número de pautas definidas por atributo, obteniendo en total 60 pautas. A modo de ejemplo, en la Tabla 3 se relacionan parcialmente el conjunto de pautas que se utilizaron como restricción de diseño del sistema Kronos, y que fueron usadas en el diseño, la construcción y la evaluación de este sistema, según se explica en las siguientes secciones. Estas pautas fueron seleccionadas como resultado de la aplicación de técnicas de ingeniería de requisitos, en el proceso de construcción de la aplicación Kronos.

Atributo de usabilidad	Número de pautas
Eficiencia	8
Efectividad	3
Facilidad de Aprendizaje	12
Capacidad de Ser Recordado	8
Satisfacción	10
Errores	11
Carga Cognitiva	8
Total	60

Tabla 2 – Número de pautas por atributo

Se observa que la efectividad es el atributo con menor número de pautas definidas (tres), mientras que los demás atributos tienen 8 o más. Esto se debe a que la efectividad, a diferencia de los otros atributos depende del cumplimiento adecuado de las tareas dentro del aplicativo, es decir que el aplicativo responda a la necesidad del usuario, por lo cual depende más de la funcionalidad que del diseño de la interfaz de usuario. En este sentido, en los demás atributos se pueden encontrar elementos que desde el diseño pueden contribuir a mejorar la eficiencia, el manejo de errores, la capacidad de recordación, la satisfacción y la carga cognitiva.

Atributo	Pautas de ejemplo
Eficiencia	P1. Kronos debe contar con controles y opciones de uso frecuente que sean visibles para el usuario. P2. Kronos debe permitir hacer y deshacer las diferentes tareas realizadas por el usuario. Es decir, el usuario deberá sentir que tiene control de las tareas que realice dentro de la aplicación. P3. Kronos debe proveer de manera eficiente, un sistema de ayuda al usuario para comprender funciones complejas.
Efectividad	P1. Kronos debe cumplir de manera adecuada con las funciones básicas para el usuario, tales como la gestión de actividades académicas. P2. Kronos debe cumplir de manera adecuada con las acciones básicas correspondientes a la gestión de actividades (agregar, guardar, modificar y eliminar) para el usuario. P3. Kronos debe brindar información de manera adecuada, basada en el contexto del dominio la aplicación.
Facilidad de Aprendizaje	P1. Al momento de ingresar a Kronos se deben realizar tareas básicas como: iniciar sesión o registrarse, si es la primera vez que el usuario la ejecuta. P2. Kronos debe contener iconos ilustrativos para la navegabilidad del usuario en la aplicación, por ende, debe ser sencillo y con imágenes que definan la acción a realizar, con el fin de facilitar la interacción con el sistema. P3. Kronos debe manejar palabras conocidas y bastante familiares para el usuario, es decir, se deben utilizar palabras propias del dominio del problema en todas las interfaces de la aplicación.
Capacidad de Ser Recordado	P1. Kronos debe contar con iconos convencionales y que permitan al usuario inferir la actividad que desee realizar. P2. Kronos debe tener fácilmente visibles las opciones más utilizadas para realizar una tarea. P3. Kronos debe mantener un diseño consistente en cuanto a la información y la disposición de los elementos presentados en las diferentes vistas.
Satisfacción	P1. Los elementos de la interfaz de Kronos deben estar distribuidos de manera adecuada y consistente. P2. Kronos debe evitar los mensajes publicitarios en las diferentes tareas del sistema. P3. Kronos debe emplear colores en la interfaz que permitan descansar la vista.
Error	P1. Kronos debe ser capaz de prevenir y solucionar los posibles errores que el usuario puede cometer en la interacción. P2. Kronos debe evitar el uso de palabras hostiles o violentas en los mensajes de error. P3. Kronos debe usar un lenguaje libre de tecnicismos para presentar la solución.
Carga Cognitiva	P1. Kronos debe evitar solicitar varias veces la misma información en las diferentes acciones que realiza el usuario. P2. Las funciones y controles más importantes del aplicativo móvil deben estar siempre disponibles al usuario. P3. El texto empleado en las diferentes vistas del aplicativo móvil debe ser conciso y claro.

Tabla 3 – Ejemplo de pautas usadas en el diseño, desarrollo y evaluación de Kronos

5. Diseño del Aplicativo Móvil

A partir de esta sección se presentan los resultados de la evaluación de las pautas propuestas, en tres momentos diferentes: El diseño del aplicativo, la implementación y la inspección de usabilidad en calidad de prueba del sistema. El equipo de diseño propuso las interfaces de alto nivel que se presentan en esta sección, tomando como referencia las pautas establecidas en las restricciones de diseño. Se usó la herramienta de maquetado JustInMind. En la figura 2 se presentan las interfaces de alto nivel de aplicativo relacionadas con el inicio de sesión y la gestión de actividades.

Para iniciar sesión el usuario debe ingresar a través desde la cuenta Google para tener acceso a la interfaz principal del aplicativo, donde se le presenta un tablero kanban, en el cual se distinguen tres secciones: pendiente, en progreso y realizado. Dentro de estas tres secciones el usuario puede ir agregando y clasificando sus actividades académicas usando el ícono "+" disponible en la parte superior derecha. Del mismo modo, las actividades podrán ser eliminadas utilizando el bote de basura disponible en la parte superior derecha. Para mover las actividades a uno de los posibles estados basta con arrastrarlas desde la interfaz táctil.



Figura 2 – Interfaces de inicio de sesión y de gestión de tareas

Del mismo modo, en la figura 3, se presentan las interfaces de alto nivel del aplicativo móvil, relacionadas con los tips de gestión del tiempo e información del perfil del usuario. Así, el usuario puede observar los aspectos más relevantes para cumplir con la adecuada gestión de las actividades, además de consultar la información del perfil de usuario.



Figura 3 – Interfaces de tips e información del usuario

6. Implementación del Aplicativo

El equipo de desarrollo de Kronos usó el Framework Ionic, el cual cuenta con un kit de herramientas de interfaz de usuario de código abierto que permite crear aplicaciones móviles y de escritorio de alta calidad con tecnologías web (HTML, CSS y JavaScript) con integraciones para tecnologías de desarrollo populares como Angular y ReactJS, las cuales ayudan a la generación de código fuente. De este modo, a través de Ionic se realizaron las siguientes interfaces a nivel de implementación, en donde se evidencia la aplicación de pautas de usabilidad basadas en el modelo PACMAD.

En la figura 4 se observa las interfaces de inicio de sesión y la interfaz principal del aplicativo móvil, en las cuales se hizo uso de iconos y botones que ofrece el framework Ionic para la construcción de las interfaces, tales como: los iconos para identificar la pestaña de actividades, la opción de tips y para las opciones de agregar, eliminar, reordenar y finalmente para el identificador del menú. Del mismo modo en la Figura 4, se aprecia el ícono de la aplicación, el cual hace alusión a la gestión del tiempo.

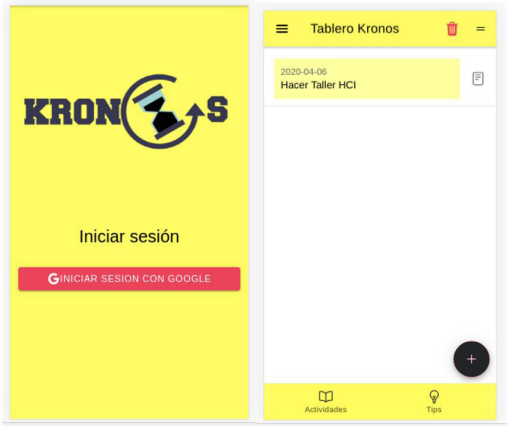


Figura 4 – Interfaces de inicio de sesión y de gestión de actividades.

Por otra parte, en la figura 5 se pueden apreciar las interfaces “Agregar Actividad” y la interfaz que se presenta al momento de seleccionar la opción de reordenar. Por lo tanto, dentro de la vista de: “Agregar Actividad” el usuario puede seleccionar uno de los tres posibles estados en los que se encuentra la tarea que desea agregar de acuerdo al tablero Kanban (Pendiente, En curso, Realizada), los cuales hacen uso de un mismo icono, pero con diferente color. De este modo, se asignó el color rojo para las actividades pendientes, azul para las actividades en curso y verde para las actividades terminadas. También se observan dos campos, uno para asignar la fecha y otro para agregar el nombre de la actividad, así como los íconos de guardar y cancelar la actividad (visto bueno y equis).

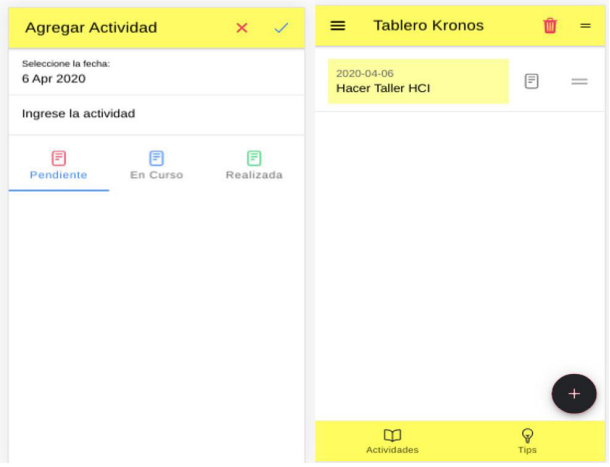


Figura 5 – Interfaces de agregar actividades y reordenar actividades

Siguiendo con la descripción del diseño del aplicativo, en la figura 6, se presenta en la interfaz el menú desplegado, desde el cual se puede acceder a la opción de ayuda, información de la aplicación y la opción de salir de la aplicación. Del mismo modo se encuentra la opción de ver actividades pendientes, en curso y terminadas. Así mismo, se muestra la interfaz de tips, en la cual se exponen imágenes que presentan información sobre técnicas de estudio y gestión del tiempo.

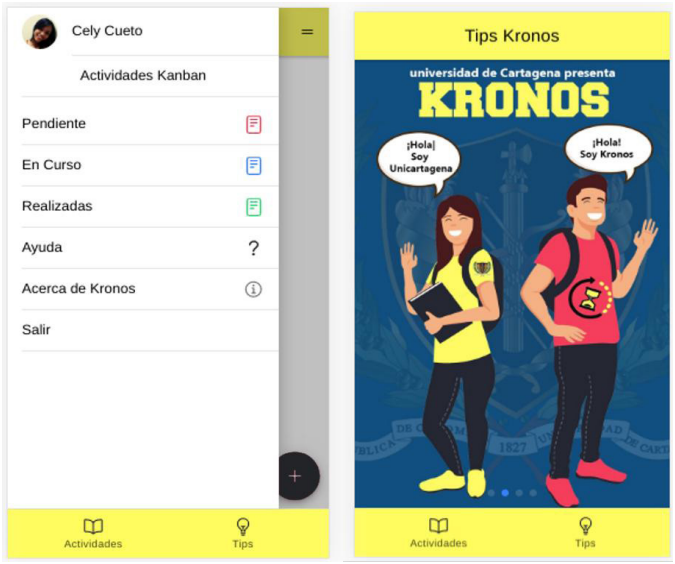


Figura 6 – Interfaces de menú Kanban y opción Tips Kronos

7. Inspección de Usabilidad

El aplicativo móvil propuesto fue evaluado en el curso electivo de Interacción Humano Computador (IHC) del Programa de Ingeniería de Sistemas de la Universidad de Cartagena, dentro de la temática de Principios de Usabilidad, la cual aborda el estudio de los principios de Nielsen, Schneiderman, Constantine, etc. Para realizar la inspección de usabilidad, se definió un conjunto de cinco preguntas para cada uno de los seis principios de usabilidad de Constantine: Estructura, Simplicidad, Visibilidad, Retroalimentación, Tolerancia y Reutilización. Las 30 preguntas se formularon con base en las pautas propuestas y sirvieron de criterio de evaluación de la usabilidad, al realizar la inspección por parte del grupo de estudiantes, con el fin de verificar el cumplimiento de cada principio (Enriquez & Casas, 2013).

El principio de Estructura hace referencia al hecho de facilitar las tareas comunes que el usuario realiza de manera habitual. El principio de Simplicidad consiste en realizar de manera sencilla las tareas que el usuario realiza de manera habitual. El principio de Visibilidad se refiere a darle prioridad a la presentación de las funciones relevantes en

pantalla por sobre las funciones opcionales. El principio de Retroalimentación implica brindar retroalimentación constante a las diferentes acciones que realiza el usuario. El principio de Tolerancia reside en dar la posibilidad de deshacer y rehacer las acciones que realiza el usuario dentro del software. Finalmente, el principio de Reutilización se fundamenta en la reducción de la necesidad de que los usuarios recuerden información de una vista a la otra.

A partir de los criterios de usabilidad definidos para cada uno de los principios de Constantine, se diseñó un formulario en Google Forms mediante el cual un grupo de 15 estudiantes del curso de HCI evaluó el cumplimiento de los principios de usabilidad considerados dentro del aplicativo móvil Kronos (ver Figura 7).

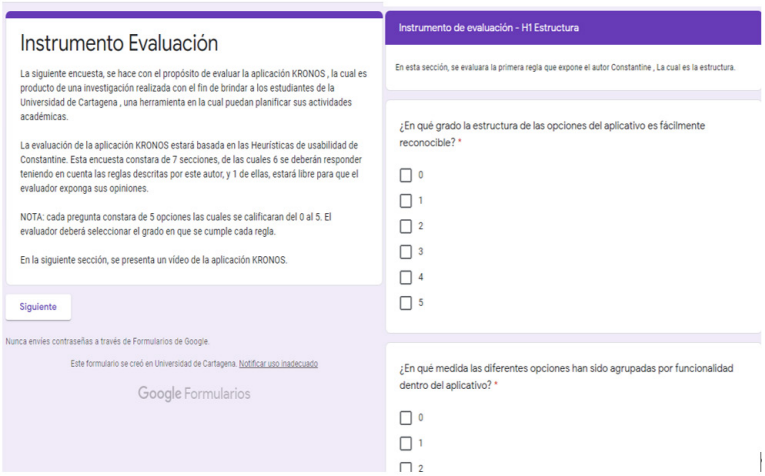


Figura 7 – Instrumento de evaluación

En la figura 7 se presenta parcialmente el formulario diseñado para la inspección de usabilidad, conformado por 6 secciones, de las cuales las 5 primeras corresponden a los criterios de evaluación definidos para evaluar el cumplimiento de los principios de Constantine; mientras que la última sección corresponde a los comentarios abiertos sobre los elementos positivos y por mejorar en el aplicativo móvil propuesto. En la Figura 8 se presentan los resultados de la evaluación, resaltando las valoraciones promedio obtenidas en cada uno de los principios de usabilidad de Constantine.

Se observa que el principio que menos se cumple es el H5 (Tolerancia) con una valoración promedio de 2.9 y un porcentaje de cumplimiento del 58%. Del mismo modo el principio que mejor se cumple es el H3 (Visibilidad) con una valoración promedio de 4.025 y un porcentaje de cumplimiento de 80.5 %. Así mismo, se obtiene como promedio general de usabilidad un valor de 3.393, lo que equivale a un porcentaje general de usabilidad de 67.86%. En este sentido es importante dar prioridad a la heurística H5, en cuanto a la mejora de los aspectos relacionados con el control de la interfaz por parte del usuario, lo

cual incluye la posibilidad de deshacer y rehacer acciones, así como gestionar de manera adecuada los errores que se puedan presentar en el aplicativo (prevención y manejo de los errores).

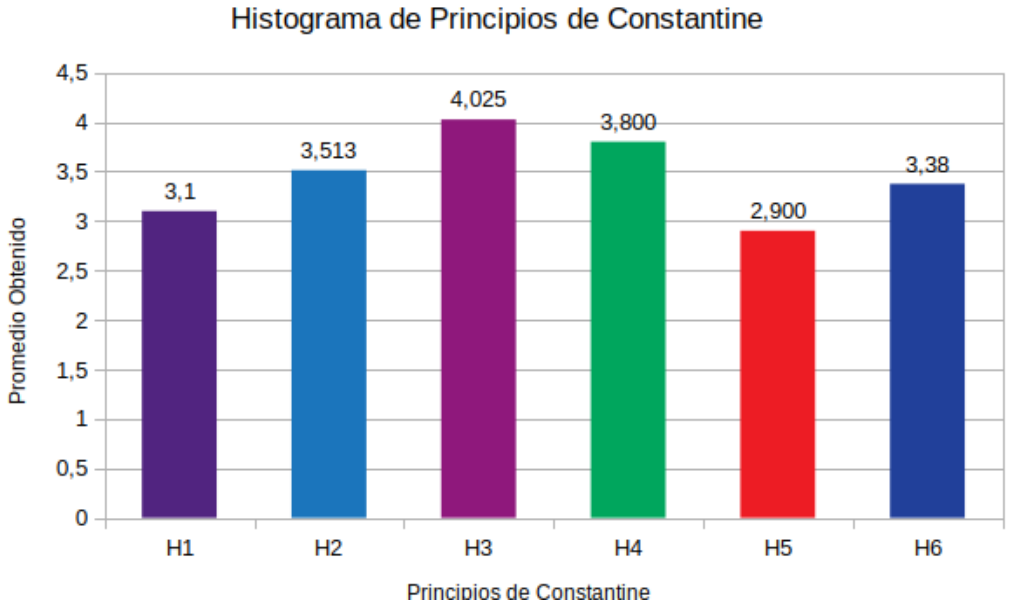


Figura 8 – Valoraciones promedio obtenidas en la evaluación

De acuerdo a los resultados mencionados anteriormente, en la Tabla 4 se presentan algunas recomendaciones obtenidas a partir de la evaluación realizada por los estudiantes del curso de IHC, las cuales han sido agrupadas por cada uno de los principios de Constantine.

Heurística	**Recomendación**
Estructura	R1. Se recomienda mejorar la estructura del aplicativo, de tal modo que guarde relación con otros aplicativos del contexto educativo y de la gestión del tiempo.
Simplicidad	R2. Se recomienda incluir atajos para algunas tareas del usuario, de tal manera que se disminuya el número de *touch* al momento de gestionar las actividades. R3. Se recomienda incluir la funcionalidad de deshacer y rehacer al momento que el usuario agrega una actividad.
Visibilidad	R4. Se recomienda mantener la homogeneidad en cuanto a los colores del tablero Kanban.
Retroalimentación	R5. Se recomienda que, para cada pestaña de actividad pendiente, en curso y terminada, aparezca el título de la sección. R6. Se recomienda incluir una ayuda inicial para que el usuario pueda agregar sus actividades por primera vez más fácilmente.

Heurística	Recomendación
Tolerancia	R7. Se recomienda que el aplicativo mejore la prevención de errores en las diferentes interacciones del usuario con la aplicación. R8. Se recomienda que el aplicativo posibilite la recuperación de información ante posibles errores en la gestión de las actividades. R9. Se recomienda incluir la funcionalidad de deshacer y rehacer en cuanto a la gestión de las actividades.
Reutilización	R10. Se recomienda agregar títulos en las diferentes vistas del aplicativo. R11. Se recomienda reutilizar datos o información previamente suministrada por el usuario en la gestión de actividades.

Tabla 4 – Recomendaciones de usabilidad obtenidas

8. Conclusiones y Trabajos Futuros

Aunque existen diferentes heurísticas para diversos contextos de aplicación, no se han evidenciado un número amplio de heurísticas específicas para el contexto de las aplicaciones móviles (Saleh & Ismail, 2015). En este sentido el modelo PACMAD representa un aporte de cara a la generación de aplicaciones móviles centradas en el usuario. Es por lo anterior, que el aporte principal del presente artículo es la especificación de un conjunto de pautas de diseño, asociadas a los diferentes atributos del modelo PACMAD, con el fin de facilitar su articulación en el proceso de desarrollo de software, como quedó evidenciado en la evaluación de las pautas, durante el diseño, la construcción y las pruebas del aplicativo Kronos.

En este artículo se propuso un conjunto de pautas de usabilidad asociadas a los atributos del modelo PACMAD, las cuales guiaron el proceso de construcción de un aplicativo móvil para la gestión de actividades académicas de los estudiantes de la Universidad de Cartagena. Estas pautas pueden ser extrapoladas a otro tipo de aplicaciones móviles de diferentes contextos de aplicación.

La inspección de usabilidad realizada sobre el aplicativo móvil, por parte de los estudiantes de IHC del programa de Ingeniería de Sistemas de la Universidad de Cartagena, teniendo en cuenta los principios de Constantine, permitió evidenciar que el principio “Visibilidad” (H3) obtuvo el mejor puntaje, mientras que por el contrario el principio “Tolerancia” no cumple en su totalidad la función de prevenir y ayudar a corregir los errores en la interacción, por lo que obtuvo la menor valoración. Del mismo modo los principios: Estructura(H1), Simplicidad(H2), Retroalimentación(H4) y Reutilización(H6), obtuvieron puntajes aceptables, por lo que se puede inferir que, gracias a la utilización de las pautas propuestas, durante el proceso de desarrollo, Kronos cumple con el atributo de calidad usabilidad. No obstante, cabe destacar que se determinaron aspectos a mejorar, tales como: homogeneidad de los colores, inclusión de atajos, mantener títulos en cada sección e implementar la función para la prevención y corrección de errores.

Como trabajo futuro, se propone definir un conjunto de criterios en términos de preguntas, los cuales puedan verificar el cumplimiento de los atributos del modelo PACMAD.

Referencias

Abascal, J., Aedo, I., Cañas, J., Gea, M., Gil, A., Lóres, J., Martínez, A., Ortega, M., Valero, P.,Vélez, M. (2002). *La interacción persona ordenador*. Lleida: AIPO. Obtenido de: https://aipo.es/?q=content/libro-aipo.

Ahmad, M., Markkula, J., & Oivo, M. (2013). Kanban in Software Development: A Systematic Literature Review. *Software Engineering and Advanced Applications (SEAA)* (págs. 9-16). Santander, Spain: IEEE. doi:10.1109/SEAA.2013.28.

Alarcón-Aldana, A., Díaz, E., & Callejas-Cuervo, M. (2014). Guía para la evaluación de la Usabilidad en los Entornos Virtuales de Aprendizaje (EVA). *Información Tecnológica, 25*(3), 135-144. doi:10.4067/S0718-07642014000300016.

Brown, T. (2008). Design thinking. Harvard business review, 86(6), 84-92. Obtenido de https://emprendedoresupa.files.wordpress.com/2010/08/p02_brown-design-thinking.pdf.

Castellano, L. (2019). Kanban. Metodología para aumentar la eficiencia en los procesos. *3C Tecnología. Glosas de innovación aplicadas a la pyme*, 8(1), 30-41. Obtenido de https://dialnet.unirioja.es/servlet/articulo?codigo=6866058.

Chanchí, G., Gómez, M., & Campo, W. (2020). Criterios de usabilidad para el diseño e implementación de videojuegos. *Revista RISTI*(E26), 461-474. Obtenido de http://www.scopus.com/inward/record.url?scp=85079441074&partnerID=8YFLogxK.

Chanchí, G., Gómez, M., & Campo, W. (2020). Propuesta de una herramienta de inspección según los atributos de usabilidad de Nielsen. *Revista RISTI*(E26), 448-460. Obtenido de https://www.scopus.com/record/display.uri?eid=2-s2.0-85079185108&origin=inward&txGid=8178ca63e114dcd94384148525fd6bca.

Enriquez, J., & Casas, S. (2013). Usabilidad en aplicaciones móviles. *Informe Científico Técnico UNPA, 5*(2), 25-47. Obtenido de https://dialnet.unirioja.es/servlet/articulo?codigo=5123524.

González-Sánchez, J., Montero-Simarro, F., & Gutierrez-Vela, F. (2012). Evolución del concepto de usabilidad como indicador de calidad de software. *El profesional de la información*, 21(5), 529-536. Obtenido de http://www.elprofesionaldelainformacion.com/contenidos/2012/septiembre/13.html.

Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science, 1*(1). doi:10.1186/2194-0827-1-1.

ISO. (2018). *ISO 9241-11:2018(en) Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*. ISO. Obtenido de https://www.iso.org/obp/ui/es/#iso:std:iso:9241:-11:ed-2:v1:en.

Muller-Roterberg, C. (2018). *Handbook of design thinking* . Christian Muller-Roterberg. Obtenido de https://www.researchgate.net/publication/329310644_Handbook_of_Design_Thinking.

Ramírez, Y., & Otálora, J. (2016). Medición de usabilidad en el desarrollo de aplicaciones educativas móviles. *Revista Virtual Universidad Católica del Norte*(47), 128-140. Obtenido de https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/746/1272.

Saleh, A., & Ismail, R. (2015). Usability evaluation frameworks of mobile application: A mini-systematic literature review. *The 3rd Global Summit on Education GSE 2015*. Kuala Lumpur-Malasya. Obtenido de https://pdfs.semanticscholar.org/c4c7/4917e7433ecc650297d094487906c6c78c18.pdf?_ga=2.49207142.1350326223.1597369420-1977161414.1597369420.

Sánchez, W. (2011). La usabilidad en ingeniería de software: definición y características. *Ing-novación*, 2, 7-21. Obtenido de http://www.redicces.org.sv/jspui/handle/10972/1937.

Sendil, C., & Panneerselvam, R. (2007). Literature review of JIT-KANBAN system. *International Journal of Advanced Manufacturing Technology*, 32, 393-408. doi: doi.org/10.1007/s00170-005-0340-2.

Serrano, M., & Blazquez, P. (2013). *Design Thinking. Lidera el Presente. Crea el Futuro*. ESIC. Obtenido de https://www.academia.edu/37936016/Desing_thinking_lidera_el_presente_crea_el_futuro_pdf.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 14/08/2020
Aceitação/Acceptance: 19/10/2020

Mitigación del ataque replay en automóviles usando SDR y CNN

Juan Carlos Martínez Quintero, Edith Paola Estupiñán Cuesta,
Saira Vanessa Rodriguez Diaz

juan.martinezq@unimilitar.edu.co; edith.estupinan@unimilitar.edu.co; u1401255@unimilitar.edu.co

Universidad Militar Nueva Granada, Carrera 11 # 101-80, 111071, Bogotá, Colombia.

Pages: 373-388

Resumen: Este documento presenta un sistema capaz de detectar el ataque replay en sistemas de seguridad de automóviles mediante SDR y CNN. El sistema propuesto pretende diferenciar la señal de radiofrecuencia de la llave original, de la señal retransmitida usando un HackRF. La retransmisión de señales se configuró con dos niveles distintos de ganancia (alta y baja). Las señales fueron adquiridas con un RTL-SDR, se procesaron en el dominio del tiempo, se extrajo el espectro en frecuencia y se convirtieron a formato de imagen. Tres redes neuronales convolucionales se entrenaron para diferenciar la señal original versus la retransmitida de ganancia alta, baja o ambas. Se evaluó la eficiencia de las redes mediante matrices de confusión analizando sus métricas y se graficó la curva ROC la cual define la eficiencia de la red mediante su área bajo la curva. Se obtuvo como resultado en todas las redes una exactitud superior al 90%.

Palabras-clave: Ataque replay; Redes Neuronales Convolucionales; RKES; SDR.

Replay attack mitigation in car vehicles using SDR and CNN

Abstract: This document presents a system capable of detecting replay attack in automobile security systems using SDR and CNN. The proposed system aims to differentiate the radio frequency signal from the original key fob and the signal retransmitted using a HackRF. Signal retransmission is configured with two different levels of gain (high and low). The signals were acquired with an RTL-SDR, processed in time domain, the frequency spectrum was extracted and converted into an image format. Three convolutional neural networks were designed to differentiate the original versus the retransmitted signal from high gain, low gain, or embassies. The efficiency of the networks is evaluated by means of confusion matrices analyzing their metrics and the ROC curve is plotted which defines the efficiency of the network by its area under the curve. As a result, an efficiency of over 90% is obtained in all networks.

Keywords*:* Replay attack; Convolutional Neural Networks; RKES; SDR.

1. Introducción

En los últimos años la tecnología SDR (Software defined radio) ha permitido a investigadores y académicos el acceso a la práctica de las comunicaciones inalámbricas. Gracias a esta tecnología el procesamiento de señales puede realizarse en un ordenador convencional y el desarrollo y prueba de tecnologías de RF (radiofrecuencia) es posible sin tener que usar hardware dedicado. Las plataformas de SDR actuales trabajan en un amplio rango de frecuencias y pueden captar o transmitir señales con gran ancho de banda, permitiendo además la configuración de estos y otros parámetros. El uso de este tipo de dispositivos ha permitido la captura y análisis de señales que viajan libremente por el espacio. En este sentido, cualquier persona que disponga de una plataforma de SDR puede sintonizarla en determinada frecuencia y analizar o almacenar la señal que viaja sin restricciones. Esto puede aplicarse a señales de radio comercial, televisión, telefonía móvil y todos los servicios que usan el espectro radioeléctrico para conectar dos o más puntos. Muchos de los sistemas de seguridad de automóviles usan el espectro radioeléctrico para transmitir la señal de bloqueo o desbloqueo del vehículo usando un mando o llave de forma remota. Los sistemas de seguridad más comunes de este tipo se denominan RKES (Remote Keyless Entry Systems). Una plataforma de SDR puede fácilmente captar, almacenar y retransmitir la señal de RF emitida por la llave haciéndole creer al sistema del vehículo que se trata de la señal de la llave original. Este ataque es conocido como ataque de retransmisión o ataque replay. Esta investigación parte de la hipótesis de que las señales pueden contener una huella del hardware que las transmite. Esta huella puede verse reflejada en parámetros como error o cambio leve de la frecuencia, aparición de armónicos o frecuencias no deseadas o simplemente cambio en los niveles del piso de ruido. En diferentes aplicaciones se ha demostrado que técnicas de inteligencia artificial como Machine Learning y Deep Learning tienen alto grado de eficiencia en la clasificación de señales. Particularmente las redes Neuronales Convolucionales conocidas como CNN por sus siglas en inglés son ideales para la clasificación, reconocimiento y procesamiento de imágenes y voz, entregando resultados muy eficientes. Se propone la implementación de este tipo de redes para detectar cambios en las señales de desbloqueo de un vehículo para mitigar el ataque replay.

La aplicación de técnicas de inteligencia artificial para clasificación de señales de radio ha estado en gran parte limitada al reconocimiento y clasificación de modulaciones digitales y análogas en muchos casos usando CNN mediante la integración de librerías como Keras y Tensorflow. En Karra et al. (2017) se identificó el tipo de dato, clase de modulación y orden de modulación. Como resultado se obtuvo que la CNN alcanzó una precisión promedio del 80% en las predicciones. En O'Shea et al. (2017) se tomaron muestras de señales moduladas en función del tiempo y su respectivo diagrama de constelación para el entrenamiento de la red. Finalmente, la red entregó una precisión de más del 90%. En Hauser et al. (2017) las muestras para identificar el tipo de modulación fueron datos I/Q sin procesar para cuantificar el impacto que tienen los datos dentro de la red. La red obtuvo una precisión entre el 80% y el 90%. En Ziegler et al. (2017) se emplearon muestras I/Q en bruto, espectro de potencia y densidad de correlación espectral para el entrenamiento de la red neuronal convolucional. Se obtiene un sistema capaz de identificar el tipo de modulación en señales transmitidas en vivo. En Wang et al. (2019)

las redes fueron entrenadas con diagramas de constelación y señales de componentes en fase y cuadratura. Como resultado se obtuvo una CNN con una precisión del 95% en la identificación de las modulaciones.

Por otro lado, en Riyaz et al, (2018) se propone una red neuronal convolucional para identificar un radio transmitiendo entre dispositivos similares. Las muestras fueron tomadas mediante SDR y se detectaron las señales generadas por una USRP. El trabajo presentó una precisión entre el 90% y 99%.

Desde el punto de vista de ataque replay en sistemas de seguridad de vehículos y su posible mitigación se toman como referencia los siguientes: En Patel et al. (2018) se presentó un nuevo sistema RKE mediante una Raspberry Pi3 y la biblioteca piVirtualWire. Se evaluó y se obtuvo un sistema capaz de detectar y mitigar el ataque replay y otros ataques. En Selvakumar et al. (2018)se propuso un mecanismo de criptografía para prevenir el ataque replay en vehículos. Se genera un código de 64 bits para la identificación de la señal enviada entre la llave remota y el vehículo. Finalmente se obtuvo como resultado un sistema capaz de mitigar el ataque replay. En Merco et al, (2018) se propuso un algoritmo de diagnóstico para la detección del ataque replay en vehículos. El sistema propuesto consta de la simulación de 6 vehículos y se verificó la efectividad del algoritmo para la detección de un ataque replay. En Greene et al, (2020) se presentó un prototipo que permite fortalecer la seguridad en sistemas RKE. Se realizó la captura y retransmisión de las señales de bloqueo/desbloqueo mediante LimeSDR y GNURadio. Se obtuvo como resultado un prototipo funcional capaz de evitar ataques replay. En Joo et al. (2020) presenta un método de huellas digitales de RF llamado Hold the Door. Es un método de sub-autenticación que soporta los sistemas de seguridad de los vehículos. Los resultados demuestran que este método detecta de manera competente y confiable los ataques a estos sistemas de seguridad.

De las investigaciones anteriores no hay evidencia de mitigación del ataque replay basado en SDR y CNN. Se encuentran múltiples investigaciones de clasificación de señales de radiofrecuencia usando CNN y otras investigaciones de métodos de mitigación del ataque usando SDR pero no la integración de los dos conceptos para mitigar este ataque en sistemas de seguridad en automóviles. Esta investigación propone un sistema de detección del ataque replay en vehículos con sistemas de seguridad RKES que no emplean la técnica "rolling code" usando SDR y CNN para mitigar el riesgo de robo. El método propuesto usa un RTL y un computador para la captura y procesamiento de la señal. Se definieron tres CNNs, capaces de identificar entre una señal original y su réplica o retransmisión a diferentes ganancias (baja, alta o ambas) y se determinó una precisión de más del 90% en las tres CNNs propuestas. Las muestras para el entrenamiento y ejecución de las redes fueron tomadas de la llave de un solo vehículo de gama media, modelo 2016 que maneja un sistema de desbloqueo RKES las retransmisiones se hicieron usando un HackRF one. Se tomaron muestras a diferentes distancias del detector. Las distancias estuvieron entre 0.5 y 5 metros variando la polarización de las antenas de forma aleatoria. El artículo está organizado inicialmente con una introducción teórica y posteriormente se desarrolla el enfoque planteado: adquisición de muestras, preprocesamiento, diseño de las CNN, análisis de resultados y conclusiones del trabajo realizado.

2. Contexto teórico

2.1. Ataques a sistemas de seguridad de vehículos

Un ataque se define como una acción deliberada ante un sistema de seguridad, llevado a cabo por terceros con el fin de tener acceso no autorizado a información, dispositivos o elementos que son propiedad de una persona (BA-CSIRT, 2018). En la actualidad los automóviles modernos se enfrentan ante múltiples ataques tal como lo describen en Sommer et al. (2019) estos ataques pueden ser: replay, de denegación de servicio, manipulación o modificación del firmware, entre otros. Para este caso se destaca el ataque replay. Este ataque se produce cuando se intercambia información entre la llave del vehículo y el vehículo. Es allí cuando un usuario mal intencionado percibe la señal entre las dos partes, la replica hacia el receptor y haciéndose pasar por el emisor original logra acceder al vehículo. Este ataque explota el hecho de que la comunicación se realiza solo en un sentido (de la llave hacia el vehículo), en este caso, la criptografía no juega un papel importante porque la información transmitida por el sistema RKES sin "rolling code" siempre se repite de la misma forma para desbloquear el vehículo. Este ataque se produce afectando dos principios de un sistema automotriz: la autenticación y la seguridad del vehículo (Rizvi et al., 2017).

2.2. Radio Definido por Software (Software Defined Radio - SDR)

Tal como lo indica Collins et al. (2018) en su libro, el concepto de SDR adopta avances en todos los campos como procesamiento digital de señales, algoritmos de comunicación digital, inteligencia artificial, diseño de hardware de radiofrecuencia (RF), entre otras, para obtener en un solo dispositivo un transmisor y receptor eficientes de acuerdo con su implementación. Los dispositivos SDR permiten la transferencia de información de forma inalámbrica mediante radiación electromagnética, ofreciendo una mayor flexibilidad y reconfigurabilidad a comparación de sistemas de radio tradicionales.

2.3. Deep Learning (DL)

El aprendizaje profundo es un tipo de aprendizaje automático, en el que se crea un modelo, el cual aprende a realizar labores de clasificación mediante imágenes, texto o sonido. Una red tradicional se conforma de solo 2 o 3 capas, mientras que en una red profunda se pueden tener cientos. Esta red se conforma por tres tipos de capas, capa de entrada, capas intermedias u ocultas, las cuales emplean la salida de la capa anterior como entrada y una capa de salida la cual determina la clasificación, cada capa está interconectada mediante nodos o neuronas (The MathWorks, 2018). Deep Learning está conformado por múltiples redes que pueden ser implementadas en conjunto o individualmente de acuerdo con el enfoque que se tenga, algunas de las redes se detallan a continuación: Perceptrón multicapa o MLP por sus siglas en inglés se define como una red neuronal capaz de agrupar sus neuronas en tres tipos de capas: capa de entrada, capas ocultas y capa de salida (Olatunji et al., 2019), otro tipo de red son las redes neuronales recurrentes o RRN por sus siglas en inglés, estas redes se caracterizan por tomar la salida o estado oculto, como entrada, son útiles ya que pueden almacenar información sobre sus entradas pasadas durante cierto periodo de tiempo (Mallya, 2016), finalmente otro tipo de red es la red neuronal convolucional o CNN la cual se detalla en la sección 2.4.

2.4. Redes neuronales convolucionales (Convolutional Neural Networks - CNN)

Las CNNs son un subconjunto de las redes de aprendizaje profundo, que son empleadas para el reconocimiento de imágenes, detección de objetos o la segmentación semántica. Cuentan con una capa de entrada y una capa de salida, así como múltiples capas ocultas, las cuales se dividen en tres tipos de capas convolucional, pooling o agrupación y una capa completamente conectada (Arsenov et al., 2018). La capa convolucional equivale a un mapa de características de una capa oculta de entrada con una matriz de peso que da lugar a un mapa de características de salida (Sakib et al., 2018), la capa de agrupación o de pooling se encarga de la reducción de dimensiones, la cual equivale al promedio o el máximo valor de una sub-imagen, al igual que en la capa convolucional. A diferencia de la capa convolucional, esta capa no contiene parámetros, ya que maneja operadores de agrupación deterministas (Zhang et al., 2020), finalmente la capa totalmente conectada se encarga de la clasificación dentro de la red neuronal convolucional basada en los atributos que se extrajeron en las anteriores capas (convolucional y de agrupación), esta se compone de múltiples capas totalmente conectadas, es decir, que cada una de las neuronas de la primera capa se conectará con cada neurona de la segunda capa, esto suele verse en las redes neuronales Perceptrón Multicapa (MLP) convencionales (Google LLC, 2018).

2.5. Herramientas de Deep Learning

Dentro de las herramientas presentes en deep learning se destacan dos librerías empleadas para la construcción de redes, estas son Tensorflow y Keras. Tensorflow fue desarrollada por el equipo de Google Brain y se conforma por una arquitectura de tres partes: preprocesamiento de datos, construcción del modelo y entrenamiento, y estimación del modelo (Tensorflow, 2020). Keras es una plataforma de código abierto empleado para el aprendizaje automático, permitiendo la creación de prototipos de redes convolucionales, redes recurrentes o una combinación de las dos (Keras, 2020). Otra herramienta empleada para la evaluación del rendimiento de la red es la matriz de confusión la cual se compone de una tabla de dimensiones NXN donde se resume e indica el nivel de éxito de las predicciones de la clasificación, es decir, la relación entre la etiqueta y la clasificación del modelo. Esta matriz permite determinar patrones de error en el modelo de clasificación (Google LLC, 2018). Los parámetros definidos en una matriz de convolución son cuatro y a continuación se definen: Verdaderos Negativos (VN) que equivalen a un falso real, que se predijo como falso, Falsos Positivos (FP) que hacen referencia a un falso real, que se predijo como verdadero. Falsos Negativos (FN) los cuales hacen referencia a un verdadero real, que se predijo como falso y por último Verdaderos Positivos (VP) los cuales hacen referencia a un verdadero real, que se predijo como verdadero. Las siglas de los parámetros previamente definidos se emplearán en la sección 3.4 del presente trabajo.

3. Enfoque propuesto

El proceso de mitigación del ataque replay usando SDR y CNN se define en la figura 1. Se realiza inicialmente la adquisición de muestras, seguido del procesamiento de la señal.

Una vez procesada la señal se diseñan y entrenan 3 redes neuronales convolucionales y finalmente se analiza la eficiencia de las redes CNNs, mediante la matriz de confusión y sus métricas más relevantes.

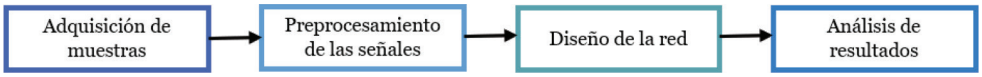


Figura 1 – Metodología propuesta

3.1. Adquisición de muestras

El escenario para la recolección de las muestras se ilustra en la figura 2. El sistema se compone de tres secuencias: Captura de la señal original transmitida por la llave del vehículo, captura de la señal retransmitida y la identificación por el sistema propuesto de una señal original o su retransmisión. El proceso implica:

A. Generar una señal radioeléctrica desde una llave remota de vehículo para almacenarla usando un dispositivo HackRF haciendo la función de receptor. Este dispositivo presenta un rango de operación en frecuencia desde 1 MHz hasta 6 MHz.
B. Retransmisión de la señal original mediante un sistema SDR, el cual está conformado por un dispositivo HackRF haciendo la función de transmisor y un computador.
C. Captura de las dos señales para su posterior procesamiento e identificación, mediante un RTL-SDR el cual opera como receptor de señales RF con un rango de operación en frecuencia de 20 MHz a 2 GHz.



Figura 2 – Escenario para la adquisición de muestras.

3.2. Preprocesamiento de las señales

El preprocesamiento de las muestras de entrenamiento y prueba de la red se realiza como se ve en el diagrama de la figura 3.

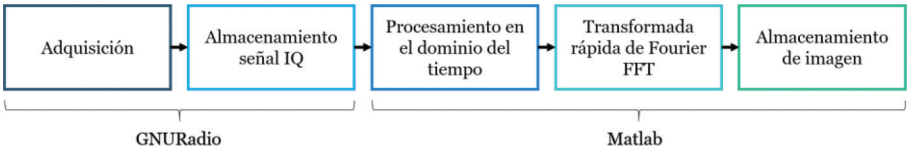


Figura 3 – Preprocesamiento de las muestras.

El diagrama de bloques implementado en el software GNURadio para la adquisición de la señal se muestra en la figura 4. Se utilizó un PC conectado a un módulo RTL-SDR para leer la señal de RF, moverla a banda base y digitalizarla. La frecuencia de muestreo se configuró a 2Msps con una ganancia de RF de 20dB. La frecuencia de muestreo seleccionada es suficiente para adquirir la señal de RF de interés y capturar armónicos que se generan alrededor de esta, algunos de ellos debido a la construcción del hardware que transmite la señal. La frecuencia de la señal de la llave del vehículo de prueba está por encima de los 433,427MHz, así que se configura el oscilador de referencia del RTL a esta frecuencia. En el diagrama aparece el bloque de configuración del RTL "RTL-SDR Source", el bloque para almacenar la señal "File Sink" y un bloque para visualizar el espectro en tiempo real de la señal "QT GUI Sink". Este último bloque solo se usa para verificar de forma visual la lectura de la señal.

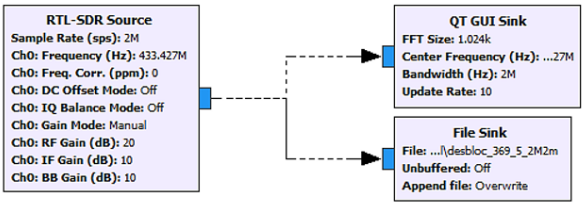


Figura 4 – Diagrama de bloques en GNURadio para leer la señal de RF.

La señal de radio puede ser transmitida por la llave original del sistema de seguridad del vehículo de prueba o también puede ser retransmitida por el hardware con el que se ejecuta el ataque, en este caso la plataforma HackRF. Este hardware se conecta a un segundo PC que cuenta con el software GNURadio. El diagrama de bloques de la figura 5 permite la retransmisión de la señal previamente almacenada con un diagrama similar al de la figura 4. El bloque "File Source" lee la señal almacenada y el bloque "osmocom Sink" envía la señal al HackRF para ser transmitida. Este último bloque permite configurar la frecuencia de transmisión (433,427MHz), la frecuencia de muestreo (2Msps) y la ganancia de transmisión.

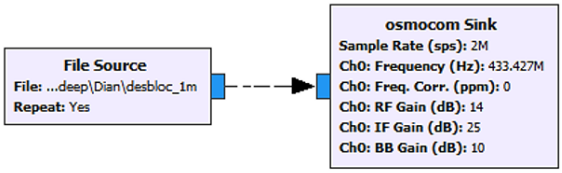


Figura 5 – Diagrama de bloques en GNURadio para retransmitir la señal Fake.

En el presente trabajo se configuraron dos niveles de ganancia para la retransmisión de la señal de RF (ataque replay). Los tres parámetros de ganancia para cada escenario replay se muestran en la tabla 1. Cuando se configura la ganancia alta, la señal recibida por el RTL mediante el diagrama de la figura 4, es una señal saturada.

Parámetro	Ganancia Baja	Ganancia Alta
RF Gain	14	14
IF Gain	25	40
BB Gain	10	62

Tabla 1 – Configuración de niveles de ganancia.

La señal digital recibida en el tiempo puede tener una amplitud entre -1 y 1 (figura 6), la saturación ocurre cuando a la entrada del receptor hay una señal que puede superar estos valores, en este caso la señal se recorta y no supera estos valores tope. El recorte de la señal puede ser visto como una distorsión de la señal recibida, sin embargo, es raro que afecte la demodulación y decodificación de la señal de RF por parte del sistema de seguridad. Con el aumento de la distancia al sistema de recepción, aumenta la atenuación de la señal y por tanto disminuye la saturación.

Las pruebas realizadas no superan un radio de 5 m al receptor (RTL) y en todos los casos la señal con alta ganancia llegó saturada. En el caso de la baja ganancia la amplitud de la señal está entre -1 y 1, sin alcanzar los valores máximos como en la figura 6.

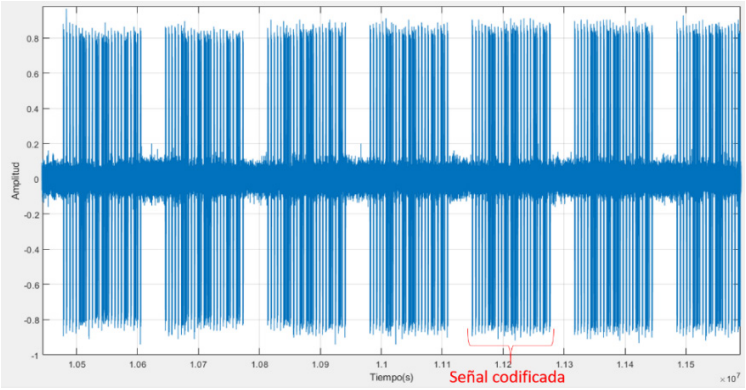


Figura 6 – Señal de desbloqueo del vehículo en el dominio del tiempo.

Cada una de las muestras proviene de la llave real, del HackRF con alta ganancia o del HackRF con baja ganancia. Estas muestras se toman a diferentes distancias, variando la polarización de las antenas, luego se almacenan con el diagrama de la figura 4 y posteriormente se leen usando el software Matlab. La figura 6 muestra la señal en el dominio del tiempo. Esta señal está conformada por varias repeticiones del código de desbloqueo del sistema de seguridad con modulación ASK. Para el procesamiento de la señal en el tiempo en Matlab, se separa una de estas repeticiones usando un algoritmo que detecta el espacio en blanco (sin transmisión) antes del inicio de la repetición. La ventana de la señal obtenida se procesa usando el algoritmo de transformada rápida de

Fourier. El resultado se muestra en la figura 7. La imagen del espectro se almacena con formato mapa de bits.

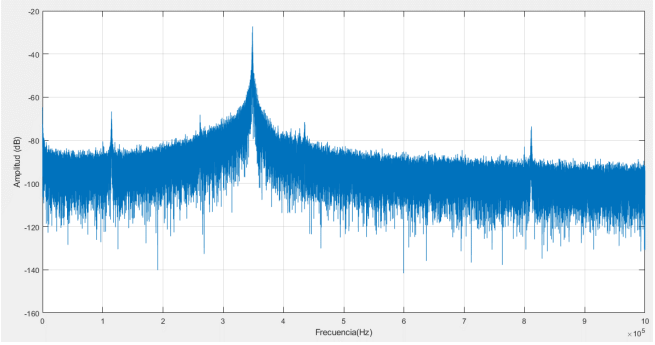


Figura 7 – Resultado de la FFT para una repetición de la señal.

Previo del ingreso de la muestra a la red para el entrenamiento, se realiza una compresión de píxeles, convirtiendo la imagen a las dimensiones de 150 x 150 píxeles, como se indica en la figura 8. De forma general los pixeles suelen tener un valor de 0 a 255, sin embargo, para que el entrenamiento de la red fuera más eficiente, se procesó la imagen para que cada píxel en ella tuviera un valor de 0 a 1.

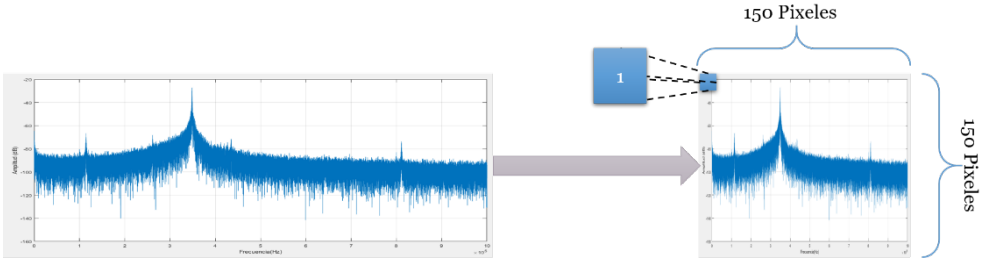


Figura 8 – Procesamiento de la señal.

3.3. Diseño de la red

A partir de la arquitectura de una red neuronal convolucional se propone el diseño del detector, haciendo uso de las imágenes de espectro previamente procesadas. Con el fin abarcar los posibles parámetros con los que se generaría una señal retransmitida en un ambiente real se diseñaron tres CNNs. Cada CNN identifica y clasifica diferentes tipos de categorías como se describe a continuación:

- CNN1 clasifica una señal original y su retransmisión con una ganancia alta.
- CNN2 clasifica una señal original y su retransmisión con una ganancia baja.
- CNN3 clasifica una señal original y su retransmisión tanto con una ganancia baja como con una ganancia alta.

Las redes CNN1 y CNN2 se entrenaron con once muestras para cada categoría (Señal original, señal retransmitida ganancia baja y señal retransmitida con ganancia alta), mientras que la red CNN3 se entrenó con la combinación de las anteriores muestras. La estructura de las tres CNNs es la misma. Las muestras se ingresan a la CNN la cual se conforma por un mapa de características por cada muestra/imagen, dos capas convolucionales y de agrupación, y finalmente una capa completamente conectada. La capa completamente conectada está compuesta por 256 neuronas en su capa de entrada, ocho capas ocultas y una capa de salida binaria en donde se identifica el tipo de señal (real o retransmitida). En la figura 9 se muestra el modelo general para el proceso de entrenamiento y ejecución de las CNNs en la identificación de los dos tipos de señales.

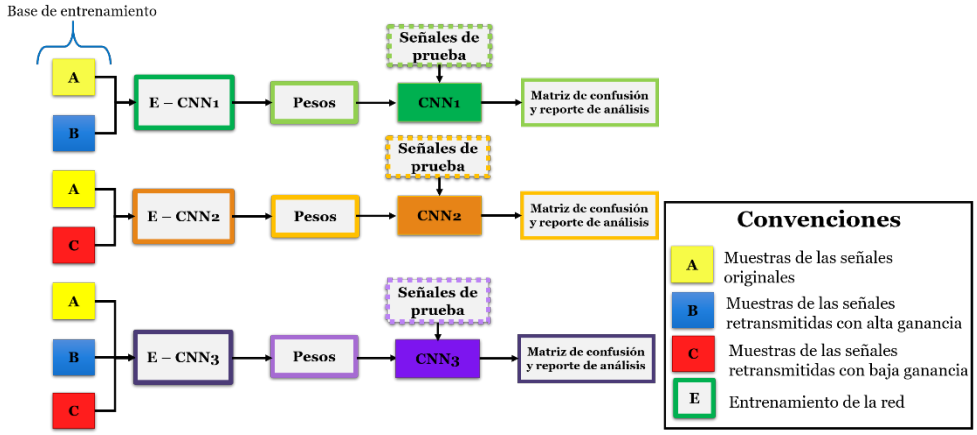


Figura 9 – Modelo general del entrenamiento.

La arquitectura de todas las redes se muestra en la figura 10. Cada CNN se conformó por una entrada o mapa de características, la cual corresponde a la imagen previamente procesada de la señal, seguida de dos capas convolucionales y dos capas de pooling. Dicha arquitectura se implementó en Google Colab empleando las librerías de Google Tensorflow y Keras.

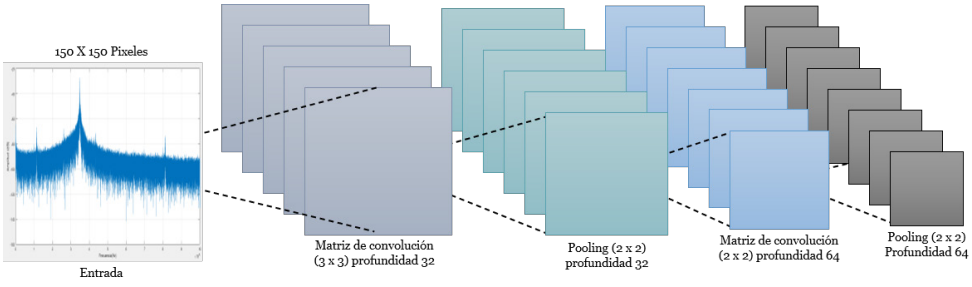


Figura 10 – Arquitectura implementada.

La primera capa de convolución se conforma por una matriz de 3X3X32 píxeles, seguida de una capa de pooling o agrupación, la cual se encarga de tomar el valor máximo en la matriz, permitiendo extraer las características más relevantes en ella. Esta capa tiene unas dimensiones de 2X2X32 píxeles. Luego nuevamente se tiene una capa de convolución de dimensiones 2X2X64 píxeles y una capa de pooling de 2X2X64 píxeles. Posteriormente se convierte la salida de la última capa de agrupación en una matriz unidimensional.

Finalmente, se toma la salida unidimensional y se pasa a la capa totalmente conectada o MLP (figura 11). La capa de entrada se conforma por 256 neuronas. Posteriormente se tienen ocho capas ocultas las cuales se van reduciendo en un 50% en cada paso, con el fin de evitar que la red tome el mismo camino siempre. Por último, se tiene una sola neurona en la capa salida, la cual tiene una clasificación binaria que identifica entre una señal original o una señal retransmitida.

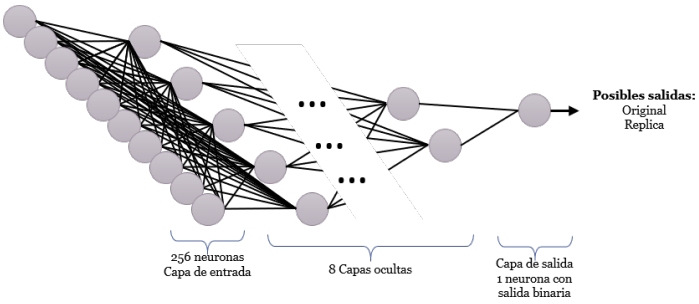


Figura 11 – Capa totalmente conectada o MLP.

3.4. Análisis de resultados

Para evaluar la eficiencia de cada red se empleó la matriz de confusión, en ella se puede observar la cantidad de datos que fueron identificados correctamente y los que se clasificaron erróneamente. La matriz se ilustra en la figura 12 y se muestra tanto en mapa de calor como en valores numéricos.

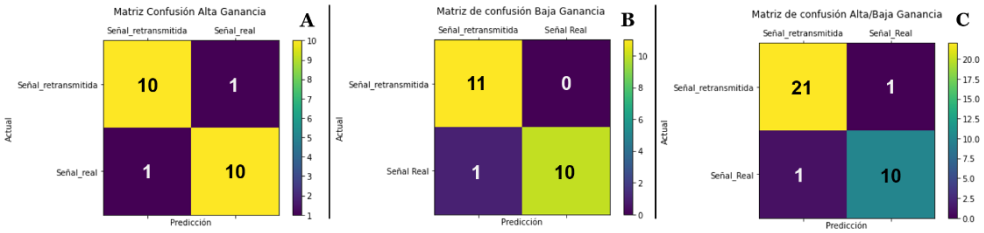


Figura 12 – Matriz de confusión. A) Señal real vs Señal retransmitida con alta ganancia.
B) Señal real vs Señal retransmitida con baja ganancia.
C) Señal real vs Señal retransmitida con baja/alta ganancia.

Se emplearon los valores obtenidos en la tabla de confusión para realizar los cálculos de las siguientes métricas que se resumen en la tabla 2 y así evaluar el rendimiento de cada CNN. En la tabla 3 se resumen los resultados obtenidos.

Métrica	Definición	Fórmula	Ec.
Sensibilidad (S)	Este parámetro permite identificar el porcentaje de señales originales que fueron clasificadas correctamente, basado en las muestras ingresadas. Se calcula a partir de la ecuación 1.	$S = \frac{VP}{VP + FN}$	Ec. (1)
Especificidad (Es)	Está definida como la métrica que permite identificar el porcentaje de las señales retransmitidas que fueron identificadas correctamente. Se calcula a partir de la ecuación 2.	$Es = \frac{VN}{VN + FP}$	Ec. (2)
Valor de predicción positiva (VPP)	Este parámetro determina la probabilidad de que, al ingresar una señal original, efectivamente sea clasificada de dicha forma y se calcula a partir de la ecuación 3.	$VPP = \frac{VP}{VP + FP}$	Ec. (3)
Valor de predicción negativa (VPN)	Define la probabilidad de que, dada una señal retransmitida, el sistema en realidad la clasifique de forma correcta y se calcula mediante la ecuación 4.	$VPN = \frac{VN}{VN + FN}$	Ec. (4)
Error de clasificación (EDC)	Esta variable permite obtener el porcentaje de predicciones que fueron clasificadas erróneamente por la red, de acuerdo con las muestras ingresadas. Se calcula a partir de la ecuación 5.	$EDC = \frac{FP + FN}{VP + VN + FP + FN}$	Ec. (5)
Exactitud (EX)	Es el porcentaje de la red que se clasificó correctamente tanto para la señal original, como para la señal retransmitida y se calcula a partir de la ecuación 6.	$EX = \frac{VP + VN}{VP + VN + FP + FN}$	Ec. (6)
AUC – Área bajo la curva ROC	Esta métrica se determina a partir del área bajo la curva ROC (receiver operating characteristic curve). Permite identificar la calidad de las predicciones de la red. Se toma un ejemplo aleatorio de la señal original o señal retransmitida y se evalúa qué tan eficiente es la red al clasificarla.		

Tabla 2 – Métricas evaluadas para cada CNN.

Métrica	Valor obtenido en % CNN 1	Valor obtenido en % CNN 2	Valor obtenido en % CNN 3	Conclusiones
Sensibilidad	90,9	90,9	90,9	Todas las redes clasificaron el mismo porcentaje de señales originales correctamente, con respecto a las muestras ingresadas
Especificidad	90,9	100	95,5	Estos resultados permiten identificar como la red CNN2 presentó una gran eficiencia al momento de clasificar las señales retransmitidas, de acuerdo con las muestras ingresadas.

Métrica	Valor obtenido en % CNN 1	Valor obtenido en % CNN 2	Valor obtenido en % CNN 3	Conclusiones
Valor de predicción positiva	90,9	90,9	90,9	Las tres redes presentan la misma probabilidad para identificar una señal original
Valor de predicción negativa	90,9	100	95,5	La red CNN2 presenta una mayor probabilidad para identificar las señales retransmitidas correctamente.
Error de clasificación	9,09	4,54	6,06	Se observó que la red CNN2 es la que presenta mayor eficiencia al tener un porcentaje de error bajo, a comparación de las demás redes.
Exactitud	90,9	95,45	93,93	La red CNN2 es quien mayor exactitud al momento de identificar las señales presenta.
AUC – Área bajo la curva ROC	91	95	94	Se determina que la CNN2 es la que mejor logró identificar y clasificar las señales ingresadas. La gráfica se muestra en la figura 13.

Tabla 3 – Parámetros evaluados en cada CNN.

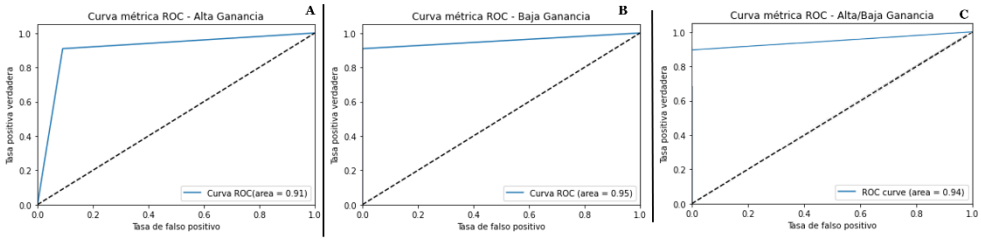


Figura 13 – Curva ROC A) CNN1, B) CNN2, C) CNN3

Con los datos que evidencian las tablas 2 y 3 se muestra que el método propuesto para mitigar ataques replay presenta valores por encima del 90% en cuanto a su sensibilidad, especificidad, valor de predicción positiva o negativa y exactitud, demostrando que las tres redes pueden ser usadas para la predicción de señales retransmitidas. Adicional a esto se analiza que todas las redes presentan un error de clasificación por debajo del 10%, lo cual indica que la red clasifica erróneamente un pequeño porcentaje en comparación con las clasificaciones correctas. La AUC permite conocer la red que mejor probabilidad de predicción tiene, para este caso la CNN2 y la CNN3 tienen la mejor probabilidad.

4. Conclusiones

Se propuso un sistema capaz de mitigar el ataque replay en sistemas de seguridad RKES, incorporando dispositivos de SDR para la adquisición de las señales de RF

originales y sus réplicas. Se diseñaron tres CNNs, capaces de identificar entre una señal radioeléctrica original y una señal retransmitida. La primera red pretende identificar señales provenientes de un dispositivo atacante HackRF configurado con ganancia alta que, dada su cercanía con el receptor lo puede saturar modificando su espectro en frecuencia en alguna medida sin afectar la apertura del vehículo. La segunda red opera en casos específicos en los cuales el atacante configura el dispositivo SDR HackRF con una ganancia baja, esta señal recibida también puede interpretarse como una señal transmitida desde lejos y que llega al receptor con un nivel aceptable. La última red se entrenó teniendo en cuenta los dos casos, por tanto, se considera que es la más completa. La segunda red entregó los mejores resultados, identificando el 100% de las señales retransmitidas y el 90.9% de las señales originales con error de clasificación total del 4.54%. La tercera red tuvo el segundo mejor desempeño con error de clasificación del 6.06%, mientras que la primera red contó con un error de clasificación del 9.09%. En general los resultados indican una exactitud superior al 90% en todas las redes propuestas. Esto último indica que las redes neuronales implementadas logran detectar diferencias en el espectro en frecuencia de una señal transmitida por la llave original del sistema de seguridad y un dispositivo SDR de bajo costo como el HackRF one. Se puede concluir que el hardware usado para la retransmisión introduce diferencias con la señal original como errores de frecuencia y frecuencias indeseadas que pueden ser detectadas usando técnicas de inteligencia artificial, por tanto, al entrenar la red en cualquier vehículo con un sistema de seguridad RKES que no emplea la técnica "rolling code" se identificarán diferencias similares y se obtendrán resultados del mismo orden. Finalmente se concluye que el sistema propuesto tiene un alto nivel de efectividad en la mitigación de ataques replay a sistemas RKES usando SDR y que con los resultados presentados en este artículo es posible generar una alarma para evitar un posible robo.

5. Agradecimientos

Esta investigación fue desarrollada al interior del grupo de investigación GISSIC como producto derivado del proyecto de investigación INV ING 2388 financiado por la vicerrectoría de investigaciones de la Universidad Militar Nueva Granada y desarrollado en el semillero Maxwell registrado y avalado por la vicerrectoría de investigaciones de la Universidad Militar Nueva Granada.

Referencias

Arsenov, A., Ruban, I., Smelyakov, K., & Chupryna, A. (2018). Evolution of Convolutional Neural Network Architecture in Image Classification Problems. *CEUR Workshop Proceedings*, *2318*, 35–45.

BA-CSIRT. (2018). *Ataques cibernéticos*, 1-6. https://www.ba-csirt.gob.ar/files/boletines/B46_AtaquesCiberneticos.pdf

Collins, T. F., Getz, R., Pu, D., & Wyglinski, A. M. (2018). Software-defined radio for Engineers. In *Alcatel Telecommunications Review*. Artech House Publishers.

Google LLC. (2018). *Glosario sobre aprendizaje automático | Google Developers*. https://developers.google.com/machine-learning/glossary/

Greene, K., Rodgers, D., Dykhuizen, H., McNeil, K., Niyaz, Q., & Shamaileh, K. Al. (2020). Timestamp-based defense mechanism against replay attack in remote keyless entry systems. *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, *2020-Janua*, 20–23. https://doi.org/10.1109/ICCE46568.2020.9043039

Hauser, S. C., Headley, W. C., & Michaels, A. J. (2017). Signal detection effects on deep neural networks utilizing raw IQ for modulation classification. *Proceedings - IEEE Military Communications Conference MILCOM*, *2017-Octob*, 121–127. https://doi.org/10.1109/MILCOM.2017.8170853

Joo, K., Choi, W., & Lee, D. H. (2020). Hold the Door! Fingerprinting Your Car Key to Prevent Keyless Entry Car Theft. *Network and Distributed Systems Security (NDSS) Symposium 2020*, *February*. https://doi.org/10.14722/ndss.2020.23107

Karra, K., Kuzdeba, S., & Petersen, J. (2017). Modulation recognition using hierarchical deep neural networks. *2017 IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN 2017*, 1–3. https://doi.org/10.1109/DySPAN.2017.7920746

Keras. (2020). *About Keras*. https://keras.io/about/

Mallya, A. (2016). *Introduction to RNNs*. http://slazebni.cs.illinois.edu/spring17/lec02_rnn.pdf

Merco, R., Biron, Z. A., & Pisu, P. (2018). Replay Attack Detection in a Platoon of Connected Vehicles with Cooperative Adaptive Cruise Control. *Proceedings of the American Control Conference*, *2018-June*, 5582–5587. https://doi.org/10.23919/ACC.2018.8431538

O'Shea, T. J., Pemula, L., Batra, D., & Clancy, T. C. (2017). Radio transformer networks: Attention models for learning to synchronize in wireless systems. *Conference Record - Asilomar Conference on Signals, Systems and Computers*, 662–666. https://doi.org/10.1109/ACSSC.2016.7869126

Olatunji, O., Akinlabi, S., Madushele, N., Adedeji, P. A., & Felix, I. (2019). Multilayer perceptron artificial neural network for the prediction of heating value of municipal solid waste. *AIMS Energy*, *7*(6), 944–956. https://doi.org/10.3934/energy.2019.6.944

Patel, J., Das, M. L., & Nandi, S. (2018). On the Security of Remote Key Less Entry for Vehicles. *International Symposium on Advanced Networks and Telecommunication Systems, ANTS*. https://doi.org/10.1109/ANTS.2018.8710105

Riyaz, S., Sankhe, K., Ioannidis, S., & Chowdhury, K. (2018). Deep Learning Convolutional Neural Networks for Radio Identification. *IEEE Communications Magazine*, *56*(9), 146–152. https://doi.org/10.1109/MCOM.2018.1800153

Rizvi, S., Willet, J., Perino, D., Marasco, S., & Condo, C. (2017). A Threat to Vehicular Cyber Security and the Urgency for Correction. *Procedia Computer Science*, *114*, 100–105. https://doi.org/10.1016/j.procs.2017.09.021

Sakib, S., Ahmed, Jawad, A., Kabir, J., & Ahmed, H. (2018). *An Overview of Convolutional Neural Network: Its Architecture and Applications*. *November*. https://doi.org/10.20944/preprints201811.0546.v1

Selvakumar, M. S., Purushothkumar, R., Ramakrishnan, A., & Raja, G. (2018). Effective Cryptography Mechanism for Enhancing Security in Smart Key System. *2018 10th International Conference on Advanced Computing, ICoAC 2018*, 190–195. https://doi.org/10.1109/ICoAC44903.2018.8939095

Sommer, F., Dürrwang, J., & Kriesten, R. (2019). Survey and classification of automotive security attacks. *Information (Switzerland)*, *10*(4). https://doi.org/10.3390/info10040148

Tensorflow. (2020). *Why TensorFlow*. https://www.tensorflow.org/

The MathWorks. (2018). *Introducing Deep Learning with MATLAB*. https://www.mathworks.com/content/dam/mathworks/ebook/gated/80879v00_Deep_Learning_ebook.pdf

Wang, Y., Liu, M., Yang, J., & Gui, G. (2019). Data-Driven Deep Learning for Automatic Modulation Recognition in Cognitive Radios. *IEEE Transactions on Vehicular Technology*, *68*(4), 4074–4077. https://doi.org/10.1109/TVT.2019.2900460

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2020). Dive Into Deep Learning. *Journal of the American College of Radiology*. https://doi.org/10.1016/j.jacr.2020.02.005

Ziegler, J. L., Arn, R. T., & Chambers, W. (2017). Modulation recognition with GNU radio, keras, and HackRF. *2017 IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN 2017*, 1–3. https://doi.org/10.1109/DySPAN.2017.7920747

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 14/08/2020
Aceitação/Acceptance: 28/10/2020

Entornos Virtuales para Asistir el Aprendizaje de las Pruebas Software: Meta-Análisis para una Familia de Experimentos

Juan Ucán Pech[1], Omar S. Gómez[2], Julio Díaz Mendoza[1], Raúl Aguilar Vera[1 *]

juan.ucan@correo.uady.mx; ogomez@espoch.edu.ec; julio.diaz@correo.uady.mx; avera@correo.uady.mx

[1] Facultad de Matemáticas, Universidad Autónoma de Yucatán, C.P. 97000, Mérida, Yucatán, México.

[2] Facultad de Informática y Electrónica, Escuela Superior Politécnica de Chimborazo, Riobamba 060155, Ecuador.

*Autor para correspondencia.

Pages: 389-403

Resumen: El artículo reporta los hallazgos encontrados del Meta-Análisis realizado a una familia de experimentos, en dicho análisis se examina la efectividad en la detección de faltas en el código, con y sin apoyo de un Entorno Virtual Colaborativo (EVC). La Familia se integra por un experimento original (e1) y dos replicaciones (r1, r2) hechas en distintos tiempos. Como técnica de síntesis se estimaron los tamaños de efecto individual y global, empleando el enfoque de diferencia de medias ponderada, así como la determinación del grado de heterogeneidad en los experimentos de esta familia. Los resultados sugieren una equivalencia en la efectividad de la detección de defectos con respecto a usar o no usar un EVC, por lo que se sugiere que el uso de un EVC es una alternativa igualmente efectiva frente al modelo tradicional de verificación en colaboración (presencia física entre los miembros de un equipo).

Palabras-clave: Entornos Virtuales de Aprendizaje; Experimentación en Ingeniería de Software; Meta-Análisis; Pruebas de Software.

Intelligent Virtual Environments to Assist Software Testing Learning: Meta-Analysis for a Family of Experiments

Abstract: The article reports the findings of the Meta-Analysis carried out to a family of experiments, in this analysis the effectiveness in detecting errors in the code is examined, with and without the support of a Collaborative Virtual Environment (CVE). The Family is made up of an original experiment (e1) and two replications (r1, r2) made at different times. As a synthesis technique, the individual and global effect sizes were estimated, using the weighted mean difference approach, as well as the determination of the degree of heterogeneity in the experiments of this family.

The results suggest an equivalence in the effectiveness of defect detection with respect to using or not using an EVC, therefore it is suggested that the use of an EVC is an equally effective alternative to the traditional collaborative verification model (physical presence between members of a team).

Keywords: Meta-Analysis; Software Engineering Experimentation; Software Testing; Virtual Learning Environments.

1. Introducción

El estudio sobre el aprendizaje a través de Entornos Virtuales es un área de creciente interés de la Informática Educativa, particularmente para aquellos interesados en la línea del Aprendizaje Colaborativo Asistido por Computadora (Computer Supported Collaborative Learning: CSCL). Los entornos virtuales desarrollados bajo el paradigma del CSCL, se centran en el uso de las tecnologías de la información y de la comunicación, como herramientas de mediación en los métodos de colaboración de la instrucción, y tienen como finalidad, la construcción del conocimiento compartido a través de la interacción social de los miembros del grupo que lo integran (Stahl, Koschmann & Suthers, 2006).

En el ámbito de la Educación en Ingeniería de Software, particularmente en el área de la Programación, existen diversas propuestas de Entornos Virtuales Colaborativos diseñados exprofeso para facilitar su aprendizaje bajo un enfoque colaborativo (Bani-Salameg et al 2010; Esteves et al, 2011; Hupfer et al, 2004).

En el caso de la propuesta de Ucán (2015), de la cual se deriva la familia de experimentos que se analiza en el presente artículo, se estableció como objetivo de investigación, desarrollar un modelo para mejorar el proceso de aprendizaje de tareas vinculadas con la programación, particularmente, la detección de faltas comunes en la programación; así, utilizando el Entorno Virtual Colaborativo (EVC) descrito en (Ucán, 2015), los autores realizaron un conjunto de experimentos con la finalidad de explorar el impacto de utilizar un enfoque colaborativo con apoyo del EVC, en contraste con los esquemas tradicionales para revisión de código.

En el presente artículo se reportan los hallazgos derivados de una síntesis cuantitativa a la información obtenida de la familia de experimentos antes citada; dicho método de meta-análisis es utilizado para combinar de forma cuantitativa los resultados de estudios, replicaciones o experimentos similares, con la finalidad de generar nuevo conocimiento (Hedges & Olkin, 1985).

2. Características del Entorno Virtual Colaborativo

El EVC integra diversas prestaciones de los sistemas de trabajo en grupo (Wilson, 1991) y es complementado con un sistema experto como componente inteligente (Hernández, 2013); dicho entorno tiene como propósito asistir al aprendiz en la detección de faltas. La figura 1 presenta una vista arquitectónica —en tres capas— del EVC utilizado en la familia de experimentos.

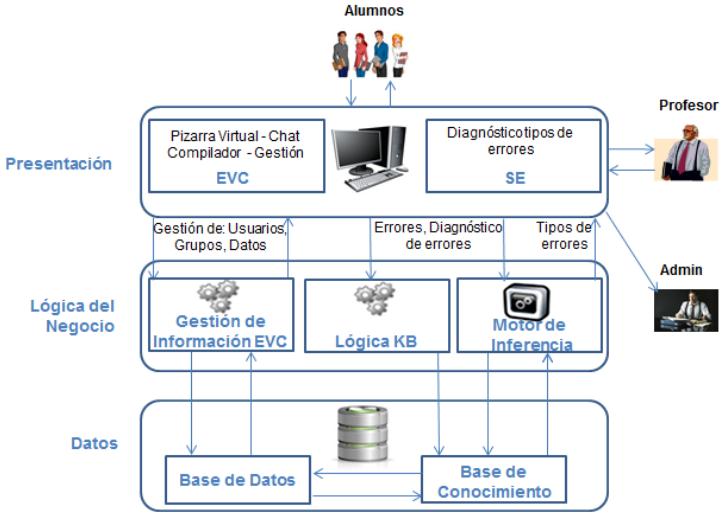


Figura 1 – Arquitectura del Entorno Virtual Colaborativo (Ucán, 2015)

De acuerdo con la capa de presentación, los estudiantes pueden visualizar —en una pizarra virtual— trozos de programas implementados bajo algún paradigma y deben encontrar las faltas incrustadas en dicho código, para dicha actividad de aprendizaje, podrán trabajar en grupos, de manera independiente, consultando al profesor, o incluso con el apoyo de un sistema experto integrado al EVC. Para el modelado del EVC, se utilizó la metodología UWE (*UML Web Engineering*) la cual contempla el modelado del análisis de requisitos, así como los modelos de contenido, navegación, representación y de proceso (Koch et al, 2008).

3. Familia de Experimentos

Los experimentos analizados en el presente reporte fueron realizados siguiendo un proceso de experimentación genérico (Gómez, O., Aguilar, R. y Ucan, J., 2019); dicha familia de experimentos consiste en un experimento original o primer experimento (E1) y dos replicaciones (R1) y (R2).

3.1. Contexto de los Experimentos

El experimento original o primer experimento (E1) se llevó a cabo en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán (UADY) en noviembre de 2014 con estudiantes que cursaban la asignatura Fundamentos de Programación de las Licenciaturas en Ingeniería de Software y Ciencias de la Computación —curso mixto; dicho experimento fue replicado en dos ocasiones (R1 y R2) en la misma institución durante el siguiente año, aunque se tuvo una modificación en el parámetro vinculado con la experiencia de los estudiantes.

La Tabla 1 presenta —para la familia de experimentos— el número de estudiantes participantes, el semestre que cursaban los estudiantes, así como el mes de ejecución del experimento.

Primer Exp. (E1)	Replicación 1 (R1)	Replicación 2 (R2)
46 estudiantes Primer semestre Noviembre de 2014	28 estudiantes Octavo semestre Mayo de 2015	23 estudiantes Sexto semestre Mayo de 2015

Tabla 1 – Familia de Experimentos

Todos los participantes en nuestra familia de experimentos, de acuerdo con la clasificación sobre experiencia en programación creada por Dreyfus y Dreyfus (1986) se pueden clasificar entre estudiantes principiantes y avanzados. Por su parte, la tipología de replicaciones para estos experimentos se acoge a las recomendaciones propuestas por Gómez (2012). En esta familia de experimentos se observa el proceso de aprendizaje de la programación en atención a la detección de faltas en programas —código— de software, en este sentido, las variables fueron las siguientes:

Variable dependiente: detección de faltas comunes de la programación.

- Desde el punto de vista operacional, la variable dependiente utilizó como métrica el porcentaje de faltas observadas en un programa (código).

Variable independiente (Factor): Estrategia de revisión de código

- Tratamiento 1: Sin el uso de una herramienta.
- Tratamiento 2: Uso del Entorno Virtual Colaborativo.

La *hipótesis de nulidad* planteada en esta familia se define como:

H_0. La efectividad medida como el porcentaje de faltas observadas es igual tanto para los grupos de participantes que emplearon el EVC como para los grupos de participantes que trabajaron Sin el EVC.

3.2. Diseño Experimental

La hipótesis de nulidad establecida en la sección previa se probó a través de las mediciones recabadas en la experimentación, dichas mediciones pertenecen a dos grupos de tratamientos, los grupos de sujetos que trabajaron con EVC y aquellos que trabajaron sin el EVC. En esta familia de experimentos se empleó un diseño cruzado con el fin de obtener un mayor número de observaciones (Kuehl, 2000), en un diseño cruzado las unidades experimentales, en este caso los grupos de sujetos conformados por tres personas reciben dos o más tratamientos y el orden de aplicación de los tratamientos está determinado por la estructura de este diseño. En particular, en cada experimento se empleó un diseño cruzado 2 x 2, dos tratamientos — sujetos Con EVC [C] y sujetos Sin EVCI [S]— en dos periodos distintos. Con dicha estructura, la mitad de los grupos de sujetos reciben el tratamiento C y en una sesión diferente reciben el tratamiento S,

mientras que la otra mitad reciben los tratamientos en orden inverso. Un inconveniente de este tipo de diseños es que los efectos de cierto tratamiento puedan extenderse más allá del periodo de aplicación —efecto remanente. Una estrategia para mitigar este inconveniente es no aplicar los tratamientos en periodos consecutivos; para esta familia de experimentos, en E1 se planificó un descanso de 4 días y para R1 y R2 un descanso de 3 días entre ambos periodos.

3.3. Ejecución del Experimento

La experimentación consistió en identificar faltas inyectadas intencionalmente en el código de dos programas de software escritos en un lenguaje de programación (para E1 en Lenguaje C y para R1 y R2 en Java), las faltas se tomaron como referencia de la clasificación propuesta en (Basili & Selby, 1987).

En relación con la asignación de las unidades experimentales a los tratamientos, se organizaron equipos de dos y tres estudiantes y se asignaron de forma aleatoria a alguna de las dos secuencias de tratamientos, tal como se ilustra en la Tabla 2.

	E1		R1		R2	
	C [EVC]	[S] EVC	C	S	C	S
Periodo 1 (Programa A)	8	8	5	4	4	4
Periodo 2 (Programa B)	8	8	4	5	4	4

Tabla 2 – Organización de Sujetos Experimentales en la Familia de Experimentos

En cada experimento de la familia, ambos grupos trabajaron con el mismo código; en el primer periodo unos equipos fueron provistos del empleo del EVC (Tratamiento C) para la identificación de las faltas, mientras que a los otros equipos se les proporcionó el programa de manera impresa, y por lo tanto, no utilizaron el EVC (Tratamiento S); en el segundo periodo los grupos de sujetos recibieron los tratamientos en orden inverso. Durante el experimento a todos los equipos se les proporcionó un documento impreso con las especificaciones del programa bajo análisis, y una plantilla impresa para el registro de la información referente a las faltas detectadas en el programa.

3.4. Análisis Experimental

En esta sección se presenta un análisis de los datos obtenidos en cada experimento, primero se presenta el estudio original con los análisis correspondientes, seguidamente se muestran los análisis de las dos réplicas de igual forma como se hizo en el experimento original.

3.4.1. Análisis del Experimento Base (Original, E1)

En la Tabla 3 se pueden apreciar las métricas obtenidas de la efectividad en la detección de defectos con respecto a los tratamientos, programas y secuencia de los tratamientos

del primer experimento; como se observa, la mayoría de los participantes detectó el mismo número de faltas en ambos tratamientos. En lo que respecta a los programas empleados, en promedio los sujetos lograron observar un mayor número de faltas en el programa B, mientras que los promedios en ambas secuencias parecen no tener diferencias significativas, por lo que pudiera intuirse la ausencia de efectos remanentes en los tratamientos.

		Efectividad
Tratamiento	[C]	43.45%
	[S]	45.33%
Periodo	Prog. A	37.88%
	Prog. B	50.90%
Secuencia	C→S	41.07%
	S→C	47.71%

Tabla 3 – Efectividad respecto a Tratamientos, Programas y Secuencia en E1

Para probar las hipótesis del estudio original, se analizó un modelo que se describe en la ecuación (1).

$$y_{ijk} = \mu + \alpha_i + b_{ij} + \gamma_k + \tau_d + \lambda_c + \varepsilon_{ijk} \quad (1)$$

Donde μ es el promedio general, αi es el efecto de la secuencia, b_{ij} es el efecto aleatorio para cada sujeto con promedio 0 y varianza σb2, γ_k es el efecto del periodo, τ_d es el efecto directo del tratamiento, λ_c es el efecto remanente, y ε_{ijk} es el error aleatorio independiente con promedio 0 y varianza σ^2. En la Tabla 4 se muestran los resultados de la prueba paramétrica del Análisis de Varianza (ANOVA) con respecto a la efectividad medida como el porcentaje de faltas observadas por los grupos de sujetos. Dados los componentes analizados con el ANOVA, se observa una diferencia significativa a un nivel α del 0.05 en el programa. Con respecto al tratamiento y al efecto remanente no se observan diferencias significativas.

Componente	**SC Parcial**	**gl**	**SM**	**F**	**p-value**
Tratamiento	290.22	1	290.22	0.9	0.3517
Periodo (programa)	1546.17	1	1546.17	4.78	0.0373
Efecto remanente	353.02	1	353.02	1.09	0.3051
Residuos	9058.20	28	323.51		

Tabla 4 – Resultados del ANOVA (Efectividad)

La medida de separación entre el tratamiento y el efecto remanente arroja un valor cercano al 30% (29.2893%), esta medida indica el grado de ortogonalidad entre ambos efectos. Un diseño cruzado 2 x 2 como el aquí usado está expuesto a un porcentaje bajo en la medida de separación. Se recomienda emplear este diseño cuando se intuya la ausencia de efectos remanentes. Dado que no se ha observado un efecto remanente significativo, los resultados del ANOVA pueden considerarse confiables. Para considerar válidos los resultados en E1, se deben verificar los supuestos de normalidad, homocedasticidad e independencia de los datos; los primeros dos supuestos fueron validados mediante las pruebas de Shapiro-Wilk y Levene, respectivamente; en el caso del tercer supuesto, se generaron gráficos de residuos vs. secuencia para evaluar visualmente; por restricciones de espacio no serán incluidos.

3.4.2. Análisis de la primera Replicación (R1)

La primera réplica continúa con el análisis de efectividad, eficiencia y costo promedio en la búsqueda de faltas con respecto a los tratamientos, programas y secuencia, sin embargo, esta prueba se realiza con nuevos sujetos. Como se indica en el contexto del experimento, en este caso se emplearon estudiantes avanzados de pregrado. La Tabla 5 presenta las observaciones acerca de la efectividad.

		Efectividad
Tratamiento	[C]	53.70%
	[S]	50.00%
Periodo	Prog. A	61.10%
	Prog. B	42.59%
Secuencia	C→S	54.16%
	S→C	50.00%

Tabla 5 – Efectividad respecto a Tratamientos, Programas y Secuencia en R1

Como se observa en la Tabla 5, los participantes detectaron un porcentaje ligeramente mayor de faltas con el uso del EVC [C] que sin uso del EVC [S]. Por lo que respecta a los programas empleados, en promedio los sujetos lograron observar un mayor número de faltas en el programa A, mientras que los promedios en ambas secuencias parecen no tener diferencias sustanciales por lo que pudiera intuirse la ausencia de efectos remanentes en los tratamientos.

Para probar las hipótesis de R1, se utilizó y analizó el mismo modelo que se describió la ecuación (1). En la Tabla 6 se muestran los resultados del análisis de varianza con respecto a la efectividad medida como el porcentaje de faltas observadas por los grupos de sujetos. Dados los componentes analizados en la ANOVA, se observa una diferencia significativa a un nivel α del 0.05 en el programa. Con respecto al tratamiento y al efecto remanente no se observan diferencias significativas.

Componente	SC Parcial	gl	SM	F	p-value
Tratamiento	222.20	1	222.20	0.98	0.3384
Periodo (programa)	1209.89	1	1209.89	5.35	0.0365*
Efecto remanente	77.15	1	77.15	0.34	0.5685
Residuos	3166.57	14	226.18		

Tabla 6 – Resultados del ANOVA en R1

3.4.3. Análisis de la segunda Replicación (R2)

En forma similar al experimento original y al primer experimento, en la segunda réplica se analizó la variable efectividad en la búsqueda de faltas con respecto a los tratamientos, programas y secuencia.

		Efectividad
Tratamiento	[C]	45.83%
	[S]	56.25%
Periodo	Prog. A	60.41%
	Prog. B	35.41%
Secuencia	C→S	41.66%
	S→C	54.16%

Tabla 7 – Efectividad respecto a Tratamientos, Programas y Secuencia en R2

Como podemos observar en la Tabla 7, los participantes detectaron un porcentaje ligeramente menor de faltas con el uso del EVC [C] que sin uso del EVCI [S]. Similar a R1, en lo que respecta a los programas empleados, en promedio los sujetos lograron observar un mayor número de faltas en el programa A, mientras que los promedios en ambas secuencias parecen no tener diferencias sustanciales por lo que pudiera intuirse la ausencia de efectos remanentes en los tratamientos.

En la Tabla 8 se muestran los resultados del análisis de varianza con respecto a la efectividad. De acuerdo con los componentes analizados, se mantiene la observación de una diferencia significativa a un nivel α del 0.05 en el tratamiento, con respecto al programa y al efecto remanente se continúa sin observar diferencias significativas.

Componente	SC Parcial	gl	SM	F	p-value
Tratamiento	1701.19	1	1701.19	4.82	0.0485*
Periodo (programa)	0.00	1	0.00	0.00	1.0000
Efecto remanente	1406.25	1	1406.25	3.98	0.0691
Residuos	4235.89	12	352.99		

Tabla 8 – Resultados del ANOVA en R2

4. Síntesis Cuantitativa de la Familia de Experimentos

La síntesis cuantitativa es un método de meta-análisis utilizado para combinar de forma cuantitativa los resultados de estudios, replicaciones o experimentos similares para generar nuevas piezas de conocimiento, las cuales serán más generales y fiables en virtud de que se encuentran sustentadas por una mayor cantidad de evidencia empírica (Hedges & Olkin, 1985).

De manera general, el meta-análisis consiste en:

1. estimar el tamaño del efecto de un par de tratamientos por cada uno de los experimentos que se desean examinar,
2. estimar un tamaño de efecto global a partir de los tamaños de efecto estimados individualmente, y
3. estimar el nivel de heterogeneidad del conjunto de tamaños de efecto observados.

Para poder aplicar el método de síntesis cuantitativa es necesario calcular los promedios, tamaños de muestra y desviaciones estándar de los pares de tratamientos de los diferentes estudios que conformarán el meta análisis. La Tabla 9 presenta los promedios, tamaños de muestra y desviaciones estándar calculadas con base en los datos de la familia de experimentos descrita en secciones previas; como se puede apreciar, los tamaños de muestra empleados en los experimentos de esta familia pareciera ser una limitación, no obstante, en el ámbito de la experimentación en ingeniería de software, suele ser una limitación contar con suficientes sujetos en la realización de experimentos, es por ello que de manera paulatina comienzan a desarrollarse familias de experimentos en los que se aplican técnicas de síntesis cuantitativa para minimizar esta limitación (Santos, Gómez & Juristo, 2018).

Estudio	n.evc	avg.evc	sd.evc	n.sevc	avg.sevc	sd.sevc
E1	16	43.45%	20.03	16	45.33%	17.70
R1	9	53.70%	20.03	9	50.00%	14.43
R2	8	45.83%	21.36	8	56.25%	23.47

Tabla 9 – Estadísticos descriptivos de la familia de experimentos

4.1. Estimación de tamaños de efecto individual

El tamaño del efecto es un índice que se utiliza para medir las diferencias entre dos variables, en este caso un par de tratamientos.; el tamaño del efecto permite conocer qué tanto es mejor un tratamiento frente a otro. Para calcular el tamaño del efecto de cada replicación de la partición seleccionada se utiliza la técnica de Glass (1976), conocida como diferencia de medias ponderada WMD (*Weighted Mean Difference*). En la Tabla 10 se muestran los tamaños de efectos calculados en cada estudio, así como el intervalo de confianza en el que se encuentran estos estimados; en nuestro caso, los tamaños de efecto se han calculado entre usar un EVC con respecto a no utilizarlo. Tamaños de efecto negativos indican un tamaño de efecto en favor de no usar un EVC, mientras que tamaños de efecto positivo indican un tamaño de efecto a favor del uso de un EVC.

Estudio	MD	95% - CI
E1	-1.8800	[-14.9774; 11.2174]
R1	3.7000	[-12.4283; 19.8283]
R2	-10.4200	[-32.4106; 11.5706]

Tabla 10 – Tamaños de efecto estimados en los experimentos de la familia

4.2. Estimación del tamaño del efecto general

Una vez estimados los tamaños de los efectos y sus intervalos de confianza, a continuación se combinan estos efectos para obtener un efecto global de los estudios en cuestión. El efecto global es un promedio ponderado del conjunto de tamaños de efecto estimados; a cada tamaño de efecto estimado se le asigna un determinado nivel de ponderación o peso de acuerdo con el nivel de precisión que éste tiene. Los tamaños de efecto que tienen mayor nivel de precisión se les asigna mayor peso, cabe mencionar que la precisión en el tamaño de efecto está en función, principalmente, del tamaño de muestra utilizado en el experimento. A mayor tamaño de muestra los intervalos de confianza del tamaño de efecto estimado se reducen, por lo que el tamaño de efecto estimado se aproxima mucho más al tamaño de efecto verdadero. Por otra parte, a menor tamaño de muestra los intervalos de confianza del tamaño de efecto estimado se amplían, por lo que es más probable que el tamaño de efecto estimado se aproxime menos al tamaño de efecto verdadero. Como se observa en la Tabla 11, E1 ha recibido un grado mayor de ponderación (49.6%); este experimento es el que cuenta con mayor tamaño de muestra, por el contrario, R2 ha recibido la menor ponderación (17.6) y contiene un tamaño de muestra menor, aunque los intervalos de confianza son más anchos.

Estudio	n	MD	95% - CI	%W
E1	16	-1.8800	[-14.9774; 11.2174]	49.6
R1	9	3.7000	[-12.4283; 19.8283]	32.7
R2	8	-10.4200	[-32.4106; 11.5706]	17.6

Tabla 11 – Ponderación y Tamaño de muestra en los experimentos de la familia

Dos de los modelos estadísticos más utilizados para calcular el efecto global de un conjunto de estudios son, el modelo de efectos fijos y el modelo de efectos aleatorios, la elección de un modelo u otro depende del supuesto que se tenga acerca de la distribución de los tamaños de los efectos que provienen del conjunto de replicaciones, en este caso de la partición (Hedges & Olkin, 1985). En el modelo de efectos fijos, se supone que todas las réplicas —de la partición— comparten el mismo tamaño de efecto verdadero. En otras palabras, supone un tamaño de efecto verdadero que es fijo y que subyace a todas las repeticiones, de modo que las diferencias observadas en los tamaños de efecto estimados son producto del error de muestreo y no el producto de la influencia de otros

factores o de las condiciones del experimento. En este modelo, el efecto global es un estimador del tamaño de efecto verdadero que es común en las replicaciones de la partición. Por el contrario, en el modelo de efectos aleatorios se asume que el tamaño de efecto verdadero puede variar de un experimento a otro, es decir, en las replicaciones de la partición pueden existir condiciones que influyen en el tamaño del efecto verdadero. El modelo se asume que los tamaños de efecto verdaderos se distribuyen de acuerdo con la distribución normal, y el efecto global es entonces un estimador del promedio de la distribución de tamaños de efecto verdaderos.

En la práctica, es muy poco probable tener replicaciones exactas de algún experimento porque siempre existirán condiciones o factores que varían entre la(s) replicación(es) y el experimento base (Gómez, 2012). Al haber en las replicaciones cambios tanto deliberados como no deliberados, los tamaños de efecto verdaderos varían entre las replicaciones. Ante a esta situación, el modelo estadístico que utilizamos para calcular el efecto global en nuestra familia de experimentos es el modelo de efectos aleatorios.

Como se observa en la Tabla 12, el tamaño de efecto global estimado se encuentra entre los tamaños de efecto -10.7856 y 7.6715; de acuerdo con el intervalo de confianza estimado, el tamaño de efecto global verdadero puede ser 0, por lo que estos resultados sugieren que la efectividad en la detección de defectos es equivalente con el uso de un ECV o sin el uso de éste.

Estudio	MD (Global)	95% - CI	Z	p-value
Modelo de Efectos Aleatorios	-1.5571	[-10.7856; 7.6715]	-0.33	0.7409

Tabla 12 – Tamaño de efecto global en esta familia de experimentos

Los resultados antes descritos pueden también ser representados, a través del uso de un diagrama de bosque (*forest plot*), como se muestra en la figura 2.

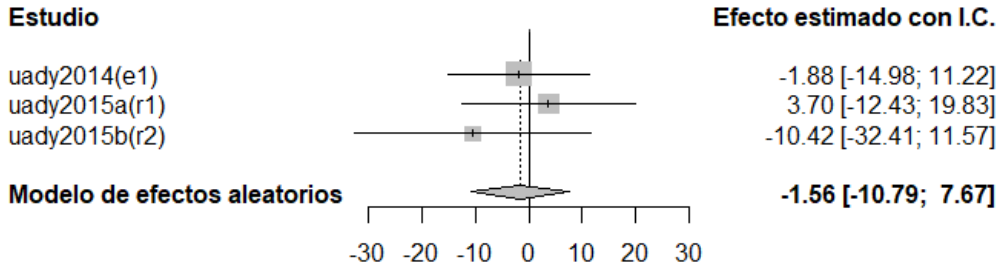


Figura 2 – Diagrama de bosque con los efectos estimados en esta familia de experimentos

4.3. Grado de Heterogeneidad

La prueba de heterogeneidad utiliza el índice Q para evaluar si existe heterogeneidad estadísticamente significativa en un conjunto de tamaños de efecto estimados, esta

prueba parte de la hipótesis nula de la existencia de un tamaño de efecto que es común en los tamaños de efecto de un conjunto de experimentos. Bajo la hipótesis nula Q sigue una distribución χ2 (chi-cuadrada) con k – 1 grados de libertad donde k es el número de experimentos; el no rechazar la hipótesis nula asume que los efectos estimados son homogéneos, es decir, que no hay variabilidad en los resultados.

No obstante, esta prueba estadística tiene una potencia baja para detectar una verdadera heterogeneidad cuando se tiene un conjunto reducido de experimentos (Borenstein et al, 2009). Debido a esta situación se utiliza también el estadístico T2 descrito anteriormente, que consiste en estimar la varianza inter-estudios de un conjunto de experimentos. El otro estadístico que también se utiliza es el índice I2 (Higgins & Thompson, 2002) que se interpreta como el porcentaje de variabilidad total en un conjunto de tamaños de efecto estimados, que se debe a una verdadera heterogeneidad. Los autores proponen tres valores de referencia para este índice, sugieren que valores alrededor de 25%, 50% y 75% representan respectivamente un nivel de heterogeneidad bajo, medio y alto.

Como se observa en la Tabla 13, el valor p resultante no es significativo, sugiriendo una homogeneidad en los resultados de la familia de experimentos. Complementariamente también se han estimado los valores de Tau^2 e I^2 obteniendo 0, y 0.0%, respectivamente. Sugiriendo también una homogeneidad entre los tamaños de efecto de los experimentos de esta familia.

Q	df	p-value
1.03	2	0.5962

Tabla 13 – Prueba de Heterogeneidad

5. Discusión

De acuerdo con los hallazgos aquí reportados, los resultados en los tres experimentos sugieren una equivalencia con respecto a la detección de defectos tanto para los grupos de participantes que trabajaron de forma colaborativa virtual (con el EVC) como para aquellos que trabajaron de manera tradicional. Para este tipo de tarea estudiada —detección de faltas en el código— los resultados arrojados sugieren que el uso de esta herramienta resulta igualmente de efectiva que trabajar de forma colaborativa bajo un enfoque tradicional —colaboración e interacción presencial entre miembros de los equipos.

Al llevar a cabo la síntesis cuantitativa, el efecto global estimado sugiere una efectividad equivalente entre ambos enfoques de colaboración —tamaño de efecto global no significativo— a través de un EVC y de modo tradicional; estos resultados son soportados por la prueba de heterogeneidad la cual sugiere que los tamaños de efecto estimados de manera individual —en los tres experimentos— son consistente entre sí.

En cuanto a las limitaciones del estudio, con respecto a las amenazas de validez interna, en los tres experimentos las sesiones se llevaron a cabo acorde a lo planificado, sin presentar algún incidente —efecto de historia— y dado que se planificó un periodo de descanso

entre la aplicación de tratamientos, se puede asumir un efecto de maduración mínimo. El efecto de sesgo por selección se ha reducido debido al mecanismo de aleatorización empleado en el diseño experimental. Durante las sesiones del experimento se observó interés por parte de los estudiantes en realizar las actividades por lo que también puede asumirse un efecto de desmoralización mínimo. Por otro lado, en lo referente a las amenazas de validez externa los resultados aquí presentados son generalizables a estudiantes con características similares a los estudiantes empleados en el experimento aquí reportado, gracias a las replicaciones efectuadas, el tamaño de la muestra se ha incrementado, así también, la confiabilidad observada en los resultados consistentes entre los experimentos de la familia.

6. Conclusiones

Con base en los resultados del meta-análisis para la familia de experimentos analizada, se puede concluir que existe equivalencia en la efectividad para la detección de faltas en el código, entre los grupos de estudiantes que trabajaron de forma colaborativa virtual (con el EVC), y aquellos que trabajaron de manera tradicional (sin el EVC).

Adicionalmente, algunas de las ventajas observadas tras utilizar el EVC son las siguientes:

1. No se requiere que los estudiantes estén situados en un mismo espacio físico, lo cual, para la educación mixta o para la virtual, la disponibilidad de un EVC resulta apropiada; ni que decir para escenarios como el de la pandemia del COVID-19, en la cual las sesiones presenciales han tenido que migrar a sesiones virtuales.
2. La prestación de una pizarra virtual en el EVC, dicho componente tiene integrado un IDE para escribir y compilar trozos de código de un lenguaje de programación en forma colaborativa, bajo el enfoque tradicional de aprendizaje, dicho escenario es imposible de generar.
3. La prestación del componente inteligente; dicho módulo —sistema experto— ha sido desarrollada para asistir la interacción entre los estudiantes durante actividades de aprendizaje de la programación, así como asistir a los estudiantes durante la identificación de defectos en una serie de programas instrumentados.

Como indican los hallazgos, tener una familia de experimentos permite corroborar los hallazgos anteriores, lo que aumenta la confiabilidad de los resultados. En futuros estudios, tenemos la intención de desarrollar más repeticiones en otros contextos para comprender mejor los límites externos para mantener resultados consistentes con nuevos tipos de réplicas del experimento original.

Agradecimientos

Agradecemos el apoyo brindado por la Secretaría de Educación Pública (México) a través del proyecto P/PROFEXCE-2020-31MSU0098J-13.

Referencias

Bani-Salameh, H., Jeffery, C. & Al-Gharaibeh, J. (2010) A Social Collaborative virtual environment for software development. *Proceedings of the International Symposium on Collaborative Technologies and Systems*. Chicago, IL. DOI: 10.1109/CTS.2010.5478525

Basili, V. & Selby, R. (1987) Comparing the effectiveness of software testing strategies. *IEEE Transactions on Software Engineering*. Vol. SE-13. No. 12. 1278-1296. DOI: 10.1109/TSE.1987.232881

Borenstein, M., Hedges, L. V., Higgins, J. & Rothstein, H. (2009) *Introduction to MetaAnalysis*. John Wiley & Sons, Ltd, United Kingdom.

Dreyfus, H. & Dreyfus, S. (1986) *Mind over Machine. The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Basil Blackwell.

Esteves, M. , Fonseca, B. , Morgado, L. & Martins, P. (2011) Improving teaching and learning of computer programming through the use of the Second Life virtual world. *British Journal of Educational Technology*. https://doi.org/10.1111/j.1467-8535.2010.01056.x

Glass, G. (1976) Primary, secondary, and meta-analysis of research. *Educational Researcher*. Vol. 5(10), 3–8.https://doi.org/10.3102%2F0013189X005010003

Gómez, O. (2012) *Tipología de Replicaciones para la Síntesis de Experimentos en Ingeniería del Software*. (Tesis Doctoral). Universidad Politécnica de Madrid, Madrid: España.

Gómez, O., Aguilar, R. y Ucán, J. (2019) Experimentación en Ingeniería de Software. En Aguilar, R. (Editor). *Ingeniería de Software en México: Educación, Industria e Investigación* (pp. 205-230). CDMX México: Academia Mexicana de Computación.

Hedges, L. & Olkin, I. (1985) *Statistical Methods for Meta-analysis*. Academic Press, Orlando:USA.

Hernández G. (2013) *Análisis del uso de la inteligencia colaborativa como herramienta para la construcción de bases de conocimiento consensuadas en procesos de diagnóstico médico*. (Tesis Doctoral). Universidad Carlos III de Madrid. Madrid: España.

Higgins, J. & Thompson, S. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. Vol. 21(11), 1539–1558. https://doi.org/10.1002/sim.1186

Hupfer, S., Cheng, L., Ross, S. & Patterson, J. (2004) Introducing Collaboration into an Application Development Environment. *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. https://doi.org/10.1145/1031607.1031611

Koch, N, Knapp A., Zhang, G., and Baumeister, H. (2008) UML-based web engineering. In *Web Engineering: Modelling and Implementing Web Applications*. Springer, London.

Kuehl, R. (2000) *Design of experiments: statistical principles of research design and analysis*. Duxbury/Thomson Learning.

Santos, A., Gómez, O. & Juristo, N. (2018) *Analyzing Families of Experiments in SE: A Systematic Mapping Study*. https://arxiv.org/abs/1805.09009. Online; accessed 22 May 2020.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), Cambridge handbook of the learning sciences (pp. 409-426). Cambridge, UK.

Ucán, J. (2015) *Aprendizaje de la Programación Asistido con Entornos Virtuales Colaborativos Inteligentes*. (Tesis Doctoral). Universidad del Sur, Mérida: México.

Wilson, P. (1991) *Computer supported cooperative work: An introduction*. Springer.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 16/09/2020
Aceitação/Acceptance: 05/11/2020

Algoritmo de optimización de colonia de hormigas aplicado a TSP, una revisión sistemática

Francisco Obando-Vidal[1, a], Néstor Díaz-Mariño[2], Ember Martínez-Flor[3]

fobando@unicomfacauca.edu.co; nediaz@unicauca.edu.co; eumartinez@unicauca.edu.co

[1] Tic Unicomfacauca, Faculta de ingenierías, Corporación Universitaria Comfacauca - Unicomfacauca (Popayán-Cauca, Colombia)

[a] Estudiante de maestría en computación, FIET, Universidad del Cauca, (Popayán-Cauca, Colombia)

[2] Grupo de inteligencia Computacional – GICO, FIET, Universidad del Cauca, (Popayán-Cauca, Colombia)

Pages: 404-417

Resumen: El algoritmo de optimización de colonia de hormigas (ACO) es utilizado en diferentes problemas de optimización combinatoria; este algoritmo está basado en el comportamiento de las hormigas, sin embargo, solo se utilizan algunas de las características observadas en el comportamiento de la hormiga natural. El objetivo de la revisión es conocer las diferentes implementaciones de ACO para dar solución al problema del agente viajero (TSP) y con base a esto establecer una base para la creación de una nueva variante del ACO que incluya características de las hormigas arrieras o cultivadoras de hongos.

Palabras-clave: ACO; Revisión sistemática; Algoritmo; TSP; optimización Combinatoria

A systematic review of the state of the art of the ant colony optimization algorithm applied to TSP.

Abstract: The ant colony optimization algorithm (ACO) is used in different combinatorial optimization problems; This algorithm is based on the behavior of ants, however, only some of the characteristics observed in the behavior of the natural ant are used. The objective of the review is to know the different implementations of ACO to solve the traveler salesman's problem (TSP) and based on this, establish a basis for the creation of a new variant of ACO that includes characteristics of the leafcutter ants or mushroom cultivator ants.

Keywords*:* ACO; Systematic review; algorithm; TSP; combinatorial optimization

1. Introducción

Los algoritmos de optimización buscan obtener mejores resultados a diferentes problemas de optimización combinatoria, entre estos, se encuentra el problema del vendedor viajero (TSP siglas en inglés) que se refiere al conjunto de ciudades que un

vendedor debe visitar una a una (sin pasar por la misma ciudad) completando todo el recorrido obteniendo el menor costo posible (Tito Chura, Silva Delgado, Alfaro Gonzales, & Fajardo Espinoza, 2019).

Una metaheurística que se destaca por sus resultados en la solución a este problema es el algoritmo de optimización de colonia de hormigas (ACO siglas en inglés). ACO está basado en el comportamiento de hormigas reales para encontrar el camino más corto por medio del rastro de una sustancia llamada feromona, que es depositada por la hormiga durante su recorrido. Para TSP cada hormiga de ACO genera un recorrido escogiendo ciudades de acuerdo con una regla de transición, donde las hormigas prefieren desplazarse a ciudades que están conectadas a caminos de bajo costo y con un mayor rastro de feromona. A este proceso se le conoce como "Construcción de la solución".

La revisión sistemática que se presenta busca identificar las diferentes variantes de ACO, y a partir de estas plantear una nueva solución ACO con hormigas artificiales con capacidad de distinguir buenas soluciones que sirvan a la colonia en el proceso de cultivar el hongo para intensificar la búsqueda esperando encontrar la mejor solución en el menor tiempo. Esta nueva variante busca complementar el ACO tradicional, incorporando un nuevo concepto natural característico de la subespecie de hormiga denominada *Attini* (Fernández, Valentina, & Francisco, 2015)(Leal, Wirth, & Tabarelli, 2014). Las Attini son cultivadoras de hongos y cuentan con la capacidad de identificar alimento perjudicial para la colonia.

En el estado del arte se encuentras diferentes resúmenes y revisiones de literatura como en (Cordon, Herrera, & Stiitzle, 2002), que presentan la definición de ACO junto con las diferentes modelos para su implementación, adicional presenta la relación con otras metaheurísticas y cuáles son los diferentes campos de aplicación entre ellos TSP (1989 a 2002). El mayor exponente de ACO es (Dorigo & Blum, 2005) presenta un resumen de los diferentes problemas donde se puede aplicar ACO, sugiere a los investigadores nuevos métodos y tendencias que hasta 2005 fueron presentados en este campo (1992 a 2004).

Por otro lado en (Chandra Mohan & Baskaran, 2012), realizan una revisión de ACO considerando trabajos publicados hasta el año 2012, entre ellos surgen las estrategias paralelas enfocadas a la construcción de la solución de ACO, híbridos con otras metaheurísticas, además de las posibles tendencias para solución de diferentes problemas de ACO, y proponen un nuevo ACO para Network simulator 2 (NS2), presentado mejores resultados a los obtenidos en la literatura para el periodo de observación (2008 a 2011). En (Rave & Álvarez, 2013), realizan una revisión del problema del vendedor viajero (TSP), con una ventana de observación del año 2005 al 2012, donde se describen métodos e investigaciones relevantes entre ellas el uso metaheurísticas como ACO, sin embargo, no se presenta una revisión enfocada en la solución de TSP solo con ACO

2. Método

Para la revisión sistemática de literatura (en adelante RSL), se utilizó como metodología la planteada en (Kitchenham, Pearl Brereton, Budgen, et al., 2009). Esta metodología

permite definir los objetivos de la RSL aportando una guía en el proceso de exploración del estado del arte de ACO, permitiendo verificar si se incluyen otros comportamientos de las hormigas naturales, en específico de la hormiga arriera o cultivadora de hongos (subespecie Attini). Se procura analizar específicamente las diferentes variantes de ACO más citados dentro del estado del arte que se hayan aplicado al problema del agente viajero o TSP.

La RSL se inicia estableciendo las preguntas de investigación basándose en una previa revisión de literatura como lo establece (Levy & Ellis, 2006). Posteriormente se realiza una obtención de la palabras claves según las preguntas de investigación como es sugerido en (Kitchenham, Pearl Brereton, Budgen, et al., 2009). Cómo paso siguiente se aplican criterios de exclusión e inclusión sobre los documentos obtenidos. Finalmente se realiza la comparación de diferentes estudios, se eliminan los repetidos y se exponen los resultados del RSL. La búsqueda de documentos se realizó en bases de datos digitales como Scopus, ScienceDirect y ACM digital.

2.1. Planeación RSL

El objetivo de la RSL es describir el estado del arte de ACO, donde se identifiquen las variantes usadas para dar solución al problema de TSP. Para ellos se requiere identificar los conceptos que representen la problemática, y dar forma al problema en cuestión. A continuación, se describen los pasos necesarios para desarrollar la metodología.

2.1.1. Preguntas de investigación

Para dar solución al objetivo se plantean las preguntas de investigación de la Tabla 1 de esa forma seleccionar y agrupar los diferentes estudios acerca del ACO aplicado a TSP, además para identificar la utilización previa del comportamiento natural de las hormigas cultivadoras de hongos.

Identificador	Pregunta para la RSL
P1	¿Cuáles algoritmos de optimización de colonia de hormigas permite resolver el problema del vendedor viajero (TSP)?
P2.	¿Qué tipos de algoritmos de optimización de colonia de hormigas existen que permiten resolver TSP?
P3.	¿Qué algoritmos de optimización de colonias hormigas incorporan características naturales de hormigas cultivadoras de hongos?

Tabla 1 – Preguntas para la Revisión sistemática

2.1.2. Cadenas de búsqueda

El siguiente paso fue obtener las palabras claves de las preguntas de la RSL como lo expone (Kitchenham, Pearl Brereton, Budgen, et al., 2009), con la finalidad de construir las cadenas de búsqueda (CB). Se usan conectores lógicos “AND” y “OR”, por cada pregunta como muestra la Tabla 2.

Las CB se ejecutaron sobre bases de datos digitales cómo Scopus, ScienceDirect y ACM Digital; En CB1 se busca identificar cuáles han sido las soluciones de ACO para el problema de TSP, en CB2 determinar cuáles son los diferentes tipos de ACO para el problema de TSP en los últimos años y en CB3 determinar si existen soluciones de ACO que incorporen conceptos o características de hormigas cultivadoras de hongos. Cada cadena fue adaptada según las opciones avanzadas de búsqueda cada que brindan los motores de las fuentes seleccionada.

Identificador	CB por Pregunta
CB1	(“ant colony optimization” OR aco OR “ant colony*” OR “ant system”) AND (“travel salesman problem” OR “travelling salesman problem”OR TSP)
CB2.	(“ant colony optimization” OR aco OR “ant colony” OR “ant system”) AND (“travel salesman problem” OR “travelling salesman problem” OR tsp) AND (types OR trends OR variant)
CB3.	(Atta OR attini OR Ant-Fungus OR Garden OR LeafCutter OR leaf-cuttimg OR Fungus- Gardening OR farming) AND (“ant colony optimization” OR aco OR “ant colony” OR “ant system”)

Tabla 2 – Cadena de búsqueda por pregunta de RSL

2.1.3. Ejecución de CB

Para la búsqueda de la información, se usaron las bases de datos digitales mencionadas anteriormente, la Tabla 3 presenta la configuración realizada por cada motor de búsqueda aplicando criterios de inclusión y exclusión para la RSL. Se obtuvo en total 976 trabajos, 576 relacionados con P1, 344 relacionados con P2 y 59 con P3, la Tabla 4 muestra los resultados obtenidos por cada CB. Después de obtener los estudios de cada una de las bases de datos se procede a realizar la revisión aplicando criterios de inclusión y exclusión (Kitchenham, Pearl Brereton, Budgen, et al., 2009).

Fuente	Configuración Búsqueda Avanzada
Scopus	TITLE-ABS-KEY PUBYEAR > AAAA LIMIT-TO LANGUAGE
Science Direct	Title,abstract or author-specified keywords Year AAAA-AAAA
ACM Digital	Publication Title: Abstract: Keywords: Publication Date:

Tabla 3 – Configuración de búsqueda avanzada por base digital.

Fuentes	Resultado CB1	Resultado CB2	Resultado CB3
Scopus	475	308	39
ScienceDirect	66	22	2
ACM Digital	5	0	1
	20	5	15
	7	0	2

Tabla 4 – Resultados de Cadenas de búsqueda por pregunta.

2.2. Proceso de revisión

Después de obtener los estudios de cada una de las bases de datos se procede a realizar la revisión aplicando criterios de inclusión y exclusión como lo indica (Kitchenham, Pearl Brereton, Budgen, et al., 2009), el cual se muestra en la Figura 1.

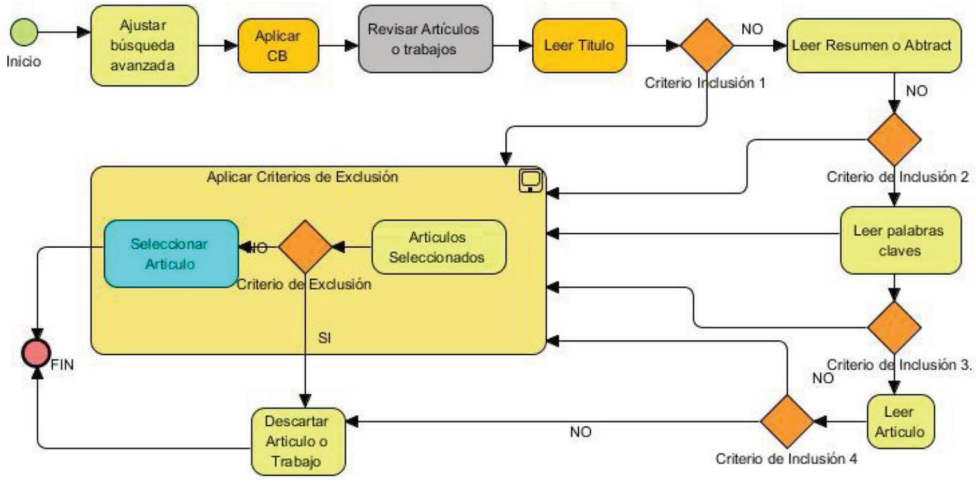


Figura 1 – Proceso de la revisión. Adaptado de (Kitchenham, Pearl Brereton, Budgen, et al., 2009).

2.2.1. Criterios de inclusión y exclusión

Los criterios de inclusión de trabajos se aplicaron en el siguiente orden, Ci1: El título de los trabajos que evidencie relación con la investigación. Ci2: Resumen (abstract) de los trabajos previamente seleccionados con relación al problema seleccionado. Ci3: Palabras claves de los trabajos evidencie la investigación relacionada Ci4: Trabajos que presenten comparación de diferentes implementaciones de ACO.

Los criterios de exclusión de los trabajos son los siguientes: Ce1: Más de 5 años en artículos científicos, excepto para la cadena de búsqueda CB3 que se amplió a 10 años para descartar posibles implementaciones con el tema en cuestión. Ce2: Referencias que no estén en inglés o español. Ce3: Referencias que no se tenga acceso completo. Ce4: Eliminar estudios repetidos. Ce5: Se excluirán aquellos artículos que no hayan sido citados excepto para la cadena de búsqueda Cb3. Ce6: Se excluyen todos los trabajos que no apliquen los criterios de inclusión.

2.3. Evaluación de calidad

La calidad de los trabajos seleccionados se evaluó respecto a 11 preguntas para determinar estudios de calidad (Dybå & Dingsøyr, 2008), las cuales buscan que estos cumplan con tres criterios de calidad necesarios. En el primero el Rigor se refiere si el estudio siguió un método de investigación adecuado, el segundo la Credibilidad que indica si los resultados son significativos y bien presentado. Y el tercer criterio la Relevancia, se refiere si los resultados obtenidos son útiles.

2.4. Recopilación de datos

Para extraer la información se aplicación los criterios de inclusión/exclusión y los criterios de evaluación de calidad. Descartando por título, palabras clave y resumen un total de 367 estudios, los artículos repetidos aplicados a las 3 preguntas formuladas fueron en total 228 y aquellos que nos cumplían con el tema principal de ACO, TSP y no cumplía los criterios de calidad en total 295 estudios. En total se seleccionaron 84 estudios distribuidos como muestra la Tabla 5 para las preguntas P1 - P2 y la Tabla 6 de la pregunta P3.

Pregunta	2019	2018	2017	2016	2015	2014	Total
P1	8	7	7	6	7	1	**36**
P2	2	3	1	0	1	0	**7**

Tabla 5 – Estudios seleccionados por año preguntas P1 y P2.

Pregunta	2019 -2018	2017-2016	2015-2014	2013-2012	2011 - 2010	2009	Total
P3	2	1	1	1	2	1	**8**

Tabla 6 – Estudios seleccionados por año preguntas P3.

La información de los estudios se condenso en un archivo de Excel, para mejorar la organización de los estudios primarios información, la estructura que se muestra en la Figura 2.

Titulo	Tipo de publicación	Autores	Resumen Temas importantes	Palabras Claves	Año	Idioma	Base de datos Digital	Enlace base de datos Revista Digital	Pregunta de investigación

Figura 2 – Estructura del archivo de información seleccionada

3. Resultados

Los estudios de investigación identificados como fuente de información primaria fueron 48 siguiendo el proceso de revisión de la Figura 1, el proceso detallado se muestra en Tabla 6, para cada una de las preguntas de investigación.

Paso	Estudios Prelimares	Actividad	Estudios Seleccionados
1	P1: 572 P2: 344 P3: 59 **Total: 976**	Identififcar estudios relavantes según criterios de Inclusión	P1: 188 P2: 82 P3: 20 **Total: 290**
2	P1:188 P2:82 P3:20 **Total:290**	Estudios que se les aplico el proceso de exclusión	P1: 47 P2: 18 P3: 8 **Total: 73**
3	P1: 47 P2: 22 P3: 8 **Total: 73**	Evaluación de calidad	P1: 34 P2: 7 P3: 7 **Total:48**

Tabla 6 – Proceso detallado selección de estudios

De la pregunta 1 se seleccionaron 34 estudios primarios de diferentes implementaciones de ACO para dar solución al problema del agente viajero, como soluciones para mejorar la evaporación y depósito de la feromona, nuevos métodos para realizar la construcción de la solución por medio de heurísticas y modelos matemáticos, soluciones combinadas con metaheurísticas para mejorar la solución, además de comparaciones de diferentes implementaciones de ACO aplicados a TSP como la librería TSPLIB, adicional ajustes de parámetros de manera automática, esta información nos brinda un panorama de cómo se ha realizado aplicado ACO para abordar el problema del agente viajero.

Por otro lado, para la pregunta 2 se seleccionaron 7 artículos que muestran nuevas tendencias, mejoras y como se han implementado ACO para TSP, como variantes con otras metaheurísticas o híbridos potenciando la profundización y diversificación en la búsqueda de la solución, además la utilización de computación paralela para optimizar la construcción de la solución o para la identificación de múltiples caminos utilizando CUDA, el uso de lógica difusa como método de control para ajustar parámetros, mejora en la población por medio de operadores genéticos del algoritmos genéticos.

Para la pregunta 3 se seleccionaron 7 artículos que muestran como en diferentes áreas de aplicación han aplicado el concepto de hormigas cultivadoras de hongos, como métodos de selección para materiales por medio del concepto de la mandíbula de las Attini, modelos de cortadora de hojas para mejora para la recolección de la información, además nos indica que poco se ha trabajado con este tipo de características de las hormigas cultivadoras de hongos. La tabla 7 muestra los resultados de los estudios primarios seleccionados agrupados por categorías,

Categoría	**Referencia**	**Cantidad**
Computación Paralela	(Dong, Li, Shen, et al., 2018; Hong, Huang, & Lin, 2014; Rafał Skinderowicz, 2016; Starzec, Starzec, Byrski, & Turek, 2019; Zhou, He, Hou, & Qiu, 2018)	5
Conceptos de hormigas cultivadoras de hongos	(Bakhtiari, Navid, Mehri, & Bochtis, 2011; Brito, Elzubair, Araújo, et al., 2017; Jafelice, Almeida, Meyer, & Vasconcelos, 2009; Li, Wang, Peng, & Liu, 2014; Motta Jafelice, Almeida, Meyer, & Vasconcelos, 2011; Strömbom & Dussutour, 2018; Wang & Wu, 2015)	7
Diseño de algoritmos bio inspirados	(Negulescu, Negulescu, & Dzitac, 2017)	1
Híbrido	(Al-Shihabi, 2016; Al-Shihabi, Arafeh, & Barghash, 2015; Liao & Liu, 2018; López-Santana, Rodríguez-Vásquez, & Méndez-Giraldo, 2018; Mavrovouniotis, Müller, & Yang, 2015; Mohsen, 2016; Thirugnanasambandam, Raghav, Saravanan, Prabu, & Rajeswari, 2019; Tseng, Chang, Lee, & Huang, 2019; Q. Xu, Chen, & Li, 2016; Yu, 2019)	11
Variantes	(Alaiso, Backman, & Visala, 2013; Alves, Neto, Ferreira, & Teixeira, 2018; Aziz, 2015; Bouzbita, El Afia, & Faizi, 2018; Chowdhury, Marufuzzaman, Tunc, Bian, & Bullington, 2019; Dahan, El Hindi, Mathkour, & Alsalman, 2019; Dend, Xu, & Zhao, 2019; El Afia, Bouzbita, & Faizi, 2017; Eldem & Ülker, 2017; Han, Ke, Gong, et al., 2018; Jayasankari & Uma Maheswari, 2019; Kencana, Harini, & Mayuliana, 2017; Luo, Guo, & Zhang, 2019; Ning, Zhang, Zhang, & Zhang, 2018; Oonsrikaw & Thammano, 2018; Sagban, Ku-Mahamud, & Abu Bakar, 2015; Sekara, Kowalski, Byrski, et al., 2015; Rafal Skinderowicz, 2015; Thepphakorn, Pongcharoen, & Hicks, 2014; Wu, Mallipeddi, & Suganthan, 2019; M. Xu, You, & Liu, 2017; Yan, Sohn, & Reyes, 2017; H. Zhang & You, 2019; Z. Zhang, Gao, Lu, Liu, & Liang, 2016)	24

Tabla 7 – Estudios primarios seleccionados por categoría.

Después de realizar la revisión de los estudios se consideró que para realizar un mejor análisis de resultados era necesario agruparlos en cinco categorías como se observa en la Tabla 7 y se describe a continuación el detalle de la conformación de las categorías:

Computación Paralela: Los estudios relacionados exponen soluciones de ACO aplicando conceptos de computación paralela enfocados en diferentes soluciones

novedosas como la inicialización de la matriz de la feromona, para evitar convergencia prematura en la solución, enfoques sistemáticos para la evaluación del algoritmo solo para problemas de TSP, nuevas construcciones de la solución para mejorar de esa forma la convergencia de la solución.

Conceptos de hormigas cultivadoras de hongos: Los estudios relacionados en éste apartado, no están enfocados en computación, sin embargo, son seleccionados para identificar de qué manera han abordado los conceptos de las hormigas cultivadoras de hongos, y de esta forma establecer los principios de solución para el ACO a implementar. Este tema es de vital importancia, ya que permite identificar con que características de hormigas cultivadoras de hongos se ha trabajado. En la naturaleza existen diferentes tipos de hormigas o castas para tareas específicas dentro de la colonia, ACO usa la casta de las exploradoras que se encargan de obtener el alimento.

Diseño de algoritmos bio-Inspirados: Este tema permite tener una guía de como se ha abordado el diseño de algoritmos inspirados en la naturaleza, de esta forma indicar como implementar el ACO con características de hormigas cultivadoras de hongos.

Híbridos: Los estudios pertenecientes a este tema tienen soluciones para mejorar ACO con la combinación de otros metaheurísticas o algoritmos Bio-Inspirados como SA con operadores de mutación para aumentar la diversidad de la colonia de hormigas, Algoritmos génicos para mejorar el mecanismo de actualización de la feromona, entre otras implementaciones. Es importante tenerlos en cuenta para identificar las soluciones ya que para el diseño e implementación del nuevo algoritmo no se busca realizar un híbrido.

Variantes: Los estudios expuestos muestran cómo mejorar ACO realizando variantes al ACO original, utilizan conceptos para mejorar la feromona que usa el algoritmo, además de nuevas propuestas para la construcción de la solución como multiobjetivos que permiten dividir el problema por secciones y aplicar mejoras para automatizar el refinamiento de los parámetros de ACO, conceptos que pueden aplicarse en el diseño del algoritmo a implementar.

4. Conclusiones y trabajos futuros

Las soluciones seleccionadas muestran diferentes implementaciones que contribuyen a la mejora de ACO aplicado a TSP, aplicando mejoras en el refinamiento de parámetros, construcción de la solución y depósito de la feromona, pero las soluciones identificadas no muestran conceptos de hormigas cultivadoras de hongos.

Esta RSL se realizó para tener una visión general para el diseño e implementación de una nueva variante de ACO que incluya características de hormigas cultivadoras de hongos o *Attini*, de acuerdo a la P3 (Tabla 1), los estudios que existen no indican la incorporación de características de *Attinis* en el área de computación, sin embargo los estudios seleccionados pueden brindar información de que características fueron seleccionadas y como fueron abordadas para tenerla en cuenta en el momento del diseño de ACO-Attini.

Como trabajo futuro, se pretende trabajar en i) el diseño de un nuevo ACO que incluya características de las hormigas cultivadoras de hongos o *Attini*, ii) Seleccionar las

características de las *Attini* como modelo de bioseguridad, cultivo de hongo o vuelo nupcial y iii) implementación de ACO con las nuevas características.

Referencias

Al-Shihabi, S. (2016). A hybrid of max–min ant system and linear programming for the k-covering problem. *Computers and Operations Research*, *76*, 1–11. https://doi.org/10.1016/j.cor.2016.06.006

Al-Shihabi, S., Arafeh, M., & Barghash, M. (2015). An improved hybrid algorithm for the set covering problem. *Computers and Industrial Engineering*, *85*, 328–334. https://doi.org/10.1016/j.cie.2015.04.007

Alaiso, S., Backman, J., & Visala, A. (2013). Ant colony optimization for scheduling of agricultural contracting work. In *IFAC Proceedings Volumes (IFAC-PapersOnline)* (Vol. 4, Issue PART 1). IFAC. https://doi.org/10.3182/20130828-2-SF-3019.00041

Alves, D. R. D. S., Neto, M. T. R. S., Ferreira, F. D. S., & Teixeira, O. N. (2018). SIACO: A novel algorithm based on ant colony optimization and game theory for travelling salesman problem. *ACM International Conference Proceeding Series*, 62–66. https://doi.org/10.1145/3184066.3184077

Aziz, Z. A. (2015). Ant Colony Hyper-heuristics for Travelling Salesman Problem. *Procedia Computer Science*, *76*(Iris), 534–538. https://doi.org/10.1016/j.procs.2015.12.333

Bakhtiari, A. A., Navid, H., Mehri, J., & Bochtis, D. D. (2011). Optimal route planning of agricultural field operations using ant colony optimization. *Agricultural Engineering International: CIGR Journal*, *13*(4), 1–10.

Bouzbita, S., El Afia, A., & Faizi, R. (2018). Parameter adaptation for ant colony system algorithm using hidden markov model for tsp problems. *ACM International Conference Proceeding Series*, *1*, 1–6. https://doi.org/10.1145/3230905.3230962

Brito, T. O., Elzubair, A., Araújo, L. S., ... Almeida, L. H. (2017). Characterization of the mandible atta laevigata and the bioinspiration for the development of a biomimetic surgical clamp. *Materials Research*, *20*(6), 1525–1533. https://doi.org/10.1590/1980-5373-MR-2016-1137

Chandra Mohan, B., & Baskaran, R. (2012). A survey: Ant Colony Optimization based recent research and implementation on several engineering domain. In *Expert Systems with Applications* (Vol. 39, Issue 4, pp. 4618–4627). https://doi.org/10.1016/j.eswa.2011.09.076

Chowdhury, S., Marufuzzaman, M., Tunc, H., Bian, L., & Bullington, W. (2019). A modified Ant Colony Optimization algorithm to solve a dynamic traveling salesman problem: A case study with drones for wildlife surveillance. *Journal of Computational Design and Engineering*, *6*(3), 368–386. https://doi.org/10.1016/j.jcde.2018.10.004

Cordon, O., Herrera, F., & Stiitzle, T. (2002). *A Review on the Ant Colony Optimization Metaheuristic-Basis Models and New Trends* (p. 35).

Dahan, F., El Hindi, K., Mathkour, H., & Alsalman, H. (2019). Dynamic flying ant colony optimization (DFACO) for solving the traveling salesman problem. *Sensors (Switzerland)*, *19*(8). https://doi.org/10.3390/s19081837

Dend, W., Xu, J., & Zhao, H. (2019). An improved ant colony optimization algorithm based on immunization strategy. *Advanced Materials Research*, *490–495*, 66–70. https://doi.org/10.4028/www.scientific.net/AMR.490-495.66

Dong, G., Li, W., Shen, J., … Guo, W. W. (2018). Solving traveling salesman problems with ant colony optimization algorithms in sequential and parallel computing environments: A normalized comparison. *International Journal of Machine Learning and Computing*, *8*(2), 98–103. https://doi.org/10.18178/ijmlc.2018.8.2.670

Dorigo, M., & Blum, C. (2005). Ant colony optimization theory: A survey. *Theoretical Computer Science*. https://doi.org/10.1016/j.tcs.2005.05.020

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology*, *50*(9–10), 833–859. https://doi.org/10.1016/j.infsof.2008.01.006

El Afia, A., Bouzbita, S., & Faizi, R. (2017). The effect of updating the local pheromone on ACS performance using fuzzy logic. *International Journal of Electrical and Computer Engineering*, *7*(4), 2161–2168. https://doi.org/10.11591/ijece.v7i4.pp2161-2168

Eldem, H., & Ülker, E. (2017). The application of ant colony optimization in the solution of 3D traveling salesman problem on a sphere. *Engineering Science and Technology, an International Journal*, *20*(4), 1242–1248. https://doi.org/10.1016/j.jestch.2017.08.005

Fernández, F., Valentina, C.-H., & Francisco, S. (2015). Hormigas Cortadoras de hojas en colombia: Acromyrmex & Atta. *Hormigas Cortadoras de Hojas de Colombia: Acromyrmex & Atta (Hymenoptera: Formicidae)*, 354.

Han, X. C., Ke, H. W., Gong, Y. J., … Zhang, J. (2018). Multimodal optimization of traveling salesman problem: A niching ant colony system. *GECCO 2018 Companion - Proceedings of the 2018 Genetic and Evolutionary Computation Conference Companion*, 87–88. https://doi.org/10.1145/3205651.3205731

Hong, T. P., Huang, L. I., & Lin, W. Y. (2014). A different perspective on parallel sub-ant-colonies. *12th International Conference on Advances in Mobile Computing and Multimedia, MoMM 2014*, 322–325. https://doi.org/10.1145/2684103.2684163

Jafelice, R. M., Almeida, C. G., Meyer, J. F. C. A., & Vasconcelos, H. L. (2009). Dispersal of leaf-cutting ants: Fuzzy mathematical modeling, numerical approximation and simulations. *2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference, IFSA-EUSFLAT 2009 - Proceedings*, 271–276.

Jayasankari, S., & Uma Maheswari, O. P. (2019). Multi improved ant colony optimization (IACO) based cluster head selection and information gathering in wireless sensor networks (WSNs). *International Journal of Advanced Science and Technology*, *28*(17), 326–337.

Kencana, E. N., Harini, I., & Mayuliana, K. (2017). The Performance of Ant System in Solving Multi Traveling Salesmen Problem. *Procedia Computer Science*, *124*, 46–52. https://doi.org/10.1016/j.procs.2017.12.128

Kitchenham, B., Pearl Brereton, O., Budgen, D., ... Linkman, S. (2009). Systematic literature reviews in software engineering - A systematic literature review. *Information and Software Technology*, *51*(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Leal, I. R., Wirth, R., & Tabarelli, M. (2014). The multiple impacts of leaf-cutting ants and their novel ecological role in human-modified neotropical forests. In *Biotropica* (Vol. 46, Issue 5, pp. 516–528). https://doi.org/10.1111/btp.12126

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science*, *9*, 181–211.

Li, X. B., Wang, Z. W., Peng, K., & Liu, Z. X. (2014). Ant colony ATTA clustering algorithm of rock mass structural plane in groups. *Journal of Central South University*, *21*(2), 709–714. https://doi.org/10.1007/s11771-014-1992-6

Liao, E., & Liu, C. (2018). A hierarchical algorithm based on density peaks clustering and ant colony optimization for traveling salesman problem. *IEEE Access*, *6*, 38921–38933. https://doi.org/10.1109/ACCESS.2018.2853129

López-Santana, E., Rodríguez-Vásquez, W. C., & Méndez-Giraldo, G. (2018). A hybrid expert system, clustering and ant colony optimization approach for scheduling and routing problem in courier services. *International Journal of Industrial Engineering Computations*, *9*(3), 369–396. https://doi.org/10.5267/j.ijiec.2017.8.001

Luo, Y., Guo, P., & Zhang, M. (2019). A Framework of Ant Colony P System. *IEEE Access*, *7*, 157655–157666. https://doi.org/10.1109/ACCESS.2019.2949952

Mavrovouniotis, M., Müller, F. M., & Yang, S. (2015). An ant colony optimization based memetic algorithm for the dynamic travelling salesman problem. *GECCO 2015 - Proceedings of the 2015 Genetic and Evolutionary Computation Conference*, *Cci*, 49–56. https://doi.org/10.1145/2739480.2754651

Mohsen, A. M. (2016). Annealing Ant Colony Optimization with Mutation Operator for Solving TSP. *Computational Intelligence and Neuroscience*, *2016*. https://doi.org/10.1155/2016/8932896

Motta Jafelice, R., Almeida, C., Meyer, J. F. C. A., & Vasconcelos, H. L. (2011). Fuzzy parameters in a partial differential equation model for population dispersal of leaf-cutting ants. *Nonlinear Analysis: Real World Applications*, *12*(6), 3397–3412. https://doi.org/10.1016/j.nonrwa.2011.06.003

Negulescu, A. E., Negulescu, S. C., & Dzitac, I. (2017). Balancing between exploration and exploitation in ACO. *International Journal of Computers, Communications and Control*, *12*(2), 265–275. https://doi.org/10.15837/ijccc.2017.2.2879

Ning, J., Zhang, Q., Zhang, C., & Zhang, B. (2018). A best-path-updating information-guided ant colony optimization algorithm. *Information Sciences*, *433–434*, 142–162. https://doi.org/10.1016/j.ins.2017.12.047

Oonsrikaw, Y., & Thammano, A. (2018). Modified ant colony optimization algorithm for multiple-vehicle traveling salesman problems. *International Journal of Networked and Distributed Computing*, *7*(1), 29–36. https://doi.org/10.2991/ijndc.2018.7.1.4

Rave, J. I. P., & Álvarez, G. P. J. (2013). Relevant literary space on travelling salesman problem (TSP): Contents, classification, methods and fields of inspiration. *Producao*, *23*(4), 866–876. https://doi.org/10.1590/S0103-65132013005000003

Sagban, R., Ku-Mahamud, K. R., & Abu Bakar, M. S. (2015). ACO ustic: A nature-inspired exploration indicator for ant colony optimization. *Scientific World Journal*, *2015*. https://doi.org/10.1155/2015/392345

Sekara, M., Kowalski, M., Byrski, A., … Lenaerts, T. (2015). Multi-pheromone Ant Colony Optimization for socio-cognitive simulation purposes. *Procedia Computer Science*, *51*(1), 954–963. https://doi.org/10.1016/j.procs.2015.05.234

Skinderowicz, Rafal. (2015). *Population-Based Ant Colony Optimization for Sequential*. *1*(April), 471–472. https://doi.org/10.1007/978-3-319-24306-1

Skinderowicz, Rafał. (2016). The GPU-based parallel Ant Colony System. *Journal of Parallel and Distributed Computing*, *98*, 48–60. https://doi.org/10.1016/j.jpdc.2016.04.014

Starzec, M., Starzec, G., Byrski, A., & Turek, W. (2019). Distributed ant colony optimization based on actor model. *Parallel Computing*, *90*, 102573. https://doi.org/10.1016/j.parco.2019.102573

Strömbom, D., & Dussutour, A. (2018). Self-organized traffic via priority rules in leaf-cutting ants. *PLoS Computational Biology*, *14*(10), 1–13. https://doi.org/10.1371/journal.pcbi.1006523

Thepphakorn, T., Pongcharoen, P., & Hicks, C. (2014). An ant colony based timetabling tool. *International Journal of Production Economics*, *149*, 131–144. https://doi.org/10.1016/j.ijpe.2013.04.026

Thirugnanasambandam, K., Raghav, R. S., Saravanan, D., Prabu, U., & Rajeswari, M. (2019). Experimental analysis of ant system on travelling salesman problem dataset TSPLIB. *EAI Endorsed Transactions on Pervasive Health and Technology*, *5*(19), 1–15. https://doi.org/10.4108/eai.13-7-2018.163092

Tito Chura, H. E., Silva Delgado, C. A., Alfaro Gonzales, E. E., & Fajardo Espinoza, E. (2019). Aplicación Del Algoritmo De Colonia De Hormigas Al Problema Del Agente Viajero. *Ciencia & Desarrollo*, *20*, 98–102. https://doi.org/10.33326/26176033.2015.20.521

Tseng, H. E., Chang, C. C., Lee, S. C., & Huang, Y. M. (2019). Hybrid bidirectional ant colony optimization (hybrid BACO): An algorithm for disassembly sequence planning. *Engineering Applications of Artificial Intelligence*, *83*(May), 45–56. https://doi.org/10.1016/j.engappai.2019.04.015

Wang, Y., & Wu, H. (2015). Stability of plant-pollinator-ant co-mutualism. *Applied Mathematics and Computation*, *261*, 231–241. https://doi.org/10.1016/j.amc.2015.03.061

Wu, G., Mallipeddi, R., & Suganthan, P. N. (2019). Ensemble strategies for population-based optimization algorithms – A survey. *Swarm and Evolutionary Computation*, *44*(August 2018), 695–711. https://doi.org/10.1016/j.swevo.2018.08.015

Xu, M., You, X., & Liu, S. (2017). *Dual Population Ant Colony Optimization Algorithm*. *5*.

Xu, Q., Chen, S., & Li, B. (2016). Combining the ant system algorithm and simulated annealing for 3D/2D fixed-outline floorplanning. *Applied Soft Computing Journal*, *40*, 150–160. https://doi.org/10.1016/j.asoc.2015.10.045

Yan, Y., Sohn, H. suk, & Reyes, G. (2017). A modified ant system to achieve better balance between intensification and diversification for the traveling salesman problem. *Applied Soft Computing Journal*, *60*, 256–267. https://doi.org/10.1016/j.asoc.2017.06.049

Yu, M. (2019). A solution of TSP based on the ant colony algorithm improved by particle swarm optimization. *Discrete and Continuous Dynamical Systems - Series S*, *12*(4–5), 979–987. https://doi.org/10.3934/dcdss.2019066

Zhang, H., & You, X. (2019). Multi-Population Ant Colony Optimization Algorithm Based on Congestion Factor and Co-Evolution Mechanism. *IEEE Access*, *7*, 158160–158169. https://doi.org/10.1109/ACCESS.2019.2950214

Zhang, Z., Gao, C., Lu, Y., Liu, Y., & Liang, M. (2016). Multi-objective ant colony optimization based on the physarum-inspired mathematical model for bi-objective traveling salesman problems. *PLoS ONE*, *11*(1), 1–23. https://doi.org/10.1371/journal.pone.0146709

Zhou, Y., He, F., Hou, N., & Qiu, Y. (2018). Parallel ant colony optimization on multi-core SIMD CPUs. *Future Generation Computer Systems*, *79*, 473–487. https://doi.org/10.1016/j.future.2017.09.073

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 19/09/2020
Aceitação/Acceptance: 17/11/2020

Evaluación del impacto en las guías de diseño aplicadas al contexto de ambientes virtuales de aprendizaje (AVA) para dispositivos móviles

Manuel Pastrana[1], Jhon Cano[2], Julián Mera[3]

mapastrana@admon.uniajc.edu; jhon.canob@campusucc.edu.co; julian.mera@campusucc.edu.co

[1] GrinTic– Institución Universitaria Antonio José Camacho, Avenida 6N No 28N-102, 76001, Cali, Colombia.

[2] ESLINGA– Universidad Cooperativa de Colombia, Cra. 73 #2a-80, 76001, Cali, Colombia.

[3] ESLINGA– Universidad Cooperativa de Colombia, Cra. 9 #23N-26, 190001, Popayán, Colombia.

Pages: 418-430

Resumen: Las heurísticas permiten generar una relación entre los atributos de calidad (características del comportamiento deseado del software) y los patrones de diseño (soluciones efectivas y probadas para problemas recurrentes de desarrollo de software). Esto facilita al equipo de desarrollo, la identificación de cómo construir una aplicación, logrando un mayor grado de aceptación por parte de los usuarios finales. El conjunto de heurísticas sugeridas para una solución, se le denomina guía de diseño. Las guías de diseño orientadas a la usabilidad se centran en sugerencias en torno a los aspectos pertinentes a este tópico, y son aplicadas de acuerdo a un contexto. En este trabajo, se detalla una definición de heurísticas parametrizadas en un sistema de sugerencias de guías de diseño orientadas a la usabilidad para el contexto de Ambientes Virtuales de Aprendizaje (AVA). Las sugerencias fueron implementadas, a través de la construcción de un aplicativo móvil de apoyo al aprendizaje de lecto-escritura en niños con discapacidad auditiva. Finalmente, se realiza un comparativo mediante encuestas a dos colegios, en donde se determina el impacto en la aceptación del aplicativo móvil con y sin la implementación de las guías de diseño en su proceso de desarrollo.

Palabras-clave: Usabilidad; guías de diseño; patrones de diseño; ingeniería de software.

Impact assessment on design guides applied to the context of virtual learning environments (AVA) for mobile devices

Abstract: Heuristics make it possible to generate a relationship between quality attributes (characteristics of desired software behavior) and design patterns (effective and proven solutions to recurring software development problems). This makes it easier for the development team to identify how to build an application, achieving a higher degree of acceptance by end users. The set of suggested heuristics for a solution is called the design guide. The usability-oriented design guides focus

on suggestions regarding the pertinent aspects of this topic, which are applied according to the context. In this work, a definition of parameterized heuristics in a recommender system of usability-oriented design guides for the context of Virtual Learning Environments (AVA) is detailed. The suggestions will be implemented through the construction of a mobile application to support the learning of reading and writing in children with hearing disabilities. Finally, a comparison is made through surveys of two schools, where the impact on the acceptance of the mobile application is determined with and without the implementation of the design guides in its development process.

Keywords: Usability; design guides; design patterns; software engineering.

1. Introducción

La inclusión de los dispositivos móviles a todo nivel dentro de la sociedad es cada vez más amplia. Por este motivo, para las instituciones educativas el uso de esta tecnología como parte de los procesos de enseñanza-aprendizaje, representa un gran reto. Aquí el papel de la usabilidad de los aplicativos juega un rol fundamental como estrategia de aceptación por parte de los usuarios finales según (Nielsen, 2010). De acuerdo a (Nielsen & Molich, 1990), la usabilidad es un atributo de calidad que evalúa cuán fácil son de usar las interfaces para un usuario. Sin embargo, no es una tarea fácil identificar dentro de las aplicaciones, el grado de usabilidad requerido. Se requiere de una visión arquitectural que permita identificar las mejores sugerencias, para luego ser puestas en contraste por medio de pruebas, que en su mayoría son de tipo cualitativo y no cuantitativo (Kumar & Goundar, 2019).

Una de las formas de enfrentar esta problemática, es la evaluación heurística; definida como un método de inspección que involucra a expertos y sus sugerencias para identificar problemas recurrentes y soluciones comprobadas que puedan ser aplicados a ellos de acuerdo con (Nielsen, 2010).

Para los entornos educativos, la usabilidad no se puede ver aislada por cada estudiante. Existe un cierto paralelismo e interacción en una comunidad de estudiantes, frente al aplicativo móvil, dado que combina el aprendizaje individualizado en cualquier momento y en cualquier lugar (Papadimitriou, Komis, Tselios, & Avouris, 2006) (Vavoula, Sharples, Rudman, Meek, & Lonsdale, 2009). Esto genera que las heurísticas tradicionales de desarrollo de software se vean afectadas en este contexto y los patrones de diseño que se utilizan en las implementaciones de otros contextos no necesariamente sean relevantes en el mencionado. Por lo anterior, la usabilidad, se refiere a la calidad de la experiencia de usuario al interactuar con productos o sistemas de información (D. Aniobi y E. Alu, 2016).

En complemento, el desarrollo de aplicaciones móviles para el contexto AVA (Ambientes Virtuales de Aprendizaje que puede ser entendido como un sistema integrado de lineamientos y herramientas de apoyo al aprendizaje), debe tener presente aspectos como la facilidad de aprendizaje de sus funcionalidades, la simplicidad de un diseño minimalista y la no sobre carga de contenido visual (Harrison, Flood, & Duce, 2013). En búsqueda de una nueva forma de hacer desarrollo de software orientado a la usabilidad

y para validar lo que se propone, se ha construido un aplicativo web que sirve de apoyo para generar una relación entre heurísticas, preguntas y sugerencias que ayuden a la toma de decisiones sobre cómo construir correctamente aplicativos para Ambientes Virtuales de Aprendizaje (AVA). Los resultados de las sugerencias fueron utilizados en la construcción de un aplicativo móvil para el aprendizaje de lecto-escritura en niños con discapacidad auditiva. La efectividad de la implementación de las guías de diseño orientadas a la usabilidad fue medida y comparada frente a un aplicativo móvil igual que no utilice las guías en su proceso de desarrollo.

El resto del articulo está organizado de la siguiente manera, la sección dos describe el escenario de motivación para la implementación de este modelo de desarrollo, la sección tres presenta una descripción de la estrategia propuesta, la sección cuatro presenta la discusión de los resultados, la sección cinco explica los antecedentes y trabajos relacionados, la sección seis presenta las conclusiones y el trabajo a futuro.

2. Escenario de Motivación

Este trabajo está orientado a la evaluación de las guías de diseño de usabilidad para el desarrollo de soluciones móviles en AVA. Durante la primera fase del proyecto, se construye un aplicativo que cuenta con 4 opciones principales, una denominada únete a nuestra comunidad (join us), que permite unirse al sistema y poder aportar retroalimentaciones al aplicativo, contribuir con heurísticas y participar de actividades asociadas a la comunidad del aplicativo. Una segunda opción, reflejada en la figura 1, es denominada documentación (documentation). Aquí se puede leer información de cómo obtener modelos de usabilidad o como unirse a la comunidad.

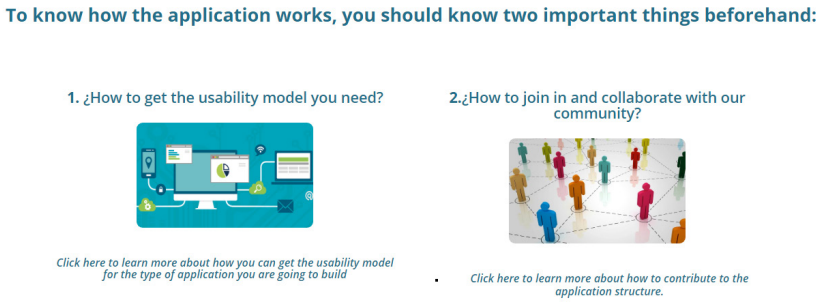


Figura 1 – Aplicativo de guías de usabilidad, sección documentación. Fuente propia

La tercera opción (get usability model) obtener modelo de usabilidad, es el eje central del aplicativo. Dependiendo del tipo de proyecto parametrizado, se pueden obtener las heurísticas seguridades de acuerdo con una serie de preguntas que enfocan al equipo de desarrollo sobre qué aspectos tiene relevancia en la construcción y así entregar las guías requeridas para alcanzar un grado de usabilidad adecuado. La última funcionalidad es un link de contacto. Adicional, el sistema tiene la opción de login permitiendo a los usuarios que están inscritos en el sistema gestionar las configuraciones mencionadas.

Gracias a esto, la fase dos se encarga de poner en funcionamiento el aplicativo, solicitando a expertos que realicen definiciones de heurística que puedan aportar sugerencias para desarrollar aplicativos en el contexto AVA, además de ser puestos a prueba para determinar la efectividad de las sugerencias. El proyecto seleccionado para trabajar con la ayuda de las guías de usabilidad se denomina Cuentil y es una aplicación que apoya el aprendizaje de lecto-escritura en niños con discapacidad auditiva.

3. Estrategia Propuesta

Para poder utilizar el sistema propuesto y medir su efectividad, es necesario realizar la parametrización del contexto AVA. La figura 2 corresponde a la pantalla de parametrización de contextos dentro del aplicativo.

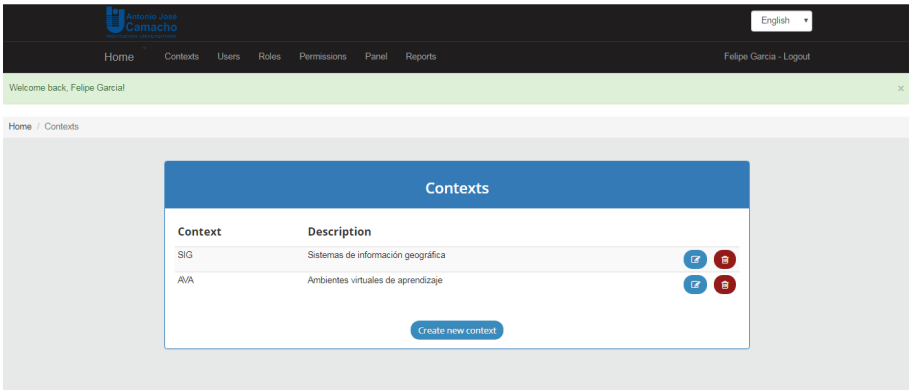


Figura 2 – Aplicativo de guías de usabilidad, parametrización de contextos. Fuente propia

Una vez realizado lo anterior, se procede a definir heurísticas asociadas al contexto. Para la definición, se utiliza como guía el trabajo de (Rodriguez González, 2018). La tabla 1 contiene las definiciones de heurísticas realizadas.

Heurística	Descripción
Adaptabilidad	Capacidad para adaptarse operativamente a diferentes ambientes sin que se necesite modificarla. Ejemplo diferentes tamaños de pantalla, diferentes navegadores.
Atractividad - Claridad arquitectónica	La interfaz de usuario debe seguir principios minimalistas y de claridad que faciliten la localización rápida de la información.
Ayudas	La información de ayuda debe ser breve, concisa, fácil de buscar y enfocada a las tareas del usuario.
Buscar la usabilidad universal	Que todas las características de la aplicación sean pensadas para que las use un principiante o un experto de la misma.
Confiabilidad	Es la medida de la habilidad de un sistema a mantenerse operativo a lo largo del tiempo.
Comprensibilidad	Capacidad del producto software de permitir al usuario comprender si el producto software es adecuado y como puede ser utilizado para tareas específicas. Ejemplo: Tooltips.

Heurística	Descripción
Conformidad	Concordancia del modelo independiente de la plataforma.
Control y libertad	Ofrece funciones de rehacer y deshacer que permitan al usuario tener el control de sus interacciones con libertad.
Consistencia y estándares	Evitar que los usuarios tengan que cuestionarse si hay frases, acciones o situaciones que tengan un mismo significado.
Contexto Cultural	Comprende el contexto de uso de una aplicación, incluyendo los usuarios y equipos que intervienen.
Efectividad	Precisión y completitud alcanzadas al realizar una tarea específica.
Eficiente	Se refiere a los recursos utilizados para alcanzar una tarea con precisión y completitud.
Estética y diseño minimalista	Hace referencia a la disposición de iconos que reflejen la realidad del mundo de los usuarios; los cuadros de dialogo no deben contener información irrelevante o que se use poco.
Facilidad de Aprendizaje	Indica la capacidad del producto software que permite al usuario aprender la funcionalidad del sistema. Ejemplo: Tooltips
Flexible	En este principio se hace referencia a la facilitación de tareas al usuarios, lo que significa que las más usadas por ellos deben estar en una pantalla de fácil acceso.
Intuitivo	El usuario puede interactuar con la aplicación de manera natural y sin necesidad de orientación previa o constante
Manejo de errores	Hace referencia a lograr recuperarse de los errores sin necesidad de salir de la aplicación y volver a entrar.
Memorabilidad	Evitar que el usuario deba recordar la información que se le da en alguna parte del proceso para poder avanzar, las instrucciones deben estar a la vista o ser de fácil acceso en el momento que se considere deben usarse
Operatividad	Capacidad del producto software que permite al usuario operar con él y configurarlo.
Privacidad	Capacidad del software de asegurar que cualquiera de sus características y/o contenidos estén enmascarados u ocultos de entidades no autorizadas.
Portabilidad	Capacidad de transferencia del producto de software de un entorno a otro. El cual puede ser de tipo organizacional, hardware o software.
Relación con la realidad	Hace referencia al entorno mediante el uso de imágenes, frases y conceptos que sean familiares para los usuarios.
Satisfacción	Sentimiento de agrado o desagrado al interactuar con una aplicación.
Simplicidad	Prescindir de elementos poco significativos en nuestra interfaz, reduciendo la carga visual
Soporte	El sistema debe incorporar mecanismos para reconocer, diagnosticar y solucionar errores.

Tabla 1 – Definición de Heurística para Ambientes Virtuales de Aprendizaje - AVA

Una vez completada la definición de heurísticas, estas son parametrizadas en el aplicativo web para poder hacer uso de ellas. La figura 3. refleja el resultado de la parametrización en el sistema para el contexto AVA.

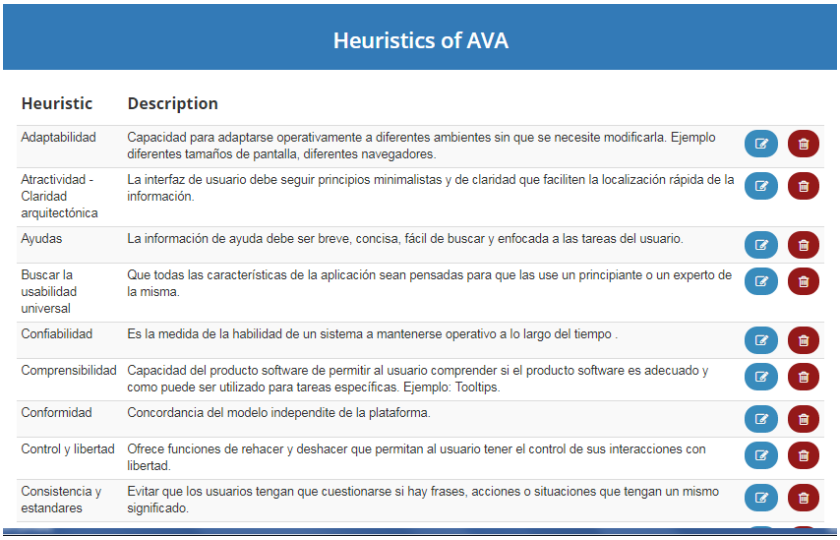


Figura 3 – Heurísticas parametrizadas en el aplicativo. Fuente propia

A continuación, por cada heurística se definen preguntas que permiten generar la orientación sugerida por la guía. Estas preguntas tienen dos respuestas posibles, sí y no. Dependiendo de la respuesta, se recomiendan unos posibles patrones de diseño orientados a la usabilidad, que pueden ser implementados para lograr satisfacer el atributo de calidad que la heurística debe cubrir, o en caso contrario se sugiere omitir esta heurística. La tabla 2 presenta la relación entre heurística, pregunta, respuesta y guía de diseño:

Heurística	Preguntas	Res.	Guía de diseño
Adaptabilidad	¿El sistema es utilizado en distintos dispositivos?	Si	Las pantallas del sistema deben ser responsivas. Utilizar Frameworks con esta capacidad incluida. Patrones diseño como tiny tweaks, mostly fluid y responsive design pueden ser de gran utilidad.
Atractividad - Claridad arquitectónica	¿La aplicación es atractiva y clara?	No	La interfaz de usuario debe seguir principios minimalistas y de claridad que faciliten la localización rápida de la información.
Ayudas	¿El sistema debe contar con ayudas en pantalla para facilitar el uso?	Si	La información de ayuda debe ser breve, concisa, fácil de buscar y enfocada a las tareas del usuario. Ayudas en pantalla como los tooltips, FAQs del sistema y acceso al manual de usuario desde un link siempre visible incrementan la usabilidad del sistema

Heurística	Preguntas	Res.	Guía de diseño
Buscar la usabilidad universal	¿El sistema está diseñado para usuarios inexpertos?	No	Se recomienda que el prototipado del sistema sea evaluado por un pequeño grupo experimental de usuarios a fin de identificar factores de interacción con dificultad y refinarlos.
Confiabilidad	¿El sistema está siendo diseñado arquitecturalmente pensando en la operatividad a través del tiempo?	No	Se recomienda el uso del patrón multicapa. Las capas mínimas recomendadas son de presentación (Pantallas), de lógica, de acceso a datos y de manejo de errores.
Comprensible	¿Las funcionalidades del sistema son comprensibles fácilmente?	No	Se recomienda el uso de tooltips y ayudas visuales.
Conformidad	¿El sistema está diseñado pensado en las especificaciones funcionales y arquitecturales?	No	Debe existir concordancia entre el modelo arquitectural y la especificación de requisitos del sistema. La inspección de código y los ciclos de calidad son una opción pertinente.
Control y libertad	¿El sistema es capaz de deshacer y/o rehacer acciones?	No	Ofrece funciones de rehacer y deshacer que permitan al usuario tener el control de sus interacciones con libertad.
Consistencia y estándares	¿Se evita que los usuarios tengan que cuestionarse si las frases, acciones o situaciones tienen el mismo significado?	No	El sistema debe manejar un lenguaje simple y común, donde todos los términos sean acordados con el cliente para no mal interpretar significados.
Contexto Cultural	¿Los usuarios finales son del mismo contexto demográfico y cultural?	Si	Aplicar Internacionalización
Efectividad	¿Se consigue el objetivo de la funcionalidad?	No	Aplicar pruebas funcionales
Eficiente	¿Las operaciones se realizan en un tiempo adecuado?	No	El uso de patrones de diseño como MultiCapa, DAO, Cache y GRASP son útiles para incrementar el rendimiento.
Estética y diseño minimalista	¿La estética del diseño es minimalista?	No	Hace referencia a la disposición de iconos que reflejen la realidad del mundo de los usuarios; los cuadros de dialogo no deben contener información irrelevante o que se use poco.
Facilidad de Aprendizaje	¿El software será usado por personas con alguna discapacidad auditiva? ¿El software será usado por personas con alguna discapacidad visual?	Si	Se deben utilizar componentes con leguajes de señas que permitan guiar al usuario final dentro del sistema. El sistema debe contener opciones de aumentar tamaño de letra o cambiar colores para los casos de daltonismo.
Flexible	¿El sistema tiene funcionalidades de uso frecuente?	Si	En este principio se hace referencia a la facilitación de tareas al usuario, lo que significa que las más usadas por ellos deben estar en una pantalla de fácil acceso, ejemplo: para acceder al botón “buscar” no debe ser necesario pasar por 2 o más pantallas.

Heurística	Preguntas	Res.	Guía de diseño
Intuitivo	¿El sistema es intuitivo?	No	Es importante validar por medio de encuestas con el grupo de usuarios finales que tan natural resulta la interacción con el sistema.
Manejo de errores	¿El sistema debe generar mensajes de error de manera entendible para el usuario?	Si	El sistema debe tener la capacidad de capturar todas las excepciones y generar en un log el mensaje respectivo para el equipo de desarrollo indicando capa, clase, método y error. En complemento debe dar un mensaje diciente al usuario que permita identificar si el error es causado por el, por ejemplo, un dato alfanumérico en un campo solo numérico.
Memorabilidad	¿La información necesaria para operar está disponible en todo momento?	No	El uso de tooltips es una excelente opción para facilitar la operación del sistema.
Operatividad	¿El sistema cuenta con la documentación respectiva para su uso?	No	El sistema debe contar con la documentación necesaria para la operación de los usuarios del sistema posterior a la capacitación. Se recomienda evaluar que tan clara es esta documentación para el usuario final.
Privacidad	¿El sistema cuenta con autenticación basada en usuarios?	No	La implementación del patrón RBA o Rol Base Access es necesaria en el sistema si se quiere mantener la privacidad de la información para que esta no sea accedida por quien no debe.
Portabilidad	¿El sistema funcionará en distintos sistemas operativos o en distintos navegadores?	Si	El uso de lenguajes multiplataforma permite escribir el código fuente de manera genérica y exportar la aplicación dependiendo de la plataforma.
Relación con la realidad	¿La aplicación toma en cuenta el contexto cultural y demográfico de los usuarios finales?	No	Ayudas como la internacionalización que permite que la misma aplicación sea utilizada en distintos idiomas genera un incremento significativo de la usabilidad pensado desde las interacciones humano computador.
Satisfacción	¿El sistema usa un sistema de medición de satisfacción del usuario?	No	Realizar retroalimentación a través de un sistema de medición, ejemplo: Encuestas.
Simplicidad	¿La interfaz gráfica o pantalla cuenta con demasiados componentes ?	Si	Prescindir de elementos poco significativos en nuestra interfaz, reduciendo la carga visual
Soporte	¿El sistema cuenta con mecanismos de soporte?	No	Es necesario crear una línea de comunicación con soporte a través de mensajes dicientes y altamente descriptivos de los errores que faciliten identificar el problema.

Tabla 2 – Relación entre heurística, preguntas, respuestas (Res.) y guías de diseño.

4. Discusión Conclusiones

Con base en la parametrización realizada al aplicativo web de guías de diseño descrita en la sección anterior, se inicia la construcción del proyecto Cuentil. Este es un aplicativo móvil tipo videojuego, que ayuda a los niños con discapacidad auditiva al aprendizaje de lectoescritura utilizando un método de narración de cuentos con base en (Nicolaidou, Tozzi, Kindynis, Panayiotou, & Antoniades, 2019). Aquí, por medio de imágenes que simbolizan parte de la historia que se está contando mediante videos con lenguajes de señas, se narran diferentes cuentos que propenden a enseñar algunas palabras. Al final, se cuenta con un taller para poder evaluar los conocimientos adquiridos a lo largo del cuento. De esta misma manera, se hará uso de elementos del trabajo de (Flórez-Aristizábal, Cano, Collazos, Solano, & Brewster, 2019) para la construcción del aplicativo.

Para el diseño de la aplicación, se utilizó el sistema de guías de diseño para el contexto **AVA.** Las principales heurísticas sugeridas por el sistema al responder las preguntas fueron las siguientes:

- Adaptabilidad: El aplicativo debe poder funcionar en diversos dispositivos móviles con características de pantalla diferente. El diseño responsive basado en patrones de diseño adaptativos es necesario.
- Atractividad - Claridad arquitectónica: Las funciones de la app deben ser pocas y fáciles de entender.
- Buscar la usabilidad universal: Se utilizó teniendo en cuenta que los usuarios de la aplicación son niños, por lo cual puede que haya algunos que sepan utilizar muy bien un dispositivo móvil, así como puede que no.
- Consistencia y estándares: Ya que la situación de discapacidad para los usuarios en este aplicativo era la problemática, se decidió no usar sinónimos para referirse siempre a una misma palabra o diferentes símbolos para una misma función.
- Estética y diseño minimalista: Se realizó la aplicación pensando en las funcionalidades mínimas para no causar una sobrecarga cognitiva al usuario.
- Ayudas: El sistema debe tener mecanismos de ayuda visuales que faciliten a los niños con discapacidad auditiva entender las historias presentadas y las preguntas que se realizan a partir de las historias.
- Efectividad: La efectividad del aplicativo debe ser medida a través de encuestas que permitan refinar aspectos que no estén dando el resultado esperado.
- Portabilidad: El sistema debe ser capaz de funcionar tanto en IOS como en Android por lo que debe ser desarrollado bajo un lenguaje multiplataforma.

4.1. Resultados

Al finalizar el desarrollo del aplicativo móvil se tuvo en cuenta que los niños con discapacidad auditiva en la gran mayoría no entienden las palabras, por lo cual se usaron componentes intuitivos como lo son las imágenes de símbolos en los botones del software, al igual que se usaron colores llamativos que captaran la atención del usuario y genere un interés de seguir aprendiendo. En la figura 4, se aprecia que, en la pantalla de inicio, se utilizaron iconos para jugar individual y en grupo, teniendo en cuenta lo que puede ser atractivo para un niño y lo que simboliza para él lo que es jugar un videojuego.

Para el tablero de selección de niveles, se utilizaron números en lenguaje de señas en sucesión de menor a mayor para indicar que al completar un nivel se avanza al siguiente. Solamente los niveles desbloqueados aparecen con los números en señas, los demás niveles hasta no ser desbloqueados se representan con un candado.

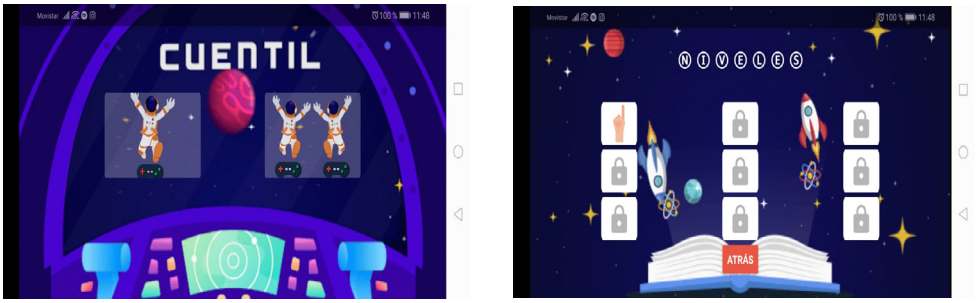


Figura 4 – Pantalla de inicio aplicación Cuentil. Fuente propia

Al seleccionar un nivel, se ingresa a la pantalla que cuenta la historia respectiva (ver Figura 5). Aquí se encuentra en el lado izquierdo el video en lenguaje de señas que narra la historia, con una imagen a la derecha que representa elementos relevantes de la historia y un texto encima para que se visualice la narración de la historia y puedan encontrar palabras que aprender y como se escriben.



Figura 5 – Pantalla de narración de cuentos. Fuente propia

Por último, se creó un taller para repasar las palabras y afianzar los conocimientos adquiridos a lo largo del cuento. Ver figura 6.

Para comprobar la efectividad del desarrollo del aplicativo móvil que implementa las guías de diseño orientadas a usabilidad, se realizó la validación frente a uno que no las utilizó en su proceso de construcción. En esta validación se cuenta con la participación de dos colegios diferentes de la ciudad de Cali, en el departamento del Valle del Cauca, Colombia. El primer colegio es el Técnico Juvenil del Sur que utiliza el aplicativo desarrollado sin las guías de diseño. El segundo colegio es la Institución Educativa Liceo Departamental sede La Gran Colombia que utiliza el aplicativo desarrollado utilizando las guías de diseño. El instrumento utilizado fueron encuestas. Para el primer colegio

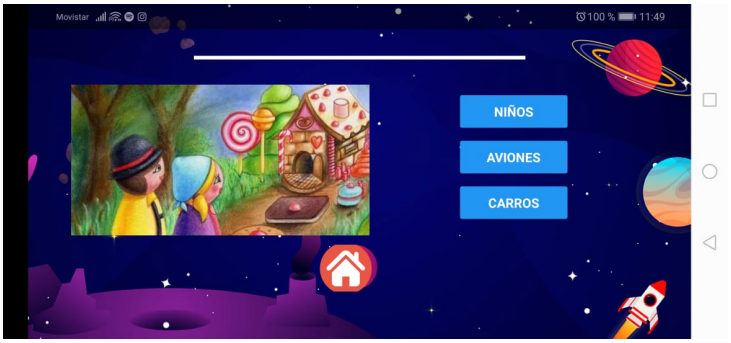


Figura 6 – Pantalla de taller de aprendizaje. Fuente propia

la distribución de edades de los encuestados fue la siguiente: 1 niños (as) de 10 años, 4 niños (as) de 9 años, 3 niños (as) de 8 años, 4 niños (as) de 7 años y 6 niños (as) de 6 años, para un total de 18 encuestados. Para el segundo colegio, la distribución fue la siguiente: 5 niños (as) de 10 años, 3 niños (as) de 9 años, 8 niños (as) de 7 años y 2 niños (as) de 6 años, para un total de 18 niños. Los resultados son expuestos en la tabla 3.

Variable a medir	**Primer Colegio**	**Segundo público**
Adaptabilidad del aplicativo	18 niños en los que el aplicativo corrió sin problemas en el dispositivo móvil	18 niños en los que el aplicativo corrió sin problemas en el dispositivo móvil
Portabilidad del aplicativo	El aplicativo corre en los diversos celulares tanto IOS como Android.	El aplicativo corre en los diversos celulares tanto IOS como Android.
Entendimiento del contexto del aplicativo	18 niños entendieron el contexto.	18 niños entendieron el contexto.
Entendimiento de los componentes en la pantalla de inicio	7 niños entendieron 11 niños no entendieron	14 niños entendieron 4 niños no entendieron
Entendimiento de los componentes en la pantalla de niveles	11 niños entendieron 7 niños no tendieron	12 niños entendieron 6 niños no entendieron
Entendimiento de los botones inferiores (Adelante, Atrás Inicio)	14 niños entendieron 4 niños no entendieron	17 niños entendieron 1 niños no entendieron
Entendimiento de las alertas en los talleres	18 niños entendieron	18 niños entendieron
Fondos (gusto)	16 niños les llama la atención 1 niño no le llama la atención 1 niño no le gusta ningún fondo del aplicativo	16 niños les llama la atención los fondos de la aplicación 2 niño no le llama la atención los fondos de la aplicación
Iconos (gusto)	16 niños les llama la atención 1 niño no le llama la atención 1 niño no le gusta ningún componente del aplicativo	18 niños les llama la atención

Tabla 3 – Encuesta realizada a colegios para evaluar el grado de aceptación de la aplicación.

Al analizar la medición del grado de aceptación, entre el aplicativo que no implementa guías (colegio 1) y el que si (colegio 2), como muestra la figura 7, se evidencia un efecto positivo para el segundo caso. En específico, los tres aspectos que mayor impacto tuvieron con el apoyo de las guías de diseño, fueron para este aplicativo: la atractividad, la estética y el uso de ayudas. La gran mayoría de los niños encuestados, lograron navegar por las interfaces del aplicativo de manera efectiva, natural e intuitiva, gracias a las ayudas visuales implementadas y a un diseño minimalista de sus pantallas. Adicionalmente, les parece atractivo el aplicativo, lo que mejora sus posibilidades de uso.

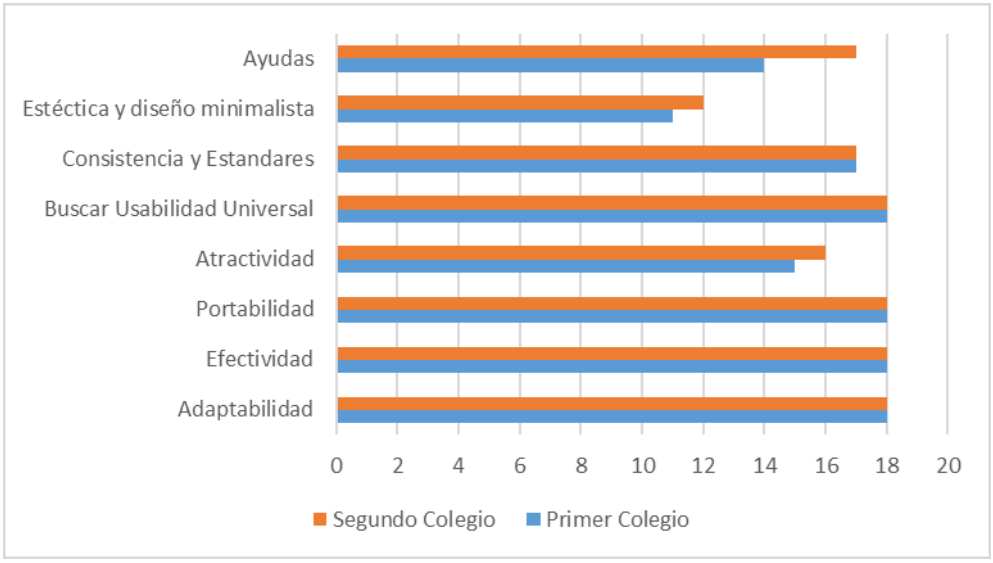


Figura 7 – Grado de aceptación de la aplicación

5. Conclusiones y resultados

Las heurísticas de usabilidad permiten generar un desarrollo guiado, organizando la información que actualmente existe, sobre problemas recurrentes (atributos de calidad) y buscando su asociación con soluciones comprobadas que ya existen para esos problemas (patrones de diseño). Al recopilar y parametrizar esta información en un sistema de recomendación, agrupándola por contextos específicos, se obtienen las guías de diseño, que pueden generar un mayor grado de aceptación por parte de los usuarios finales, de los aplicativos desarrollados bajo esta orientación.

Las evaluaciones del impacto de las guías de diseño con respecto a la aceptación del aplicativo móvil arrojaron un panorama favorable frente a las acciones realizadas por parte de los niños con respecto al aplicativo que no las utilizó. Para los niños el aplicativo que utiliza las guías resulto más intuitivo, fácil de manejar, llamativo y sobre todo útil para su proceso de aprendizaje lecto-escritor.

Es necesario a futuro probar la efectividad de las guías de diseño orientadas a la usabilidad bajo más casos dentro del contexto AVA y en otros contextos, permitiendo así refinar la información actual y alimentar el aplicativo con más información.

Referencias

D. Aniobi y E. Alu. (2016). Usability Testing and Evaluation of a Cloud Computing-based Mobile Learning App: Students' Perspective. *International Journal of Computer Science Issues*, *13*(1), 67–75. https://doi.org/10.20943/ijcsi-201602-6775

Flórez-Aristizábal, L., Cano, S., Collazos, C. A., Solano, A. F., & Brewster, S. (2019). Designability: Framework for the design of accessible interactive tools to support teaching to children with disabilities. *Conference on Human Factors in Computing Systems - Proceedings*, 1–15. https://doi.org/10.1145/3290605.3300240

Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science*. https://doi.org/10.1186/2194-0827-1-1

Kumar, B. A., & Goundar, M. S. (2019). Usability heuristics for mobile learning applications. *Education and Information Technologies*, *24*(2), 1819–1833. https://doi.org/10.1007/s10639-019-09860-z

Nicolaidou, I., Tozzi, F., Kindynis, P., Panayiotou, M., & Antoniades, A. (2019). *Development and usability of a gamified app to help children manage stress: an evaluation study*. *27*, 105–120. https://doi.org/10.17471/2499-4324/1050

Nielsen, J. (2010). What Is Usability? In *User Experience Re-Mastered*. https://doi.org/10.1016/b978-0-12-375114-0.00004-9

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/97243.97281

Papadimitriou, I., Komis, V., Tselios, N., & Avouris, N. (2006). Designing PDA mediated educational activities for a museum visit. *IADIS International Conference on Cognition and Exploratory Learning in Digital Age, CELDA 2006*.

Rodriguez González, G. C. E. (2018). Investigación e Innovación. In *Revista Logos Ciencia & Tecnología* (Vol. 10). https://doi.org/10.22335/rlct.v10i2.584

Vavoula, G., Sharples, M., Rudman, P., Meek, J., & Lonsdale, P. (2009). Myartspace: Design and evaluation of support for learning with multimedia phones between classrooms and museums. *Computers and Education*. https://doi.org/10.1016/j.compedu.2009.02.007

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: xx/08/2020
Aceitação/Acceptance: xx/10/2020

Priorização de Requisitos de Software Derivados de Modelos de Processos de Negócio: abordagem utilizando o Modelo de Kano

Rodrigo Oliveira Zacarias[1], Karina Terra de Souza[1], Thayná Albernaz Pessanha[1], Simone Vasconcelos Silva[1], Aline P. V. de Vasconcelos[1], Alline Sardinha C. Morais[1]

rodrigozacarias.ti@gmail.com, karinasouza162@gmail.com, thaynaalbernaz9@gmail.com, simonevs@iff.edu.br, apires@iff.edu.br, amorais@iff.edu.br

[1] Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Diretoria de Extensão, Pesquisa e Pós-Graduação, Programa de Pós-Graduação em Sistemas Aplicados à Engenharia e Gestão. R. Dr. Siqueira, 273, Parque Dom Bosco, 28030-130, Campos dos Goytacazes, RJ, Brasil.

Pages: 431-447

Resumo: A modelagem de processos de negócio auxilia as organizações na compreensão de suas atividades. A partir dela, é possível obter os requisitos necessários para o projeto de desenvolvimento de software capaz de automatizar as atividades. A priorização destes requisitos possui grande importância em razão de sua influência no escopo do projeto, uma vez que pode ser inviável a implementação de todos os requisitos em curto prazo. Este trabalho tem o objetivo de apresentar uma abordagem para elicitação e priorização de requisitos de software extraídos a partir da modelagem de processos por meio da técnica de priorização Modelo de Kano. A metodologia consiste em três fases: modelagem do processo, extração e priorização de requisitos. Como resultado é apresentado um estudo de caso que demonstra a aplicação da metodologia proposta.

Palavras-chave: Modelagem de Processos; Requisitos; Priorização; Modelo de Kano.

Prioritization of software requirements derived from a business process model: approach using the Kano Model

Abstract: Business process modeling helps organizations to understand their activities. Through it, it is possible to obtain the necessary requirements for the software development project capable of automating activities. These requirements prioritization is very important due to their influence on project scope, since it may be impossible to implement all requirements in a short deadline. This paper aims to present an approach for eliciting and prioritizing software requirements extracted from process modeling through the Kano Model prioritization technique. The methodology consists of three phases: process modeling, extraction and requirements prioritization. As a result, a case study is presented that demonstrates the proposed methodology application.

Keywords: Process Modeling; Requirements; Prioritization; Kano Model.

1. Introdução

Devido à alta competitividade, os sistemas de software tornam-se uma solução organizacional e face às diversas soluções tecnológicas existentes, os gestores devem ter cautela na escolha daquelas que melhor se adequem à organização. Por isso, o desenvolvimento de softwares que atendam à realidade dos processos organizacionais torna-se imprescindível (Pereira *et al.*, 2016).

No desenvolvimento de software, a gerência de requisitos é uma etapa de extrema importância que impacta na qualidade do produto final, uma vez que os requisitos devem estar em consonância com as expectativas do cliente (Cordeiro & Freitas, 2012). Para automatizar um processo de negócio, sua compreensão é essencial na identificação de possíveis melhorias que um software pode agregar.

A modelagem de processos é uma técnica que auxilia na definição e compreensão das tarefas e artefatos envolvidos e permite o monitoramento dos processos como um todo, conforme os objetivos estratégicos da organização (Pradella, 2013). Bitencourt *et al.* (2016) afirmam que a aplicação de técnicas para a extração de requisitos de software de um modelo de processo de negócio auxilia na obtenção de requisitos mais precisos, consistentes e alinhados à realidade da organização. Há diversas opções de notação para modelagem de processos de negócios disponíveis, mas o BPMN (*Business Process Model and Notation*) tem obtido destaque devido à sua aplicação em diversos projetos e contextos (Freitas, 2018).

Além da especificação, a priorização de requisitos é um fator que traz grandes impactos no escopo do projeto de software. No processo de desenvolvimento iterativo e incremental, a escolha de quais requisitos serão implementados em cada versão do software interfere na execução do projeto, reduzindo custos e tempo e, aumentando a satisfação do cliente (Cordeiro & Freitas, 2012).

Há um número considerável de métodos e técnicas de priorização de requisitos que são aplicados em diferentes projetos de software, considerando o objetivo e contexto específicos de cada escopo (Alshehri & Benedicenti, 2013; Mkpojiogu & Hashim, 2017). No entanto, grande parte desses métodos não são abordados de forma ampla e podem apresentar inconsistências relacionadas à usabilidade e escalabilidade durante a sua execução. Além disso, os mesmos não relacionam a satisfação dos usuários e a percepção de qualidade do produto final no cumprimento ou não dos requisitos (Mkpojiogu & Hashim, 2017).

Este estudo tem como objetivo apresentar uma abordagem para a elicitação e priorização de requisitos de software para automatização de processos de negócio, considerando as expectativas dos *stakeholders* em relação à otimização e às melhorias que a automatização poderá agregar ao processo. Na metodologia proposta, os requisitos são derivados de modelo de processo de negócio cuja a notação é BPMN e, posteriormente, priorizados, por meio do Modelo de Kano. De forma a complementar os resultados, é realizado um comparativo entre o Modelo de Kano e a técnica Atribuição Numérica para avaliar seus respectivos desempenhos perante requisitos extraídos de um modelo de processo de negócio.

Neste trabalho, optou-se pelo Modelo de Kano para priorização de requisitos por permitir uma avaliação sob ótica do nível de satisfação dos *stakeholders* e os impactos

na qualidade do produto final, além de apresentar um bom desempenho em relação às demais técnicas de priorização nos aspectos relatados por Mkpojiogu & Hashim (2017). E verificou-se a existência de poucos trabalhos científicos utilizando o Modelo de Kano para priorização de requisitos de software, e estes não abordaram a extração de requisitos a partir de modelos de processos de negócio.

2. Referencial teórico

2.1. Gestão de Processos

A Gestão de Processos de Negócios ou BPM (*Business Process Management*) surge como uma disciplina gerencial, que integra estratégias e objetivos de uma organização com expectativas e necessidades dos clientes. A mesma tem o objetivo de identificar, desenhar, executar, medir, monitorar e controlar processos de negócios, que podem ser automatizados ou não, a fim de alcançar resultados consistentes e alinhados aos objetivos estratégicos da organização. Através do BPM, a organização consegue promover a melhoria do desempenho organizacional, agregando valor e inovação (ABPMP, 2013; Pradella, 2013).

As técnicas de modelagem de processo de negócio proporcionam a compreensão do ambiente no qual o sistema irá funcionar, identificando os requisitos que correspondem às necessidades do negócio. A representação do processo auxilia no monitoramento e execução da estratégia, e o desenho baseado em modelo faz com que os stakeholders melhorem o processo de comunicação. A notação BPMN tem por objetivo padronizar graficamente a modelagem dos processos (ABPMP, 2013; Weilkiens *et al.*, 2017). A Figura 1 mostra os principais símbolos da notação BPMN.

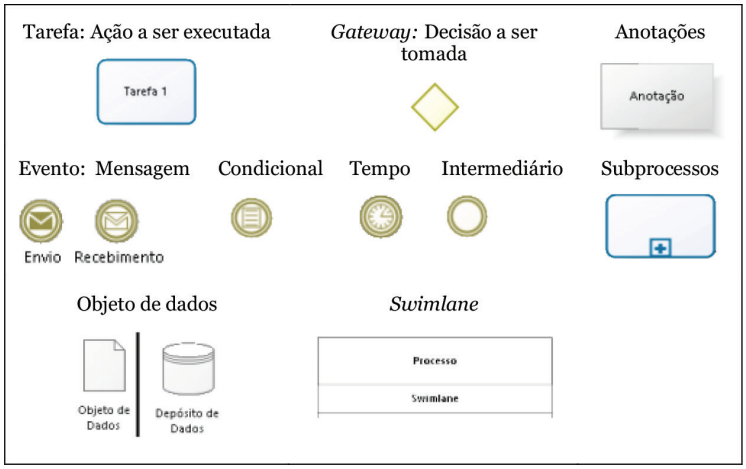


Figura 1 – Principais símbolos da notação BPMN[1].

[1] Acesse o catálogo completo de elementos de BPMN em https://www.omg.org/bpmn/Samples/Elements/Complete_BPMN_Elements.htm

2.2. Gerência de Requisitos

Não há definição precisa de requisitos de software, pois diferentes autores os veem em diferentes contextos (Olaronke *et al*., 2018). De acordo com IEEE (1990), requisito é uma condição que deve ser atendida por um sistema para atender a um contrato, padrão, especificação ou outro documento formal.

Segundo Sommerville (2011), os requisitos são classificados em funcionais (descrevem as funcionalidades que o sistema deve executar) e não funcionais (são às especificações técnicas para qualificar os requisitos funcionais). Os requisitos são especificados em consonância com as regras do negócio para que o produto final possua a aplicabilidade desejada. As regras de negócio são declarações que descrevem e/ou restringem o negócio (Melo & Muniz Silva, 2014).

Para Machado (2016) a gerência de requisitos é o processo responsável pelos atributos e definições dos requisitos de software gerando a priorização e rastreabilidade. No desenvolvimento de software a elicitação e priorização de requisitos são etapas fundamentais, e Engenharia de Requisitos diz que nelas ocorre o mapeamento dos requisitos utilizados.

Elicitação de Requisitos

Diversas abordagens são utilizadas para a elicitação de requisitos, tais como: entrevistas, observação, prototipação, *brainstorming*, análise de textos, entre outras. Os requisitos elicitados devem possuir completude, consistência, legibilidade, além de não conter ambiguidades de forma que asseguram que um projeto de software atinja seus objetivos (Pressman, 2011). Alguns requisitos elicitados pelos métodos tradicionais não refletem a dinâmica do negócio, assim, os modelos de processos de negócio são uma alternativa para a obtenção de requisitos mais precisos e adequados à organização (Bitencourt *et al*., 2016).

Existem na literatura técnicas e métodos baseados em *frameworks* e heurísticas que auxiliam na obtenção dos requisitos por meio dos modelos, tais como:

- BPMNFR: permite a extração de requisitos não funcionais a partir de extensão da notação BPMN, identificando os requisitos nos componentes dos modelos de processos (Xavier, 2009);
- *Workshops:* propõe o levantamento de requisitos a partir da análise dos modelos de processos de negócio junto aos *stakeholders* (Farias, 2010);
- REMO: apresenta heurísticas (H) que permitem a identificação e extração de requisitos funcionais (RF) e não funcionais (RNF) e regras de negócios (RN) a partir dos modelos de processos de processos (Vieira, 2012);
- REMO Adaptado: propõe o aprimoramento da técnica REMO por meio de maior detalhamento dos procedimentos de obtenção de requisitos não funcionais, e acrescenta uma heurística à técnica permitindo a análise de componentes do BPMN que se referem a subprocessos (Freitas, 2018).

Dentre os métodos apresentados, a técnica REMO destaca-se por apresentar resultados significativos quando comparada a técnicas de elicitação de requisitos tradicionais e por extrair, além dos requisitos funcionais, os não funcionais e as regras de negócio a partir

dos modelos de processos de negócio. Baseado nesse resultado, Freitas (2018) propõe melhorias para técnica REMO a fim de obter maior precisão e completude dos requisitos extraídos dos modelos de processos de negócio elaborados em BMPN. Nesta técnica REMO Adaptado, as adaptações nas heurísticas incidem em melhorias na obtenção dos requisitos não funcionais e na análise de subprocessos no modelo, conforme segue:

- H1 - Atividades/Tarefas do Processo (podem ser automatizadas através de funções que o sistema possuirá ou continuarão sendo realizadas de forma manual). RF – Análise da necessidade de informatizar a atividade. RNF – Avaliação se a há restrições na atividade a serem realizadas, utilizando de critérios (Segurança, Tempo de Resposta, Confiabilidade, Armazenamento de grande volume de dados, Eficiência e Escalabilidade);
- H2 - Condições de Decisão (identificam requisitos funcionais implícitos ou regras a serem atendidas). RF - Análise da necessidade de descrever um ou mais RF. RN – Verificação se há regras a serem atendidas pelo sistema;
- H3 - Eventos de Mensagens (o evento de lançamento e/ou captura da mensagem). RF - Análise se a descrição do envio da mensagem é um RF. RNF – Avaliação da exigência do tempo de resposta para cada mensagem;
- H4 - Eventos Condicionais (identificam uma pré-condição ou regra que deve ser atendida pelo sistema). RF - Análise se uma ação ocorre a partir de evento. RN - Verificação se o evento é uma regra a ser atendida;
- H5 - Eventos de Tempo (identificam a periodicidade a ser atendida/esperada durante o fluxo do processo). RF - Avaliação se a partir do evento descreve-se uma ação para o sistema. RN - Verificação se há descrito um tempo que pode ser uma regra a ser atendida pelo sistema;
- H6 - Eventos Intermediários (identificar uma ação/condição para o sistema). RF - Análise se uma ação é descrita para o sistema a partir do evento. RN - Verificação se o evento torna-se uma regra a ser atendida;
- H7 - Objeto de Dados (identificam um documento utilizado/produzido por uma determinada atividade). RF - Análise se o objeto deve ser mantido e se indica a necessidade de emissão de relatórios. RN - Verificação se o objeto possui informações obrigatórias a serem atendidas pelo sistema;
- H8 – Anotações (identificam informação a ser atendida). RF - Análise se a anotação permite criar uma ação para o sistema. RN - Verificação se as informações tornam-se uma regra que a ser atendida. RNF - Avaliação se as anotações descrevem qualidade específica a ser atendida;
- H9 - Envolvidos no Processo (identificam papeis utilizados pelo sistema). RF - Análise se cada *swimlane* deve ser considerada como papel a ser mantido pelo sistema. RN - Verificação se para cada *swimlane* é necessário realizar um controle de perfil. RNF – Avaliação da necessidade de descrever uma qualidade relacionada a controle de segurança;
- H10 – Subprocessos (podem gerar requisitos). RF / RN / RNF: Análise se as atividades do subprocesso devem ser traduzidas como requisitos do sistema, de acordo com as outras heurísticas. Verificação se o referido subprocesso realiza tarefas ligadas ao processo e por isso devem ser automatizadas. Avaliação, em caso contrário, deve ser feito um controle do fluxo do processo quando o subprocesso está sendo executado.

Neste trabalho, se optou por aplicar a técnica REMO Adaptado, uma vez que a mesma agrega maior acurácia a já eficaz técnica REMO, além de utilizar a notação BPMN como padrão de modelagem de processos de negócio.

Priorização de Requisitos

Muitos projetos de software adotam processos de desenvolvimento iterativo e incremental, entregando o software em *releases*. Diversos atributos são relacionados ao projeto, dificultando a seleção dos requisitos implementados em cada ciclo incremental. Para Gupta *et al.* (2017) é a equipe de desenvolvimento que decide os requisitos implementados numa versão, considerando as necessidades do cliente, tempo, custos, riscos e recursos disponíveis. Pode-se citar algumas técnicas de priorização de requisitos (Olaronke *et al.*, 2018; Asghar *et al.*, 2016):

- Atribuição numérica: escala que varia de 1 a 5, os *stakeholders* atribuem um valor a cada requisito, sendo quanto maior o valor maior a importância;
- MoScoW: os *stakeholders* agrupam os requisitos em quatro grupos de prioridades: *MUST have*, *SHOULD have*, *COULD have* e *WON'T have*, sendo esta, respectivamente, a escala de prioridade;
- Top 10: os *stakeholders* escolhem os dez principais requisitos da lista. Adequada para um pequeno número de requisitos e *stakeholders*;
- AHP (*Analytic Hierarchy Process*): método multicritério baseado na comparação paritária no cálculo do valor de importância de cada item;
- Modelo de Kano: identifica e prioriza os requisitos conforme a percepção de qualidade e satisfação dos usuários através de questionário específico.

Segundo Mkpojiogu & Hashim (2017), não se pode afirmar que uma técnica de priorização é superior às demais, pois a maior parte delas ainda não foi amplamente adotada e possui problemas de usabilidade e escalabilidade, além de não considerar a importância dos atributos de qualidade e a satisfação dos clientes no processo de priorização de requisitos. O modelo de Kano é proposto como uma alternativa às limitações mencionadas, uma vez que vincula a satisfação dos clientes ao atendimento dos requisitos que são prioridade para os usuários.

2.3. Modelo de Kano

O modelo de Kano propõe que a satisfação do cliente aumenta de forma significativa com pequenas melhorias no desempenho de um requisito, enquanto que para outros, por mais que o desempenho seja ótimo, a satisfação não sofre significativas alterações (Sauerwein *et al.*, 1996). Este modelo identifica e qualifica os requisitos de um produto ou serviço para torná-los mais atrativos aos clientes. Os requisitos são extraídos através de pesquisas de campo por meio de entrevistas a fim de compreender as necessidades dos clientes (Pinto *et al.*, 2012).

A identificação dos tipos de requisitos é realizada através de um questionário em que é apresentado um par de perguntas (Como você se sente quando o atributo x está presente no produto/serviço? E quando ele não está?) para cada característica do produto/serviço, tendo o cliente a possibilidade de responder de cinco formas diferentes, conforme o grau de satisfação (1. Eu gosto disso desta maneira, 2. Isso deve ser desta maneira, 3. Eu fico

neutro, 4. Eu posso viver com isso desta maneira, 5. Eu não gosto disso desta maneira). Investiga-se a reação do cliente quando o produto/serviço possui uma determinada característica (forma funcional), em seguida, é avaliada a reação oposta, isto é, quando a característica não é encontrada no produto/serviço (forma disfuncional) (Sauerwein *et al.*, 1996).

A partir das respostas obtidas no par de perguntas é realizada uma análise de classificação dos requisitos do modelo em seis categorias de acordo com a importância de cada requisito para determinado produto ou serviço *(*Berger *et al.*, 1993; Matzler & Hinterhuber, 1998): (A)trativo - exerce maior influência no grau de satisfação dos clientes; (O)brigatório - quando não atendido, gera significativa insatisfação por parte dos clientes; (U)nidimensional - o nível de satisfação dos clientes é diretamente proporcional ao grau de desempenho do produto/serviço; (N)eutro - sua presença ou ausência não resulta em qualquer satisfação ou insatisfação por parte do cliente; (R) everso - o baixo desempenho do atributo resulta em satisfação, e vice-versa; (Q) uestionável - indica que a pergunta não foi devidamente formulada, ou que o cliente não a compreendeu da forma adequada, ou que a resposta foi inconsistente.

A classificação dos requisitos é realizada através da combinação das respostas para as formas funcional e disfuncional da pergunta. Esta classificação deve ser realizada para cada cliente entrevistado (Roos *et al.*, 2009). Uma vez classificados os requisitos por todos os clientes, é gerada uma tabela de resultados que demonstra a distribuição geral do requisito analisado sobre todas as categorias, uma vez que a opinião dos participantes não é absoluta para uma classificação específica. A definição da categoria pode ser realizada de acordo com a maior frequência de respostas dadas pelos participantes para um determinado requisito. No entanto, quando esta identificação não é clara assume-se O>U>A>N para a determinação da categoria do requisito (Sauerwein *et al.*, 1996).

A satisfação dos clientes em relação a presença ou não de determinado requisito pode ser avaliada através de dois índices: coeficiente de satisfação (Equação 1) e coeficiente de insatisfação (Equação 2) (Berger *et al.*, 1993).

$$CS = \frac{\%A+\%U}{\%A+\%U+\%O+\%N} \qquad \text{Equação 1}$$

$$\mathrm{CI} = \frac{(\%\mathrm{U}+\%\mathrm{O})-1}{\%\mathrm{A}+\%\mathrm{U}+\%\mathrm{O}+\%\mathrm{N}} \qquad \text{Equação 2}$$

O coeficiente de satisfação varia de 0 a 1, sendo que quanto mais próximo de 1, maior será a influência do requisito na satisfação do cliente. Se o coeficiente de insatisfação atingir valores próximos a -1, então o grau de insatisfação do cliente será alto se o requisito não for alcançado; se o coeficiente assumir valores próximos a 0, o requisito não interfere na satisfação do cliente (Sauerwein *et al.*, 1996).

A utilização do modelo de Kano permite que requisitos do produto sejam melhor compreendidos, pois auxilia na identificação dos critérios que trazem maior satisfação para os clientes. A classificação dos requisitos em obrigatório, unidimensional ou atrativo pode ser usado no processo de desenvolvimento de produtos através da priorização de

investimentos em requisitos atrativos e unidimensionais, pois estes trazem maior grau de satisfação para os clientes. Esta situação é observada quando existem requisitos que não podem ser atendidos de forma simultânea, assim, torna-se necessário optar por aquele que exerce maior influência na satisfação do cliente (Matzler & Hinterhuber, 1998).

Como exemplo de utilização do modelo de Kano pode-se citar o trabalho de Duda *et al.* (2015) onde o modelo priorizou requisitos do projeto relacionado à entrega de conteúdo móvel a várias plataformas. E como resultado, a técnica agrupou os requisitos de software em três categorias (obrigatório, linear e entrega inicial).

3. Metodologia

A abordagem para extração e priorização de requisitos de software proposta neste trabalho é representada pela Figura 2, sendo composta pelas três fases.

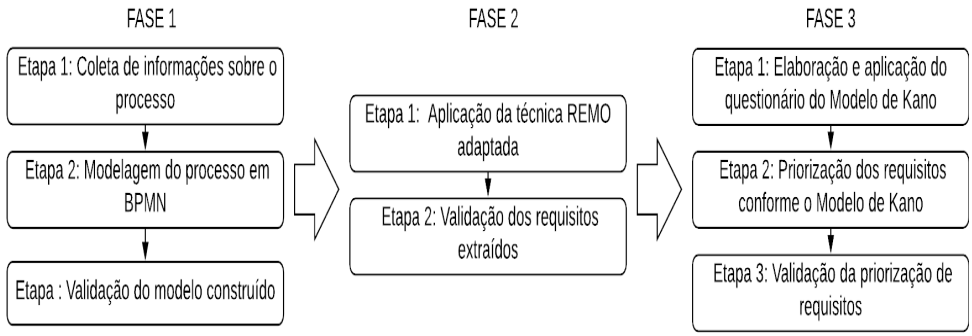


Figura 2 – Síntese da Metodologia Proposta.

A Fase 1 - Modelagem do Processo de Negócio é composta pelas etapas:

- Etapa 1 - Coleta de informações sobre o processo*:* é realizada por meio de entrevistas individuais junto aos *stakeholders*, possibilitando diversas visões das atividades e auxiliando na descrição mais fiel à realidade;
- Etapa 2 - Modelagem do processo em BPMN: é realizada com base nas informações coletadas junto aos *stakeholders* permitindo que seja desenvolvido o desenho do processo e o detalhamento de suas tarefas;
- Etapa 3 - Validação do modelo construído: por meio de reunião de *brainstorming* com os *stakeholders* e mediada pelo analista de sistemas. É realizada a leitura do modelo e o mesmo é avaliado em relação a realidade do negócio. Caso haja informações conflitantes, o modelo deve ser retificado conforme as considerações dos *stakeholders*.

A Fase 2 – Elicitação de Requisitos é composta pelas etapas:

- Etapa 1: Aplicação da técnica REMO Adaptado: propõe as heurísticas para obtenção dos requisitos de software. Esta técnica foi escolhida por possuir um melhor detalhamento dos procedimentos de extração de requisitos não-funcionais e adicionar uma heurística para componentes BPMN que se refere a subprocessos. As informações extraídas são organizadas em requisitos (funcionais e não funcionais) e regras de negócio;
- Etapa 2: Validação dos requisitos extraídos: ocorre ao término da extração dos requisitos do processo e são validados pelos *stakeholders* através de entrevistas individuais, no intuito de captar a visão de cada parte envolvida sobre o processo modelado. É verificada a qualidade dos requisitos extraídos, identificando possíveis ambiguidades, inconsistências, não completude e demais problemas de especificação. As divergências são retificadas e novas informações devem ser incluídas.

A Fase 3 – Priorização de Requisitos é composta pelas etapas:

- Etapa 1: Elaboração e aplicação do questionário do Modelo de Kano: elaboração do questionário de acordo com modelo e o preenchimento do mesmo pelos *stakeholders* de forma individual;
- Etapa 2: Priorização dos requisitos conforme Modelo de Kano: a priorização dos requisitos funcionais, assim como das regras de negócio, é realizada a partir das respostas coletadas no questionário de Kano;
- Etapa 3: Validação da priorização de requisitos: os requisitos priorizados são validados por meio de entrevistas individuais com os *stakeholders* a fim de verificar sua consistência. É importante que estes tenham a ciência de que a entrega do software será realizada conforme as prioridades elencadas. Caso haja divergências, estas são resolvidas por meio de consenso entre os envolvidos, podendo ser aplicada a técnica de *brainstorming*. Concluída a validação, o documento de requisitos é consolidado pelo analista de sistemas e dado prosseguimento ao projeto de desenvolvimento de software.

4. Estudo de Caso

Com o intuito de aplicar a metodologia proposta, foi realizado um estudo de caso em uma instituição pública estadual, localizada no estado do RJ, a qual é responsável pela gestão das atividades administrativas e pedagógicas das unidades escolares estaduais. Dentre os processos institucionais, optou-se pela análise e modelagem de um processo referente a uma das atividades interna da instituição: atendimento de requisições de material de estoque. Esse processo foi escolhido pois o mesmo é executado diariamente pelo setor responsável e apresenta dificuldades e inconsistências que poderiam ser minimizadas com a automatização.

Iniciando a Fase 1 da metodologia, foi realizada a coleta de informações por meio de entrevistas não estruturas e individuais com uma amostra de sete *stakeholders*, sendo seis membros de equipe e um coordenador do setor. Com as informações coletadas, foi

elaborado no software *Bizagi Modeler*[2] o modelo do processo de negócio. Em seguida, o mesmo foi avaliado pelos *stakeholders* e pequenos ajustes foram realizados (Figura 3). A visualização ampliada da Figura 3 encontra-se disponível para acesso (https://cutt.ly/FhTr7fA).

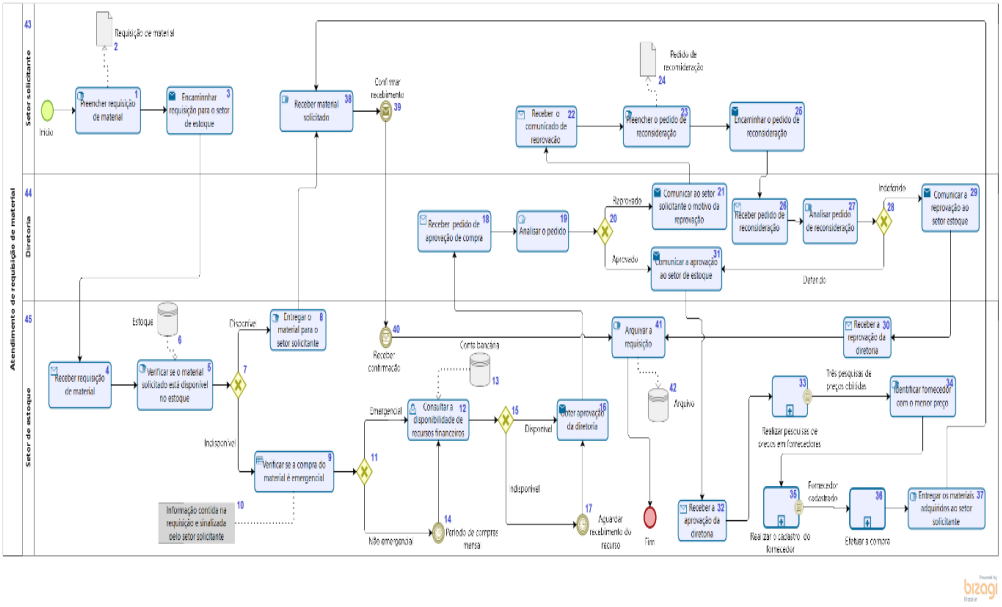


Figura 3 – Modelagem do Processo: Atendimento de requisição de material.

Com o modelo validado, foi iniciada a Fase 2 da metodologia. Aplicou-se a técnica de REMO Adaptado para a extração dos requisitos de software e regras de negócio relacionados ao processo. A partir da análise dos componentes BPMN norteada pelas heurísticas da técnica, foram identificados nove requisitos funcionais, três requisitos não funcionais, quatros regras de negócio e três atores do sistema (funcionário do setor solicitante, funcionário do setor de estoque e diretor).

Após essa etapa, a lista de requisitos e regras de negócio elaborada foi submetida à avaliação dos *stakeholders*. Os mesmos solicitaram explicações sobre a forma de especificação dos requisitos para compreender como cada um atenderia suas necessidades. Alguns deles sugeriram a inclusão de um novo requisito que não foi identificado no modelo pela técnica REMO Adaptado, o RF9 (Emitir relatório de rastreabilidade de consumo). Segundo eles, a obtenção de um histórico dos tipos de materiais consumidos por cada setor e sua frequência poderiam auxiliá-los em decisões estratégicas relacionadas às compras.

Dessa forma, a Tabela 1 apresenta os dados extraídos do modelo do processo "Atendimento de requisição de material". Na primeira coluna estão listados os requisitos funcionais

[2] Disponível em www.bizagi.com

(RF1 a RF8) e solicitado pelos *stakeholders* (RF9), requisitos não funcionais (RNF1 a RNF3) e regras de negócio (RN1 a RN4). A segunda coluna apresenta quais heurísticas foram utilizadas para a obtenção desses dados. Por fim, a terceira coluna apresenta os números que indicam os componentes BPMN (Figura 3) que originaram esses dados.

Prosseguindo para a Fase 3 da metodologia, foi elaborado o questionário do Modelo de Kano para priorização dos requisitos funcionais do software (Tabela 1). Foram elaboradas perguntas funcionais e disfuncionais para avaliar cada um dos requisitos. Associado a cada requisito foi efetuada a descrição de como cada um será representado no software para proporcionar maior entendimento aos membros do setor de estoque. Além disso, também foi solicitado aos *stakeholders* que avaliassem cada requisito segundo um grau de importância, dentro de uma escala variando de 1 (pouco importante) à 5 (muito importante).

O questionário foi preenchido pelos *stakeholders* com a supervisão do analista de sistemas para o esclarecimento de dúvidas. Os dados coletados foram tabulados e aplicou-se o modelo de Kano para a classificação de prioridade dos requisitos. Paralelamente, houve a aplicação do método de Atribuição Numérica com os valores de importância coletados para fins de comparação com o Modelo de Kano, permitindo uma avaliação complementar dos resultados.

Por fim, os requisitos priorizados foram submetidos à avaliação dos membros do setor de estoque. Cada membro avaliou o documento apresentado, com oportunidade de efetuar quaisquer considerações que julgasse necessárias. Em suma, os *stakeholders* consideraram que o levantamento de requisitos foi bem abrangente e conseguiu captar as reais necessidades de melhorias para o processo com a sua automatização, o que fez com que os requisitos apresentassem alto grau de importância e, consequentemente, o alto grau de prioridade no momento da avaliação. Nesse sentido, os resultados se mostraram coerentes com a realidade do processo. Ao final desta etapa, todos os membros concordaram com o documento, permitindo que o analista consolidasse a documentação da gerência dos requisitos.

Requisitos Funcionais (RF)	**Heurísticas**	**Elemento BPMN**
RF1 – Manter perfis de usuário de acordo com os atores	H9	43, 44 e 45
RF2 – Manter material de estoque	H1/H7	5 e 6
RF3 – Manter a requisição de material e fluxo entre setores	H1/H7	1, 2, 3, 4, 41 e 42
RF4 – Listar os materiais disponíveis no estoque	H1/H2	5, 6 e 7
RF5 – Lançar os materiais entregues	H1/H3	8 e 38
RF6 – Registrar a entrada e a saída de material do estoque	H1/H10	8, 36, 37, 38, 39 e 40
RF7 – Manter pedido de reconsideração e fluxo entre setores	H1/H7	23, 24, 25, 26, 27, 28, 29, 30 31 e 32
RF8 – Manter o status da requisição de material conforme as atividades do processo.	H1/H3/H10	1, 5, 8, 19, 33, 34, 35, 36, 37, 41 e 42
RF9 – Emitir relatório de rastreabilidade de consumo	---	Requisito não extraído do modelo

Requisitos Não Funcionais (RNF)	**Heurísticas**	**Elemento BPMN**
RNF1 – Controle da segurança dos perfis de usuário	H9	43, 44 e 45
RNF2 – Capacidade de armazenamento das requisições de material	H1/H7	1 e 2
RNF3 – O sistema deve funcionar em plataforma web	H1	Todas as tarefas
Regras de Negócio (RN)	**Heurísticas**	**Elemento BPMN**
RN1 – Classificar compras: emergenciais e não emergenciais	H1/H2/H8	9 e 11
RN2 – Efetuar compras somente com a disponibilidade de recursos financeiros	H1/H2/H5	12, 15 e 17
RN3 – As compras não emergenciais são realizadas somente no período de compras mensal	H2/H5	11 e 14
RN4 – As compras são efetuadas com aprovação da diretoria	H2	20 e 28

Tabela 1 – Dados extraídos do modelo "Atendimento de requisição de material"

4.1. Resultados

A partir dos dados coletados foi possível estabelecer a priorização dos requisitos funcionais conforme a visão dos *stakeholders* sobre as melhorias no processo (Tabela 2). De acordo com a aplicação do Modelo de Kano observa-se que todos os requisitos possuíram a mesma classificação final (U), isto é, a satisfação dos clientes tende a aumentar quando os requisitos são implementados, e vice-versa. Devido a este fato, conclui-se que a priorização de requisitos por meio do Modelo de Kano foi de significativa importância para o desenvolvimento deste estudo.

O grau de importância (Tabela 2) foi obtido através da técnica Atribuição Numérica, onde os *stakeholdes definiram* o nível de importância (1 -pouco e 5 -muito) de cada requisito no desenvolvimento do software. O grau de importância final foi calculado por meio da média simples das avaliações.

Cliente	**Requisitos**																	
	RF1		**RF2**		**RF3**		**RF4**		**RF5**		**RF6**		**RF7**		**RF8**		**RF9**	
1	N	4	N	4	N	4	U	5	U	4	U	5	U	3	O	4	A	4
2	Q	4	U	5	U	5	U	5	U	5	U	5	U	5	U	5	U	5
3	O	5	N	4	A	4	A	4	A	4	A	4	A	4	A	3	A	3
4	U	5	U	5	U	5	A	5	U	5	A	5	U	5	O	5	N	3
5	U	5	U	5	U	5	Q	5	U	5	Q	5	U	5	U	5	U	5
6	N	4	N	5	U	5	U	5	U	5	U	5	U	5	U	5	U	5
7	U	3	U	3	O	4	U	4	U	3	U	3	O	4	U	3	O	5

Tabela 2 – Classificação e Grau de importância dos requisitos

A Tabela 3 mostra os resultados obtidos com as técnicas de priorização, sendo o grau de importância (GI) utilizado em ambas as técnicas, e os coeficientes de satisfação (CS), insatisfação (CI) e a classificação (CL) dos requisitos restritos ao Modelo de Kano. Todos os requisitos apresentaram elevado grau de importância, bem como alto coeficiente de satisfação. Observa-se que os requisitos RF4, RF5 e RF6 proporcionam máxima satisfação quando atendidos, e geram significativa insatisfação quando não o são. A partir dos dados de CI, CS e GI, utilizou-se a técnica de regressão linear para estabelecer uma relação entre essas variáveis, conforme Equações 3 e 4, e assim foram calculadas as estimativas ECS e ECI.

$$ECS = 0.808\ GI - 2.81 \quad \text{Equação 3}$$

$$ECI = -0.304\ GI + 1.01 \quad \text{Equação 4}$$

O nível de prioridade foi definido por meio de quartis, que tiveram seus limiares em Q1(25%), Q2(50%) e Q3(75%) do número de requisitos (Equações 5, 6 e 7). Termo significa a posição do requisito no ranqueamento.

$$Q_1 = \frac{(n+1)th}{4}\ termo \quad \text{Equação 5}$$

$$Q_2 = \frac{(n+1)th}{2}\ termo \quad \text{Equação 6}$$

$$Q_3 = \frac{3(n+1)th}{4}\ termo \quad \text{Equação 7}$$

A partir da definição dos quartis foram estabelecidos quatro níveis de prioridade, a partir do ranqueamento (R) de CS (PCS), CI (PCI) e GI (PGI), a saber: $R < Q_1$ (prioridade muito alta), $Q_1 \leq R < Q_2$ (prioridade alta), $Q_2 \leq R < Q_3$ (prioridade média), e $R \geq Q_3$ (prioridade baixa). Estes limiares foram comparados com o ranqueamento dos coeficientes e do grau de importância em ordem decrescente. Ressalta-se que, em relação ao CI, valores mais próximos de -1 terão prioridade mais elevada, uma vez que o grau de insatisfação será maior à medida que o coeficiente se aproxima deste valor. Considerou-se a mesma posição no *ranking* quando os coeficientes apresentaram valores iguais.

Req	GI	CS	CI	CL	ECS	ECI	PCS	PCI	PGI
RF1	4.29	0.5	-0.5	U	0.66	-0.29	alta	alta	alta
RF2	4.43	0.57	-0.43	U	0.77	-0.34	alta	alta	alta
RF3	4.57	0.71	-0.29	U	0.88	-0.38	muito alta	muito alta	muito alta
RF4	4.71	1	-0.5	U	1	-0.42	muito alta	muito alta	muito alta
RF5	4.43	1	-0.14	U	0.77	-0.34	alta	alta	alta

Req	GI	CS	CI	CL	ECS	ECI	PCS	PCI	PGI
RF6	4.57	1	-0.5	U	0.88	-0.38	muito alta	muito alta	muito alta
RF7	4.43	0.86	-0.14	U	0.77	-0.34	alta	alta	alta
RF8	4.29	0.71	-0.14	U	0.66	-0.29	alta	alta	alta
RF9	4.29	0.71	-0.43	U	0.66	-0.29	alta	alta	alta

Tabela 3 – Grau de importância, coeficientes e priorização dos requisitos funcionais

4.2. Discussões dos Resultados

A partir do modelo de processo de negócio foram extraídos nove requisitos funcionais, três não funcionais e identificadas quatro regras de negócio. De acordo com Sommerville (2011), os requisitos não funcionais estão relacionados às especificações técnicas e qualitativas do software e por isso não foram priorizados. Em relação aos requisitos funcionais, foi observado que a satisfação dos clientes será proporcional à implementação dos mesmos. Alguns requisitos (RF3, RF4 e RF6) demonstraram uma maior importância através da técnica Atribuição Numérica. No entanto, quando utilizado o Modelo de Kano, os requisitos RF4, RF5 e RF6 passam a assumir maior destaque, devido seus coeficientes CS e CI.

A partir da análise do CS e CI, percebe-se que o requisito RF5 apresenta um alto nível de satisfação quando é atendido, mas um baixo grau de insatisfação quando não é, demonstrando que este não seria um requisito chave para o projeto. Assim, afirma-se que os requisitos RF4 e RF6 deveriam ser priorizados em primeiro momento.

As divergências entre os requisitos a serem priorizados, por meio das duas técnicas, são eliminadas quando a metodologia de regressão linear é aplicada. Pelo fato de alguns requisitos apresentarem iguais estimativas de CS e CI, o ranqueamento dos requisitos que deveria ser distribuído de 1 a 9, passa a ser de 1 a 4. Como os quartis estão distribuídos nos níveis 2,5; 5; e 7,5; todos os requisitos foram classificados como de prioridade alta e muito alta.

Estes altos níveis de prioridade são compatíveis com a técnica de Atribuição Numérica, uma vez que, por meio dela, os requisitos apresentaram elevado grau de importância. A partir da análise das estimativas de CS, CI e GI, os requisitos apresentaram idênticos níveis de prioridades nas duas técnicas de priorização, onde a PGI obtida pela técnica de Atribuição Numérica foi idêntica à PCS e PCI (Tabela 3), apesar da metodologia de obtenção ser completamente diversa.

5. Considerações Finais

Este trabalho propôs uma abordagem da modelagem do processo para a elicitação dos requisitos e a aplicação do Modelo de Kano para a priorização dos mesmos.

A modelagem de processos, utilizando a notação BPMN, contribuiu para um melhor entendimento das tarefas e atividades do processo de negócio, permitindo o apontamento de formas de melhorias e otimização. A aplicação da técnica REMO Adaptado permitiu

a derivação dos requisitos funcionais, não funcionais e regras de negócio através do modelo de processo de negócios, obtendo informações mais concretas, consistentes e condizente com as necessidades do processo para o desenvolvimento de um software mais robusto e funcional. A utilização dessa técnica também permitiu que a dinâmica do processo fosse explicitada, obtendo-se maior precisão e completude dos requisitos.

O modelo de Kano apresentou-se como uma técnica importante para a priorização de requisitos de software, uma vez que leva em consideração a satisfação e insatisfação dos *stakeholders,* além de considerar a percepção da qualidade do produto requerido como um fator relevante para a obtenção dos resultados.

A similaridade observada entre os resultados obtidos no estudo, pelos métodos de Atribuição Numérica e Modelo de Kano para priorização dos requisitos, ocorreu devido à pequena quantidade de requisitos e pelo fato dos *stakeholders* atribuírem elevados graus de importância para os requisitos elicitados do modelo de processo de negócio. Como trabalhos futuros sugere-se um estudo de caso associado a uma maior quantidade de requisitos e um número mais expressivo de *stakeholders*.

6. Referências Bibliográficas

ABPMP. (2013). *Guia para o Gerenciamento de Processos de Negócio. Corpo Comum do Conhecimento - CBOK V3.0*. (2ª ed). Brasil: ABPMP.

Alshehri, S., & Benedicenti, L. (2013). Using the Analytical Hierarchy Process as a Ranking Tool for User Story Prioritization Techniques. *In Proceedings of the 8th Inter. Conf. on Software Eng. Advances - ICSEA 2013, (*p. 329-335). Venice, Italy.

Asghar, A. R., Bhatti, S.N., Tabassum, A., Sultan, Z., & Abbas, R. (2016). Role of Requirements Elicitation & Prioritization to Optimize Quality in Scrum Agile Development. *International Journal of Advanced Computer Science and Applications*, 7(12), 300-306.

Berger, C., Blauth, R., & Boger, D. (1993). Kano's Methods for Understanding Customer-defined Quality. *Center for Quality Management J.*, 2(4), 3-36.

Bitencourt, A., Paiva, D., & Cagnin, M. (2016). Elicitação de Requisitos a partir de Modelos de Processos de Negócio em BPMN: Uma Revisão Sistemática. In *Anais Principais do XII Simpósio Brasileiro de Sistemas de Informação*, (pp. 200-207). Porto Alegre: SBC.

Cordeiro, A. G., & Freitas, A. L. P. (2012). Priorização de requisitos e avaliação da qualidade de software segundo a percepção dos usuários. Ciência da Informação, 40(2), 160-179.

Duda, J., Rostanski, M., Rostanski, Borczyk, W., & Grochla, K. (2015). Applying Kano model into goal/requirements elicitation for crossplatform mobile content technology. In *11th International Conference on Strategic Management and its Support by Information Systems 2015*. (pp, 356-365). Uherske Hradiste, Czech Republic: VSB-Technical University of Ostrava.

Farias, T. M. M. (2010). *Uma proposta de abordagem de levantamento de requisitos baseada em modelagem de processos de negócio*. (Dissertação de Mest. Ciência da Computação), Univ. Federal do Pernambuco (UFPE), Recife.

Freitas, R. L. (2018). *Método de derivação de requisitos de software a partir de modelos de processos de negócio otimizados*. (Dissertação de Mest. Prof. Sistemas Aplicados à Engenharia e Gestão), Instituto Federal Fluminense (IFF), Campos dos Goytacazes.

Gupta, V., Talha, M., & Gupta, C. (2017). Non-functional Requirement Prioritization Approach. International *J. of Soft. Eng. and Its Applications*, 11(1), 61–66.

IEEE. (1990). *IEEE standard glossary of software engineering terminology*. New York, N.Y: Institute of Electrical and Electronics Engineers, 1990.

Machado, F. N. R. (2016). *Análise e Gestão de Requisitos de Software – Onde nascem os sistemas*. (3. ed). São Paulo: Érica.

Matzler, K., & Hinterhuber, H. H. (1998). How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment. *Technovation*, 18(1), 25–38.

Melo, A. C., & Muniz Silva, P. S. (2014). Repositório para o rastreamento de requisitos funcionais derivados de regras de negócio. *Revista Brasileira de Computação Aplicada*, 6(2), 61-75.

Mkpojiogu, E. O. C., & Hashim, N. L. (2017). Quality Based Prioritization: An Approach for Prioritizing Software Requirements. *Journal of Telecommunication, Electronic and Computer Engineering*, 9(2), 17–21.

Olaronke, I., Rhoda, & I., Ishaya, G. (2018). An Appraisal of Software Requirement Prioritization Techniques. *Asian J. of Research in Computer Science*, 1(1), 1–16.

Pereira, R. M., Castro, S. O. C., Marques, H. R., Botelho, L. H. F, Silva, T. S., & Freitas, A. F. (2016). A Informatização de Processos em Instituições Públicas: o caso da Universidade Federal de Viçosa. *Revista Navus*, 6(1), 17–29.

Pinto, D. D., Ladeiras, F. M., Sant'Anna, T. S., & Miyashita, R. (2012). Classificação de Atributos de Satisfação de Clientes do Serviço de Telefonia Móvel. *Revista de Gestão e Operações Produtivas*, 3(1), 1-14.

Pradella, S. (2013). Gestão de processos: uma metodologia redesenhada para a busca de maior eficiência e eficácia organizacional. *Revista Gestão & Tecnologia*, 13(2), 94–121.

Pressman, R. S. (2011). *Engenharia de software: uma abordagem profissional*. (7. ed). Porto Alegre: AMGH Editora.

Roos, C., Sartori, S., & Godoy, L. P. (2009). Modelo de Kano para a identificação de atributos capazes de superar as expectativas do cliente. *Revista Produção Online*, 9(3), 536-550.

Sauerwein, E., Baiolom, F., Matzler, K., Hinterh, H.H. (1996). The Kano Model: How To Delight Your Customers. In *X International Working Seminar on Production Economics*, (pp. 313-327). Innsbruck/Igls/Austria.

Sommerville, I. (2011). *Software engineering*. (9th ed). Boston: Pearson.

Vieira, S. R. C. (2012). *REMO – Uma Técnica de Elicitação de Requisitos orientada pela Modelagem de Processos de Negócios*. (Dissertação de Mestrado em Informática). Universidade Federal do Amazonas, Manaus.

Weilkiens, T., Weiss, C., Grass, A., Duggen, K. N. (2017). OCEB 2 Certification Guide - Business Process Management - Fundamental Level. (2. ed). Morgan Kaufmann.

Xavier, L. (2009). *Integração de Requisitos não Funcionais a Processos de Negócios: Integrando BPMN e NFR*. (Dissertação de Mestrado em Ciência da Computação). Universidade Federal do Pernambuco, Recife.

Critérios Editoriais

A RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação) é um periódico científico, propriedade da AISTI (Associação Ibérica de Sistemas e Tecnologias de Informação), que foca a investigação e a aplicação prática inovadora no domínio dos sistemas e tecnologias de informação.

O Conselho Editorial da RISTI incentiva potenciais autores a submeterem artigos originais e inovadores para avaliação pelo Conselho Científico.

A submissão de artigos para publicação na RISTI deve realizar-se de acordo com as chamadas de artigos e as instruções e normas disponibilizadas no sítio Web da revista (http://www.risti.xyz).

Todos os artigos submetidos são avaliados por um conjunto de membros do Conselho Científico, não inferior a três elementos.

Em cada número da revista são publicados entre cinco a oito dos melhores artigos submetidos.

Criterios Editoriales

La RISTI (Revista Ibérica de Sistemas y Tecnologías de la Información) es un periódico científico, propiedad de la AISTI (Asociación Ibérica de Sistemas y Tecnologías de la Información), centrado en la investigación y en la aplicación práctica innovadora en el dominio de los sistemas y tecnologías de la información.

El Consejo Editorial de la RISTI incentiva autores potenciales a enviar sus artículos originales e innovadores para evaluación por el Consejo Científico.

Lo envío de artículos para publicación en la RISTI debe hacerse de conformidad con las llamadas de los artículos y las instrucciones y normas establecidas en el sitio Web de la revista (http://www.risti.xyz).

Todos los trabajos enviados son evaluados por un número de miembros del Consejo Científico de no menos de tres elementos.

En cada número de la revista se publican cinco a ocho de los mejores artículos enviados.

Os associados da AISTI recebem a RISTI gratuitamente, por correio postal. Torne-se associado da AISTI. Preencha o formulário abaixo e envie-o para o e-mail aistic@gmail.com

Los asociados de la AISTI reciben la RISTI por correo, sin costo alguno. Hazte miembro de la AISTI. Rellena el siguiente formulario y remítelo al e-mail aistic@gmail.com



Formulário de Associado / Formulario de Asociado

Nome/Nombre: _____

Instituição/Institución: _____

Departamento: _____

Morada/Dirección: _____

Código Postal: _____ Localidade/Localidad: _____

País: _____

Telefone/Teléfono: _____

E-mail: _____ Web: _____

Tipo de Associado e valor da anuidade:

☐ Individual - 35€

☐ Instituição de Ensino ou I&D/Institución de Educación o I&D - 250€

☐ Outro (Empresa, etc.) - 500€

NIF/CIF: _____

Data/Fecha: __/__/_____ Assinatura/Firma: _____

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies
©AISTI 2020 http://www.aisti.eu

