



FONDO EDITORIAL

CONGRESO INTERNACIONAL DE INGENIERÍA DE SISTEMAS



UNIVERSIDAD  
DE LIMA

# ENTORNOS HÍBRIDOS EN LA POSPANDEMIA: POSIBILIDADES PARA LAS NUEVAS TECNOLOGÍAS

ACTAS DEL V CONGRESO INTERNACIONAL DE INGENIERÍA DE SISTEMAS



Entornos híbridos en la pospandemia: posibilidades para las nuevas tecnologías  
Actas del V Congreso Internacional de Ingeniería de Sistemas



---

FONDO EDITORIAL

CONGRESO INTERNACIONAL DE INGENIERÍA DE SISTEMAS

---



UNIVERSIDAD  
DE LIMA

# **ENTORNOS HÍBRIDOS EN LA POSPANDEMIA: POSIBILIDADES PARA LAS NUEVAS TECNOLOGÍAS**

ACTAS DEL V CONGRESO INTERNACIONAL DE INGENIERÍA DE SISTEMAS

Congreso Internacional de Ingeniería de Sistemas (5.º : 2022 : Lima, Perú)

*Entornos híbridos en la pospandemia: posibilidades para las nuevas tecnologías. Actas del V Congreso Internacional de Ingeniería de Sistemas / organizado por la Carrera de Ingeniería de Sistemas de la Universidad de Lima. Primera edición digital. Lima: Universidad de Lima, Fondo Editorial, 2022.*

149 páginas: ilustraciones, diagramas, gráficos.

Texto en español e inglés.

Conferencias magistrales, paneles y ponencias del V Congreso Internacional de Ingeniería de Sistemas (CIIS), denominado “Entornos híbridos en la pospandemia: posibilidades para las nuevas tecnologías”, el cual se realizó del 10 al 12 de octubre del 2022.

Incluye referencias. Pósteres: páginas 123-137.

1. Ingeniería de sistemas -- Congresos. 2. Ingeniería del *software* -- Congresos. 3. Sistemas expertos (informática) -- Congresos. 4. Enseñanza en línea -- Congresos. 5. Aprendizaje automático (inteligencia artificial) -- Congresos. I. Actas del V Congreso Internacional de Ingeniería de Sistemas (5.º : 2022: Lima, Perú). II. Universidad de Lima. Fondo Editorial

005.1

C

5

ISBN 978-9972-45-613-8

ISSN 2810-806X (En línea)

*Entornos híbridos en la pospandemia: posibilidades para las nuevas tecnologías*

*Actas del V Congreso Internacional de Ingeniería de Sistemas*

Primera edición digital: diciembre, 2022

© De esta edición:

Universidad de Lima

Fondo Editorial

Av. Javier Prado Este 4600,

Urb. Fundo Monterrico Chico, Lima 33

Apartado postal 852, Lima 100, Perú

Teléfono: 437-6767, anexo 30131

fondoeditorial@ulima.edu.pe

www.ulima.edu.pe

Diseño, edición y carátula: Fondo Editorial de la Universidad de Lima

Imagen de carátula:

Prohibida la reproducción total o parcial de este libro, por cualquier medio, sin permiso expreso del Fondo Editorial.

ISBN 978-9972-45-613-8

ISSN 2810-806X (En línea)

Hecho el depósito legal en la Biblioteca Nacional del Perú n.º 2023-00766

## **PROGRAM COMMITTEE / COMITÉ ACADÉMICO**

- Juan Gutiérrez
- Andrea Matuk
- Nadia Rodríguez

## **ORGANIZING COMMITTEE / COMITÉ ORGANIZADOR**

- Andrea Matuk Chijner
- Carlos Torres Paredes
- Daniel Cárdenas Salas
- Franci Suni Lopez
- Guillermo Antonio Dávila Calle
- Hernan Nina Hanco
- Hernán Quintana Cruz
- José Caballero Ortiz
- Juan Gutiérrez Cárdenas
- Manuela Linares
- Nadia Rodríguez
- Pablo Rojas Jaen
- Rocío del Pilar Checa Fernández
- Rosario Guzmán Jiménez
- Víctor Hugo Ayma Quirita



# Índice

PRESENTACIÓN	11
CONFERENCIAS MAGISTRALES	15
Lenguajes y modelos subyacentes a los grafos de conocimiento <i>Renzo Angles</i>	17
Trustworthy Autonomous Systems (TAS): The Verifiability Approach <i>Mohammad Reza Mousavi</i>	27
Hybrid Work: Is it a Blessing or a Curse? <i>Tatiana Andreeva</i>	31
PANELES	35
SERO Tawa Pukllay: sistema educativo remoto <i>online/offline</i> <i>Rosario Guzmán Jiménez, Eduardo Alejandro Escotto Córdova, Álvaro Saldivar, Dhavit Prem</i>	37
Venciendo los desafíos para la inclusión de mujeres en STEM <i>Guillermo Dávila, Indira Guzmán, Claudia Quintanilla, Cristiano Maciel</i>	43
Desafíos del aprendizaje profundo en la visión por computador <i>Victor Hugo Ayma Quirita, Pedro Marco Achanccaray Díaz, Smith Washington Arauco Canchumuni, Pedro Juan Soto Vega</i>	49
Gamit!: plataforma para la gamificación basada en un sistema de recompensas para estudiantes universitarios <i>Hernán Alejandro Quintana Cruz, Elvira Rincón-Flores, Brenda Santos, Alberto Matsuura</i>	55

PONENCIAS	59
Análisis global de la producción científica respecto del aprendizaje organizacional en la educación superior <i>Liliana Pedraja-Rejas, Emilio Rodríguez-Ponce, Camila Muñoz-Fritis</i>	61
Dynamic Information and Analytical System for Assessing and Managing Acquired Knowledge <i>Yana Goncharova, Olga Sheveleva, Vladimir Dobrynin</i>	77
Estimación de la cantidad de heridos en accidentes de tránsito dentro de la provincia de Lima utilizando modelos de regresión lineal múltiple y PCA <i>Alisson Georgeth Avalos Tapia, Juan David Cárdenas Zúñiga, Rodrigo Fernando Caballero Chocano, Victor Andres Ayma Quirita</i>	91
Algoritmos supervisados de <i>machine learning</i> para determinar la ubicación de dispositivos wifi <i>Javier More Sánchez</i>	111
PÓSTERES	123
MOOC para la educación continua en sistemas de información durante la pandemia <i>Diana Concepción Mex Álvarez, Margarita Castillo Téllez, Francisco Emmanuel Alfaro Juárez</i>	124
Revisión de algoritmos de detección de <i>malware</i> ofuscado basados en <i>machine learning</i> <i>Erly Galia Villarroel Enríquez</i>	132
DATOS DE LOS AUTORES	139

# Presentación

La Carrera de Ingeniería de Sistemas de la Universidad de Lima organizó y presentó el V Congreso Internacional de Ingeniería de Sistemas (CIIS 2022) denominado “Entornos híbridos en la pospandemia: posibilidades para las nuevas tecnologías”, llevado a cabo el 10, 11 y 12 de octubre del año 2022. El congreso tuvo como objetivo la reflexión, discusión y descripción de aplicaciones relacionadas con los entornos híbridos y las nuevas tecnologías.

Los entornos híbridos toman especial relevancia en el contexto actual, caracterizado por las restricciones de movilidad, distanciamiento social y paralización de algunas actividades presenciales. Individuos, organizaciones y sociedades se han ido adaptando a este contexto y han aprendido o potenciado el aprendizaje de herramientas que facilitan la productividad del teletrabajador, el rastreo de personas infectadas, la identificación de nuevas variantes de enfermedades, entre otros. Por un lado, gran cantidad de estas iniciativas fueron desarrollándose como respuesta a urgencias, es decir, de forma reactiva. Por otro lado, las mismas soluciones generaron más interacciones y experiencias virtuales relacionadas con sus contrapartes físicas, incrementando el número de entornos híbridos.

En este contexto, es oportuno explorar de forma proactiva las posibilidades ofrecidas por las nuevas tecnologías para maximizar los resultados de estos entornos híbridos a corto, mediano y largo plazo. Entendemos por resultados no solamente los alcanzados en términos económicos, sino también el potencial innovador que puede surgir de estos entornos, el bienestar de sus miembros o la flexibilidad como competencia que permita adaptarse a futuros cambios generados por disrupciones tecnológicas y otras emergencias sanitarias o sociales.

El CIIS 2022, como las anteriores ediciones, congregó a investigadores de primer nivel y se abrió una verdadera oportunidad de aprendizaje, que es registrada en la presente memoria. En primer lugar, la doctora Tatiana Andreeva, profesora asociada en Gestión y Comportamiento Organizacional, y directora de Investigación en la Escuela de Negocios de la Universidad de Maynooth, Irlanda, impartió la ponencia magistral titulada “Trabajo híbrido: ¿bendición o carga?”, donde se exploran los retos y oportunidades que plantea el trabajo híbrido en nuestra sociedad. En segundo lugar, el doctor Mohammad Reza Mousavi, profesor en el área de ingeniería del *software* en el King’s College de London, Inglaterra, ofreció la ponencia magistral titulada “Sistemas autónomos confiables (TAS): el enfoque de la verificabilidad”, la cual aborda la importancia del uso de los sistemas autónomos en nuestra sociedad y más aún que estos no sean un peligro para las personas. En tercer lugar, el doctor Renzo Angles, profesor en la Universidad de Talca, Chile, presentó su ponencia magistral titulada “Lenguajes y modelos subyacentes a los grafos de conocimiento”, donde pone énfasis en los principios del modelado de bases de datos fundado en la estructura de datos grafo y sus principales aplicaciones en la construcción de los grafos de conocimiento. Subraya el concepto de grafos de conocimiento, que tienen como principio fundamental la abstracción basada en grafos y, de esa manera, se pueden implementar escenarios en general, de múltiples fuentes de datos y a gran escala. Finalmente, el doctor Avid Román-González, profesor principal de la Universidad Nacional Tecnológica de Lima Sur, Perú, desarrolló su ponencia magistral titulada “Procesamiento de imágenes aplicado en temas de salud”. Explica el uso de las técnicas de procesamiento de imágenes en una gran cantidad de aplicaciones en diferentes áreas y dominios. Por ejemplo, se puede usar el procesamiento de imágenes para ayudar en la solución de varios problemas que aquejan a nuestra sociedad, y justamente una de las áreas en las que se usa el procesamiento de imágenes son las ciencias de la salud. El doctor Román-González presenta algunos trabajos, metodologías, técnicas y aplicaciones del procesamiento de imágenes en temas de salud y la ayuda en la detección de ciertas enfermedades; resalta que las herramientas desarrolladas no son para hacer un diagnóstico, ya que esto les corresponde a los profesionales de las ciencias de la salud, pero sí son un gran apoyo para la detección de sospechas de casos<sup>1</sup>.

Asimismo, el CIIS 2022 incorpora en su programa secciones de paneles, donde expertos nacionales y extranjeros presentan sus puntos de vista sobre un tema específico y luego discuten entre ellos y la audiencia sobre aspectos relacionados con el tema del panel. Actúan como moderadores profesores de la Carrera de Ingeniería de Sistemas. Los paneles propuestos para el CIIS 2022 son “SER0 Tawa Pukllay: sistema educativo remoto *online/offline*”, moderado por la profesora Rosario Guzmán; “Venciendo los desafíos para la inclusión de mujeres en

---

1 La exposición del doctor Avid Román-Conzález solo está disponible en el video de evento.

STEM”, moderado por el profesor Guillermo Dávila; “Desafíos del aprendizaje profundo en la visión por computador”, moderado por el profesor Víctor Ayma; y “Gamit!: plataforma para la gamificación basada en un sistema de recompensas para estudiantes universitarios”, moderado por el profesor Hernán Quintana.

Finalmente, como es tradición en el CIIS, investigadores nacionales e internacionales presentaron sus trabajos de investigación en áreas de sistemas de información, ingeniería de *software*, ciencias de la computación, tecnologías de la información o afines a la temática del evento.

*Nadia Rodríguez Rodríguez*

Directora de la Carrera de Ingeniería de Sistemas  
Universidad de Lima



# CONFERENCIAS MAGISTRALES



# Lenguajes y modelos subyacentes a los grafos de conocimiento

Renzo Angles

rangles@utalca.cl

<https://orcid.org/0000-0002-6740-9711>

Universidad de Talca, Chile

Recibido: 7 de septiembre del 2022 / Aceptado: 10 de octubre del 2022

doi: <https://doi.org/10.26439/ciis2022.6065>

**RESUMEN.** Un grafo de conocimiento es una gran base de datos que integra información desde distintas fuentes de datos, con el objetivo de poder extraer conocimiento y transformarlo en valor para los usuarios. Esta base de datos es representada como un grafo, donde los nodos representan entidades, y cada arista representa una relación entre dos nodos o un atributo de un nodo. El objetivo de este artículo es presentar una revisión de los modelos de datos que se usan para representar grafos de conocimiento, y los lenguajes de consulta que permiten extraer la información explícita e implícita contenida en dichos grafos.

**PALABRAS CLAVE:** grafo de conocimiento, modelos de datos basados en grafos, lenguaje de consulta para grafos

## LANGUAGES AND MODELS UNDERLYING KNOWLEDGE GRAPHS

**ABSTRACT.** A knowledge graph is a large database that integrates information from different data sources, this with the objective of supporting knowledge extraction, and allowing the transformation of such knowledge into value for the users. Such a database is represented as a graph where the nodes represent entities, and each edge represents a relationship between nodes, or the attribute of a node. The objective of this article is to review the data models used to represent knowledge graphs, and the query languages that allow to extract explicit and implicit information contained in such graphs.

**KEYWORDS:** knowledge graph, graph data model, graph query language

## 1. INTRODUCCIÓN

Los términos *dato*, *información* y *conocimiento* son fundamentales en áreas del conocimiento de gran interés científico e impacto tecnológico en la actualidad. Entre dichas áreas se encuentran bases de datos (databases), datos masivos (*big data*), web semántica (*semantic web*), inteligencia artificial, entre otros.

Para explicar dichos conceptos, consideremos la siguiente frase: “Rosa Rosales cortó una rosa, que roja es la rosa de Rosa Rosales”.

En términos muy generales, un dato puede definirse como una palabra sin significado obvio, o una palabra ambigua (es decir, puede entenderse o interpretarse de diversas maneras). Por ejemplo, la palabra *rosa* puede tener varios significados: una flor, un nombre propio, un color, una marca o incluso un lugar. Ahora, si acotamos dichos significados a la frase del ejemplo, entonces, *rosa* puede ser un nombre o una flor. Luego, para determinar el significado preciso de una ocurrencia de la palabra *rosa* en la frase del ejemplo, debemos analizar las palabras que aparecen a su alrededor. Por ejemplo, si consideramos el fragmento “Rosa Rosales”, podemos inferir que *Rosa* es un nombre propio; en el caso de “que roja es la rosa”, podemos saber que se refiere a una flor. Es decir, el significado de un dato puede dilucidarse analizando las relaciones que tiene con otros datos. En este sentido, la información se puede definir como un conjunto de datos cuyo significado se infiere de sus relaciones y un contexto. Finalmente, el término *conocimiento* puede definirse como la información implícita que puede extraerse al procesar la información explícita o existente. Por ejemplo, el número de veces que cada palabra aparece y la palabra que aparece el mayor número de veces son dos ejemplos de conocimiento que pueden extraerse de la frase del ejemplo.

Desde el inicio de la era digital, a finales de los años cincuenta, se han investigado problemas y desarrollado soluciones para gestionar datos, información y conocimiento. Inicialmente, la preocupación estaba en almacenar (codificar) y procesar (leer y escribir) los datos. Con la aparición del modelo de datos relacional en 1970, se propusieron distintas formas de modelar (o representar) los datos y extraer (o consultar) la información subyacente. La creación de la Web (en 1989) marca el inicio de otra etapa, una donde las personas y los sistemas empiezan a generar información y conocimiento gracias a estándares para transferir datos (HTTP), generar información (HTML) y representar conocimiento (RDF, RDF Schema, OWL). La Web motiva el desarrollo de sistemas informáticos interconectados, los cuales originan información que se caracteriza por su volumen (cantidad), variedad (heterogeneidad) y velocidad (de generación), dando lugar al concepto de *big data*. En este punto, se empiezan a desarrollar plataformas (como Apache Hadoop) que permiten el procesamiento y análisis de datos masivos, empleando técnicas de almacenamiento distribuido de datos y computación paralela. Las organizaciones se dan cuenta del conocimiento implícito existente en los datos que producen, por lo que empiezan a impulsar proyectos de ciencia de datos (*data science*) con el objetivo de generar valor empleando métodos estadísticos y técnicas avanzadas de minería

de datos. La existencia de grandes cantidades de datos no solo permite extraer conocimiento, sino también que los sistemas de inteligencia artificial aprendan de los datos a fin de lograr predicciones más precisas. Podríamos decir que actualmente nos encontramos en “la era del conocimiento”, ya que estamos desarrollando métodos para representar, analizar y generar conocimiento, o mejor dicho, grafos de conocimiento.

## 2. GRAFOS DE CONOCIMIENTO

No existe una definición estándar para el término *grafo de conocimiento* (*knowledge graph*), pero presentaremos tres definiciones que consideramos relevantes:

- Un grafo dirigido, cuyos nodos son unidades de conocimiento (conceptos) que un estudiante debe adquirir, y cuyas aristas denotan dependencias entre dichas unidades de conocimiento (Schneider, 1973). Esta es la primera definición del término *grafo de conocimiento*, que se incluye en un artículo publicado en 1973.
- Un modelo inteligente (un grafo) que comprende entidades del mundo real y las relaciones entre ellas. Esto corresponde a un fragmento del *post* que se usó para dar a conocer el Google Knowledge Graph (Singhal, 2012).
- Un grafo de datos destinado a acumular y transmitir conocimiento del mundo real, cuyos nodos representan entidades de interés, y cuyas aristas representan relaciones potencialmente diferentes entre dichas entidades. Esta definición es parte de un artículo (Hogan et al., 2021) que revisa modelos, lenguajes, herramientas y dominios de aplicación asociados a los grafos de conocimiento.

Luego de revisar diversos trabajos en el área, nuestra definición de grafo de conocimiento es la siguiente: una gran base de datos que integra información desde distintas fuentes de datos con el objetivo de generar información adicional y conocimiento. Dicha base de datos es representada como un grafo; es decir, las entidades se representan como nodos, y las relaciones entre dichas entidades se representan como aristas.

Actualmente existen diversos modelos, lenguajes y métodos que nos permiten representar, analizar (consultar) y generar grafos de conocimiento. A continuación, describimos algunos de ellos.

## 3. MODELOS PARA REPRESENTAR GRAFOS DE CONOCIMIENTO

El desarrollo de un grafo de conocimiento es un proceso complejo. Una de las primeras tareas es el modelado conceptual, que consiste en identificar tipos de entidades, tipos de relaciones y tipos de restricciones, y representarlos usando un modelo de datos (Brodie et al., 1984). De

una manera muy general, un modelo de datos se define como una colección de herramientas conceptuales usadas para modelar representaciones de entidades del mundo real y las relaciones entre ellas (Silberschatz et al., 1996). Existen distintos modelos de datos (Beerl, 1988), los cuales pueden agruparse según la estructura abstracta en que están basados, por ejemplo, tablas, árboles o grafos.

El modelado conceptual de un grafo de conocimiento puede realizarse usando cualquier modelo de datos, pero es más natural y usual emplear un modelo basado en grafos, ya que está pensado para representar de mejor manera las conexiones entre las entidades. Existen diversos modelos de datos basados en grafos, algunos teóricos (Angles & Gutierrez, 2008) y otros más prácticos (Angles et al., 2017). Los modelos más usados en la actualidad son tres: grafo dirigido etiquetado, grafo con propiedades y grafo RDF.

Un grafo dirigido etiquetado (*directed labeled graph*) (Barceló Baeza, 2013) es una estructura compuesta de nodos y aristas, donde los nodos y las aristas pueden tener identificadores y etiquetas; cada arista conecta un par de nodos; las aristas son dirigidas, ya que tienen un nodo origen y destino; y pueden existir múltiples aristas entre dos nodos (multigrafo). Desde el punto de vista de modelado de datos, un nodo representa una entidad o un valor, y una arista representa una relación entre dos entidades o un atributo de una entidad. Este modelo es muy usado en métodos estadísticos, minería de datos, inteligencia artificial y sistemas de procesamiento masivo de datos.

Un grafo con propiedades (*property graph*) (Angles, 2018) es un grafo dirigido etiquetado, pero tiene una característica extra: cada nodo o arista puede mantener un conjunto (posiblemente vacío) de propiedades, donde una propiedad tiene un nombre (o etiqueta) y un valor. En este modelo, un nodo representa una entidad, una arista representa una relación entre dos entidades, y una propiedad representa una característica específica y propia de una entidad o una relación. Este modelo es muy usado por los sistemas de gestión de bases de datos que soportan el almacenamiento y consulta de grafos (Angles & Gutierrez, 2018).

RDF (*resource description framework*) es un estándar para describir recursos web, que define un modelo de datos basado en grafos. En un grafo RDF existen tres tipos de nodos: recursos web identificados por una especie de URL (llamada IRI), nodos blancos que representan recursos anónimos, y literales que representan datos o valores concretos (como cadenas, números y fechas). Un triple RDF es una tupla de la forma (sujeto, predicado, objeto), donde el sujeto puede ser un recurso web o un recurso anónimo; el predicado suele ser un recurso web (que referencia a un tipo de relación), y el objeto puede ser un recurso web o un valor concreto. Según lo anterior, el sujeto y el objeto serían nodos en el grafo, y el predicado es el identificador de una relación. Nótese que, en comparación con un grafo con propiedades, una arista puede representar una relación o una propiedad. Un grafo RDF es la manera estándar de representar datos en la web semántica.

Existen otros modelos que tratan de sobrellevar las deficiencias de estos modelos (Angles et al., 2022), o han sido diseñados para dominios de aplicación especiales. Por ejemplo, el modelo basado en hipergrafos (*hypergraphs*) (Iordanov, 2010) permite representar relaciones n-arias con más facilidad, pero su implementación es más compleja. RDF\* (Hartig, 2017) extiende el modelo RDF con el objetivo de permitir metadatos sobre las propiedades de un triple RDF. En un trabajo reciente (Lassila et al., en prensa), los autores plantean la idea de un único modelo de datos basado en grafos denominado *OneGraph*, el cual busca unificar RDF y los grafos con propiedades con el objetivo de permitir interoperabilidad entre ambos modelos y facilitar el desarrollo de lenguajes de consulta.

Actualmente existen diversos sistemas de gestión de datos que permiten almacenar y consultar grafos de conocimiento (DB-Engines, 2022). Estos sistemas se pueden dividir en cuatro grupos: sistemas de base de datos para grafos (*graph database systems*), como Neo4j (<http://neo4j.com/>) y TigerGraph (<https://www.tigergraph.com>), que en su mayoría soportan *property graphs*; RDF *triple stores*, como Apache Jena (<https://jena.apache.org/>) y GraphDB (<https://graphdb.ontotext.com>), que son diseñados para gestionar grafos RDF; plataformas de procesamiento distribuido de grafos (*distributed graph processing frameworks*), como Apache Giraph (<http://giraph.apache.org>) y GraphX (<https://spark.apache.org/graphx/>), que permiten manipular grafos dirigidos etiquetados de gran tamaño; y sistemas multimodelo, como Amazon Neptune (<https://aws.amazon.com/es/neptune/>) y Microsoft Azure Cosmos DB (<https://docs.microsoft.com/en-us/azure/cosmos-db/>), que soportan múltiples modelos, usualmente RDF y grafos con propiedades.

#### 4. LENGUAJES PARA CONSULTAR GRAFOS DE CONOCIMIENTO

Un componente clave de cualquier sistema de gestión de datos es su lenguaje de consulta. En términos generales, un lenguaje de consulta es un lenguaje computacional de alto nivel que permite recuperar los datos almacenados en un sistema de gestión de datos (Samet, 1981). En el caso de los sistemas de gestión de datos basados en grafos, un lenguaje de consulta para grafos (*graph query language*) está diseñado para extraer información usando operaciones orientadas a grafos, como patrones y consultas de caminos (Angles, Reutter & Voigt, 2018).

Existen varios trabajos relacionados con los lenguajes de consulta para grafos, los cuales tratan aspectos como su definición (Angles & Gutierrez, 2008; Angles, 2012), expresividad (es decir, los tipos de consultas que el usuario puede expresar) (Angles et al., 2022), complejidad computacional (esto es, el tiempo requerido para evaluar distintos tipos de consultas) (Barceló Baeza, 2013; Angles et al., 2017; Bonifati & Dumbrava, 2019), implementación (Fletcher et al., 2016) y evaluación (Ciglan et al., 2012). El libro de Bonifati et al. (2018) presenta una revisión completa sobre consultas orientadas a grafos.

A pesar de los avances en el desarrollo de lenguajes de consulta para grafos, aún no existe un estándar. Sin embargo, en septiembre del 2019, se inició el proyecto GQL Standard (<https://www.gqlstandards.org/>) cuyo objetivo es definir un lenguaje estándar que se basa en la fusión de cuatro lenguajes existentes: Cypher (Neo4j) (Cypher Query Language, s. f.), PGQL (Oracle) (Van Rest et al., 2013), G-CORE (LDBC) (Angles, Arenas et al., 2018) y GSQL (TigerGraph) (Deutsch et al., 2020).

La principal noción detrás de un lenguaje de consulta para grafos es la búsqueda de patrones (*graph pattern matching*) (Gallagher, 2006). Un patrón de grafo simple es un subgrafo donde los nodos y las aristas pueden ser entidades, relaciones y variables. Los patrones simples se pueden combinar usando operadores relacionales, como reunión (*join*), unión y diferencia, permitiendo definir patrones más complejos (Angles et al., 2022). El resultado de un patrón de grafo puede ser un conjunto de grafos que satisfacen la estructura del patrón, o una tabla donde cada columna es una variable y cada fila contiene asignaciones de variables a nodos o valores.

Aunque muchas consultas reales pueden ser expresadas como patrones de grafos, es muy frecuente necesitar de mecanismos adicionales que permitan explorar la topología del grafo (Angles et al., 2022). Esto nos lleva a la noción de consulta de camino (*path query*) (Barceló et al., 2012), que tiene que ver con navegar a través del grafo usando patrones de camino. Recordemos que, en la teoría de grafos, un camino es una secuencia de nodos y aristas que nos permiten ir desde un nodo origen a un nodo destino (Diestel, 2005). En este sentido, la forma más común de representar consultas de caminos es a través de una expresión de la forma  $(a,p,b)$ , donde  $a$  representa el nodo origen,  $b$  el nodo destino, y  $p$  es una expresión regular (Mendelzon & Wood, 1995). Los aspectos teóricos asociados a las consultas de caminos se han estudiado de manera amplia (Angles et al., 2022; Bonifati et al., 2018), y los sistemas de gestión de bases de datos soportan diversos tipos de consultas de caminos. Sin embargo, aún persiste el desafío de ejecutar eficientemente consultas de caminos en grafos de gran tamaño, debido a la complejidad computacional intrínseca del problema.

Si bien los lenguajes de consulta para grafos permiten extraer información desde los grafos de conocimiento, no son lo suficientemente poderosos para expresar consultas complejas (conocimiento) propias del análisis de grafos (*graph analytics*). Por esta razón, se han desarrollado lenguajes que mezclan operaciones declarativas con programación, como Gremlin (Apache TinkerPop, s. f.); librerías que proveen algoritmos para grafos, como The Neo4j Graph Data Science Library (Neo4j, 2021); y sistemas especiales que permiten procesamiento de grafos a gran escala, como Spark GraphX (<https://spark.apache.org/graphx/>).

## 5. CONCLUSIONES

Actualmente las organizaciones están creando muchos grafos de conocimiento con el objetivo de obtener valor de sus datos. Sin bien existen modelos y lenguajes bien establecidos para representar los grafos de conocimiento, aún hay varios desafíos prácticos relacionados con la consulta y análisis de estos grafos, si se piensa en extraer de manera eficiente el conocimiento oculto.

Además de representar y consultar los datos, necesitamos modelos para representar el conocimiento, así como métodos deductivos e inductivos que permitan generar conocimiento de manera automática. Estos temas son el foco de investigación actual en el área y representan el estado actual de la era del conocimiento.

## REFERENCIAS

- Angles, R. (2012). A comparison of current graph database models. En *4rd International Workshop on Graph Data Management: Techniques and Applications (GDM) (ICDE Workshop)*.
- Angles, R. (2018). The property graph database model. En *Proceedings of the Alberto Mendelzon International Workshop on Foundations of Data Management (AMW)* (vol. 2100). CEUR Workshop Proceedings.
- Angles, R., Arenas, M., Barceló, P., Boncz, P., Fletcher, G., Gutierrez, C., ... Voigt, H. (2018). G-CORE: A core for future Graph Query languages. En *Proceedings of the International Conference on Management of Data (SIGMOD)*. ACM.
- Angles, R., Arenas, M., Barceló, P., Hogan, A., Reutter, J., & Vrgoč, D. (2017). Foundations of modern query languages for graph databases. *ACM Computing Surveys*, 50(5).
- Angles, R., & Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys*, 40(1), 1-39.
- Angles, R., & Gutierrez, C. (2018). An introduction to graph data management. En *Graph data management* (cap. 1). Springer Nature.
- Angles, R., Hogan, A., Lassila, O., Rojas, C., Schwabe, D., Szekely, P., & Vrgoč, D. (2022). Multilayer graphs: A unified data model for graph databases. En *Proceedings of the 5th ACM SIGMOD Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA)*. ACM.
- Angles, R., Reutter, J., & Voigt, H. (2018). Graph query languages. En *Encyclopedia of big data technologies* (pp. 1-8). Springer International Publishing.

- Apache TinkerPop. (s. f.). *Gremlin Query Language*. <https://tinkerpop.apache.org/gremlin.html>
- Barceló, P., Libkin, L., Lin, A. W., & Wood, P. T. (2012). Expressive languages for path queries over graph-structured data. *ACM Transactions on Database Systems*, 37(4), 1-46.
- Barceló Baeza, P. (2013). Querying graph databases. En *Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS)* (pp. 175-188). ACM. <https://doi.org/10.1145/2463664.2465216>
- Beeri, C. (1988). Data models and languages for databases. En *Proceedings of the 2nd International Conference on Database Theory (ICDT)* (vol. 326, pp. 19-40). Springer.
- Bonifati, A., & Dumbava, S. (2019). Graph queries: From theory to practice. *SIGMOD Record*, 47(4), 5-16.
- Bonifati, A., Fletcher, G., Voigt, H., & Yakovets, N. (2018). *Querying graphs*. Morgan & Claypool Publishers.
- Brodie, M. L., Mylopoulos, J., & Schmidt, J. W. (1984). *On conceptual modelling*. Springer-Verlag.
- Ciglan, M., Averbuch, A., & Hluchy, L. (2012). Benchmarking traversal operations over graph databases. En *Proceedings of the International Conference on Data Engineering Workshops* (pp. 186-189). IEEE Computer Society.
- Cypher Query Language*. (s. f.). <http://neo4j.com/developer/cypher-query-language/>.
- DB-Engines. (2022, noviembre). *DB-Engines Ranking of Graph DBMS*. <http://db-engines.com/en/ranking/graph+dbms>.
- Deutsch, A., Xu, Y., Wu, M., & Lee, V. E. (2020). Aggregation support for modern graph analytics in TigerGraph. En *Proceedings of the International Conference on Management of Data (SIGMOD)* (pp. 377-392). ACM.
- Diestel, R. (2005). *Graph theory* (3.ª ed., vol. 173). Springer-Verlag.
- Fletcher, G. H. L., Peters, J., & Poulouvasilis, A. (2016). Efficient regular path query evaluation using path indexes. En *Proceedings of the 19th International Conference on Extending Database Technology* (pp. 636-639). OpenProceedings.org.
- Gallagher, B. (2006). Matching structure and semantics: A survey on graph-based pattern matching. En *AAAI Fall Symposium on Capturing and Using Patterns for Evidence Detection* (pp. 45-53). The AAAI Press.
- Hartig, O. (2017). Foundations of RDF\* and SPARQL\* - An alternative approach to statement-level metadata in RDF. En *Proceedings of the 11th Alberto Mendelzon*

- International Workshop on Foundations of Data Management and the Web* (vol. 1912). CEUR Workshop Proceedings.
- Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., ... Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4).
- Iordanov, B. (2010). HyperGraphDB: A generalized graph database. En *Web-Age Information Management. WAIM 2010. Lecture Notes in Computer Science* (vol. 6185, pp. 25-36). Springer-Verlag.
- Lassila, O., Schmidt, M., Hartig, O., Bebee, B., Bechberger, D., Broekema, W., ... Thompson, B. (en prensa). The OneGraph vision: Challenges of breaking the graph model lock-in. *Semantic Web*, 14.
- Mendelzon, A. O., & Wood, P. T. (1995). Finding regular simple paths in graph databases. *SIAM Journal on Computing*, 24(6), 1235-1258.
- Neo4j. (2022). *The Neo4j Graph Data Science Library*. <https://neo4j.com/docs/graph-data-science/current/>
- Samet, J. (Ed.). (1981). *Query languages. A unified approach. Report of the British Computer Society Query Languages Group*. Heyden University Press.
- Schneider, E. W. (1973). *Course modularization applied: The interface system and its implications for sequence control and data analysis*. <https://files.eric.ed.gov/fulltext/ED088424.pdf>
- Silberschatz, A., Korth, H. F., & Sudarshan, S. (1996). Data models. *ACM Computing Surveys*, 28(1), 105-108.
- Singhal, A. (2012, 16 de mayo). Introducing the Knowledge Graph: things, not strings. *The Keyword*. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
- Van Rest, O., Hong, S., Kim, J., Meng, X., & Chafi, H. (2016). PGQL: A property Graph Query Language. En *Proceedings of the Fourth International Workshop on Graph Data Management Experiences and Systems*.



# Trustworthy Autonomous Systems (TAS): The Verifiability Approach

Mohammad Reza Mousavi  
Mohammad.mousavi@kcl.ac.uk  
<https://orcid.org/0000-0002-4869-6794>  
King's College London, London, UK

Recibido: 11 de septiembre del 2022 / Aceptado: 6 de agosto del 2022  
doi: <https://doi.org/10.26439/ciis2022.6063>

**ABSTRACT.** Autonomous systems are taking over the decision-making in many crucial aspects of our lives. Trust in them will help users benefit from such systems without harming themselves. Establishing the right level of trust involves a holistic validation and verification process, accounting for aspects such as interactions with the physical world and human users. In this talk, I present our ongoing effort to provide a holistic framework for ensuring the verifiability of autonomous systems.

**KEYWORDS:** autonomous systems, trust, verifiability, validation and verification, testing

## SISTEMAS AUTÓNOMOS CONFIABLES (TAS): EL ENFOQUE DE LA VERIFICABILIDAD

**RESUMEN.** Los sistemas autónomos se están haciendo cargo de la toma de decisiones en muchos aspectos cruciales de nuestras vidas. Confiar en ellos ayudará a sus usuarios a beneficiarse de dichos sistemas sin dañarse a sí mismos. Establecer el nivel adecuado de confianza implica un proceso holístico de validación y verificación, que tiene en cuenta aspectos como las interacciones con el mundo físico y los usuarios humanos. En esta charla, presento nuestro esfuerzo continuo para proporcionar un marco holístico para garantizar la verificabilidad de los sistemas autónomos.

**PALABRAS CLAVE:** sistemas autónomos, confianza, verificabilidad, validación y verificación, *testing*

## 1. INTRODUCTION

Autonomous systems are the result of an integration of software, hardware, and communication systems that enables decision-making with minimal intervention required from their users (Mousavi et al., 2022). Examples of such systems include pacemakers and implantable defibrillators, drones and unmanned aerial vehicles (UAVs), and chatbots. Although decision making in such systems is performed autonomously, they often engage in patterns of interactions with users and hence, their usefulness crucially depends on a smooth orchestration of these interactions.

Trust and trustworthiness are crucial aspects in the development and deployment of autonomous systems: it concerns the users' belief that the system is going to be helpful and safe in challenging scenarios (Araujo et al., 2019). Trusting a system that is not trustworthy can harm users, since users will adapt the systems in challenging scenarios that the system cannot cope with. Likewise, not trusting a system that is trustworthy can lead to avoiding the system in scenarios that the system can cope with and hence, not benefitting from the system. Establishing the right level of trust involves gathering and communicating sufficient evidence for the system's safety and usefulness. A holistic validation and verification process is an essential ingredient for providing such evidence (Mousavi et al., 2022; Araujo et al., 2022).

In this talk, I will go through our verifiability framework for autonomous systems. This involves learning about system and user behavior and capturing their appropriate models (Damasceno et al., 2021); adapting the models by observing and adapting to changes (Damasceno et al., 2019; Tavassoli et al., 2022); generating structured test suites that cover different aspects of system and user behavior (Araujo et al., 2020; Biewer et al., 2022); and an analysis of the test results and explaining the patterns of interaction (Gou et al., 2022).

For each of the above-mentioned four steps, we review our latest results, and point out the challenges before us in establishing a holistic verification framework for autonomous systems.

## REFERENCES

- Araujo, H. L. S., Damasceno, C. D. N., Dimitrova, R., Kefalidou, G., Mehtarizadeh, M., Mousavi, M. R., Onime, J., Ringert, J. O., Rojas, J. M., Verdezoto, N. X., & Wali, S. (2019, October 21-23). *Trusted autonomous vehicles: An interactive exhibit* [Conference presentation]. 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), Shenyang, China. <https://doi.org/10.1109/IUCC/DSCI/SmartCNS.2019.00091>
- Araujo, H., Hoenselaar, T., Mousavi, M. R., & Vinel, A. (2020, August 31-September 3). *Connected automated driving: A model-based approach to the analysis of basic awareness*

- services* [Conference presentation]. 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, United Kingdom. <https://doi.org/10.1109/PIMRC48278.2020.9217142>
- Araujo, H., Mousavi, M. R., & Varshosaz, M. (2022). Testing, validation, and verification of robotic and autonomous systems: A systematic review. *ACM Transactions on Software Engineering and Methodology*. <https://doi.org/10.1145/3542945>
- Biewer, S., Dimitrova, R., Fries, M., Gazda, M., Heinze, T., Hermanns, H., & Mousavi, M. R. (2022). Conformance relations and hyperproperties for doping detection in time and space. *Logical Methods in Computer Science*, 18(1). [https://doi.org/10.46298/lmcs-18\(1:14\)2022](https://doi.org/10.46298/lmcs-18(1:14)2022)
- Damasceno, C. D. N., Mousavi, M. R., & da Silva Simao, A. (2019, December 2-6). *Learning to reuse: Adaptive model learning for evolving systems* [Conference presentation]. 15th International Conference, IFM 2019, Bergen, Norway. [https://doi.org/10.1007/978-3-030-34968-4\\_8](https://doi.org/10.1007/978-3-030-34968-4_8)
- Damasceno, C. D. N., Mousavi, M. R., & Simao, A. da S. (2021). Learning by sampling: Learning behavioral family models from software product lines. *Empirical Software Engineering*, 26, 4. <https://doi.org/10.1007/s10664-020-09912-w>
- Gou, M. S., Lakatos, G., Holthaus, P., Wood, L., Mousavi, M. R., Robins, B., & Amirabdollahian, F. (2022, August 29-September 2). *Towards understanding causality – a retrospective study of using explanations in interactions between a humanoid robot and autistic children* [Conference presentation]. 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy. <https://doi.org/10.1109/RO-MAN53752.2022.9900660>
- Mousavi M. R., Cavalcanti A., Fisher M., Dennis L., Hierons R., Kaddouh B., Law E. L., Richardson R., Ringert J. O., Tyukin I., & Woodcock J (2022). *Trustworthy autonomous systems through verifiability*. IEEE Software.
- Tavassoli, S., Damasceno, C. D. N., Khosravi, R., & Mousavi, M. R. (2022, September 12-16). *Adaptive behavioral model learning for software product lines* [Conference presentation]. Proceedings of the 26th ACM International Systems and Software Product Line Conference Graz, Austria. <https://doi.org/10.1145/3546932.3546991>



# Hybrid Work: Is it a Blessing or a Curse?

Tatiana Andreeva

Tatiana.andreeva@mu.ie

<https://orcid.org/0000-0002-4045-7254>

Maynooth University, Ireland

Recibido: 26 de septiembre del 2022 / Aceptado: 10 de octubre del 2022

doi: <https://doi.org/10.26439/ciis2022.6066>

**ABSTRACT.** Enabled by new technologies and informed by the experiences during COVID pandemic, many employees would like to continue working remotely (Teevan et al., 2022), as this enhances their work-life balance and well-being (George et al., 2021). Many leaders, on the contrary, would like to see their employees back to the office (Microsoft, 2022). To address this challenge, many organisations are considering moving to the hybrid work arrangements, allowing their employees to work some days in the remote settings, but requesting them to be in the office some other days (e.g., Barerro et al., 2022). This approach is often seen as a solution that meets the interests of both employees and employers and can capture the benefits of both remote and in-office work (e.g., Choudhury et al., 2022). But does hybrid work really work?

**KEYWORDS:** hybrid work, remote work, HRM, collaboration, line managers, management challenges

## TRABAJO HÍBRIDO: ¿BENDICIÓN O CARGA?

**RESUMEN.** Habilitados por las nuevas tecnologías e informados por las experiencias durante la pandemia del COVID-19, a muchos empleados les gustaría continuar trabajando de forma remota (Teevan et al., 2022), ya que esto mejora su equilibrio y bienestar entre el trabajo y la vida personal (George et al., 2021). Muchos líderes, por el contrario, quisieran ver a sus empleados de vuelta en la oficina (Microsoft, 2022). Para hacer frente a este desafío, muchas organizaciones están considerando cambiar a acuerdos de trabajos híbridos, permitiendo que sus empleados trabajen algunos días en entornos remotos, pero solicitándoles que estén en la oficina otros días (p. ej., Barerro et al., 2022). Este enfoque a menudo se ve como una solución que satisface los intereses tanto de los empleados como de los empleadores y puede capturar los beneficios tanto del trabajo remoto como en la oficina (p. ej., Choudhury et al., 2022). Pero ¿realmente funciona el trabajo híbrido?

**PALABRAS CLAVE:** trabajo híbrido, trabajo remoto, HRM, colaboración, gerentes de línea, desafíos de gestión

## 1. IN THIS TALK

This talk explores the effects of hybrid work on individual employees and organisations, building on the insights from previous research (e.g., Cristea & Leonardi, 2019; Gajendran & Harrison, 2007; Golden & Veiga, 2005; Rockmann & Pratt, 2015) and on recent emerging evidence (e.g., Choudhury et al., 2022; Zappa et al., 2022). How do work arrangements of the employees—in the office, hybrid or remote— influence their creativity and performance? How these work arrangements impact knowledge sharing and collaboration among the colleagues in the organisation? How do line managers react to the different work arrangements? Analysis of the emerging evidence on these questions suggests that hybrid work may bring not only benefits, but also serious risks—both for employees and organisations—and thus needs to be carefully managed. Some ideas on how these risks could be addressed are discussed.

## REFERENCES

- Barerro, J., Bloom, N., & Davis, S. J. (2022). *Why working from home will stick*. National Bureau of Economic Research (working paper 28731). <https://www.nber.org/papers/w28731>
- Choudhury, P., Khanna, T., Makridakis, C., & Schirmann, K. (2022). *Is hybrid work the best of both worlds? Evidence from a field experiment*. Harvard Business School (working paper 22-063). [https://www.hbs.edu/ris/Publication%20Files/22-063\\_639195cc-e7b5-47d3-9281-62d192c5b916.pdf](https://www.hbs.edu/ris/Publication%20Files/22-063_639195cc-e7b5-47d3-9281-62d192c5b916.pdf)
- Cristea, I. C., & Leonardi, P. M. (2019). Get noticed and die trying: Signals, sacrifice, and the production of face time in distributed work. *Organization Science*, 30(3), 552-572. <https://doi.org/10.1287/orsc.2018.1265>
- Gajendran, R. S., & Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: Meta-analysis of psychological mediators and individual consequences. *Journal of Applied Psychology*, 92(6), 1524-1541. <https://www.apa.org/pubs/journals/releases/apl-9261524.pdf>
- George, T. J., Atwater, L. E., Maneethai, D., & Madera, J. M. (2021). Supporting the productivity and wellbeing of remote workers: Lessons from COVID-19. *Organizational Dynamics*, 51(2), 100869. <https://doi.org/10.1016/j.orgdyn.2021.100869>
- Golden, T. D., & Veiga, J. F. (2005). The impact of extent of telecommuting on job satisfaction: Resolving inconsistent findings. *Journal of Management*, 31(2), 301-318. <https://doi.org/10.1177/0149206304271768>
- Kane, G. C., Nanda, R., Phillips, A., & Copulsky, J. (2021). Redesigning the post-pandemic workplace. *MIT Sloan Management Review*, 62(3), 12-14.

- Microsoft (2022). *Great expectations: Making hybrid work work. Work trend index: Annual report*.  
<https://www.microsoft.com/en-us/worklab/work-trend-index/great-expectations-making-hybrid-work-work>
- Mortensen, H. (2021, February 24). How to make hybrid workplace fair. *Harvard Business Review*. <https://hbr.org/2021/02/making-the-hybrid-workplace-fair>
- Rockmann, K. W., & Pratt, M. G. (2015). Contagious offsite work and the lonely office: The unintended consequences of distributed work. *Academy of Management Discoveries*, 1, 150-164. <https://doi.org/10.5465/amd.2014.0016>
- Teevan, J., Baym, N., Butler, J., Hecht, B., Jaffe, S., Nowak, K., Sellen, A., & Yang, L. (Eds.). *Microsoft New Future of Work Report 2022. Microsoft Research Tech Report MSR-TR-2022-3*. <https://www.microsoft.com/en-us/research/uploads/prod/2022/04/Microsoft-New-Future-Of-Work-Report-2022.pdf>
- Zappa, P., Andreeva, T., & Dávila, G. (2022, July 7-22). *Hybrid work arrangements, asymmetry in task dependence and knowledge network churn in the (post) COVID-19 workplace* [Conference presentation]. 38th European Group of Organization Studies Colloquium, Vienna, Austria.



# PANELES



# SERO Tawa Pukllay: sistema educativo remoto *online/offline*

Rosario Guzmán Jiménez (moderadora)

rguzman@ulima.edu.pe

<https://orcid.org/0000-0002-4550-7935>

Universidad de Lima, Perú

Eduardo Alejandro Escotto Córdova

<https://orcid.org/0000-0002-1104-8195>

Universidad Nacional Autónoma de México

Álvaro Saldívar

yachay@yupanainka.com

<https://orcid.org/0000-0003-4250-0794> Asociación

Yupanki Perú

Dhavit Prem

dhavitprem@gmail.com

<https://orcid.org/0000-0003-2353-6214>

Universidad Nacional Mayor de San Marcos, Perú

Recibido: 3 de septiembre del 2022 / Aceptado: 5 de octubre del 2022

doi: <https://doi.org/10.26439/ciis2022.6069>

**RESUMEN.** La recuperación del conocimiento matemático inca a través del método Tawa Pukllay es un recurso de la alternancia semiótica que facilita el aprendizaje de la aritmética básica. Paralelamente, la educación rural presenta muchos retos que se han puesto de manifiesto en la pandemia del COVID-19. El sistema *online/offline* fue una respuesta concreta para que, a través de un juego serio con una interfaz bilingüe (quechua y español) en tabletas, sea posible la continuidad de estudio y aprendizaje de la aritmética básica en niños de zonas rurales de una comunidad andina del Perú. SERO Tawa Pukllay estimuló la iniciación de los niños en el manejo de dispositivos electrónicos manteniendo su identidad, reforzando la lingüística

quechua, la cosmovisión andina, así como su autonomía. Esto permitió el autoaprendizaje y motivación para superar los retos planteados en su proceso de aprendizaje de lectura de números y operaciones aritméticas, soportado en un enfoque pedagógico lúdico y el desarrollo de su pensamiento matemático.

PALABRAS CLAVE: Tawa Pukllay, *yupana*, educación *online/offline*, juegos serios, matemáticas

## SERO TAWA PUKLLAY: ONLINE/OFFLINE REMOTE EDUCATIONAL SYSTEM

ABSTRACT. The recovery of ancestral Inca mathematical knowledge through the Tawa Pukllay method is a semiotic alternative that makes learning basic arithmetic operations easier. Rural education presents many challenges made evident by the COVID-19 pandemic. The online/offline system was a concrete response to some of them, allowing rural children in an Andean community in Peru to continue learning basic arithmetic by playing a serious game with a bilingual (Quechua and Spanish) interface on electronic tablets. SERO Tawa Pukllay stimulated children's initiation to use electronic devices while maintaining their identity, reinforcing the use of Quechua, their Andean worldview, and their autonomy. The process fostered self-learning and motivation as the children strove to complete the challenges posed by learning to read numbers and solve arithmetic problems presented in a playful pedagogical approach to developing their mathematical thinking

KEYWORDS: Tawa Pukllay, *yupana*, online/offline education, serious game, mathematics

## 1. INTRODUCCIÓN

En el siglo XIX, el sistema educativo estuvo alineado a la Revolución Industrial, preparando a los estudiantes para el trabajo; pero hoy se le exige a este sistema no solo la enseñanza de las competencias del siglo XIX (Organización para la Cooperación y el Desarrollo Económicos, 2019); es decir, no solo una formación en conocimientos generales, sino un aprendizaje a lo largo de toda la vida y no limitado a la escuela, sino en múltiples entornos de aprendizaje. El mandato social ha cambiado; estamos en una época de extinción de los sistemas obsoletos y del nacimiento de sistemas educativos innovadores, los cuales no deben ser el resultado de *copy-paste* de lo que hacen los países del primer mundo, porque no contamos con presupuestos para financiar estos modelos; necesitamos soluciones coherentes y flexibles a costos eficientes, que se puedan adaptar a las distintas características de nuestras poblaciones con una base cultural heterogénea, además de mecanismos que nos ayuden a cerrar la brecha existente (Aguerrondo, 2020).

En el Perú, la educación primaria rural (6-11 años) se imparte en escuelas rurales, de las cuales más del 50 % se encuentran en centros poblados que distan más de dos horas de la capital de provincia y tienen menos de 500 habitantes; la brecha de la conexión a internet en hogares es del 1,4 % en el ámbito rural frente al 23,4 % en el urbano. La educación cultural bilingüe es reducida debido a la diversidad de idiomas: el 26 % tiene como segunda lengua el aimara, el 39,8 % el quechua y sus variantes, shipibo el 6,9 %, el asháninka el 4,5 %, y el awajún el 2,1 % (Ministerio de Educación del Perú [MINEDU], 2020). Las comunidades campesinas y nativas (Ministerio de Desarrollo Agrario y Riego, 2019) están conformadas en su mayoría por campesinos u oriundos de una determinada localidad, con costumbres y tradiciones muy arraigadas; su trabajo es comunal, viven en pobreza o extrema pobreza, y muchas veces poseen un idioma o dialecto propio. Entre los profesores, el 43,2 % son polidocentes multigrado en el ámbito rural, comparado con el 4,5 % en el ámbito urbano; y el 12 % son unidocentes multigrado, comparado con el 0,3 % en el ámbito urbano (MINEDU, 2020); muchos de ellos no cuentan con los instrumentos o metodologías adecuadas para enfrentar estas realidades, además de la alta carga administrativa y el hecho de que cumplen diversidad de roles en la comunidad.

El aprendizaje es uno de los componentes fundamentales del proceso cognitivo del ser humano, que combina factores cognitivos, emocionales y ambientales para adquirir o mejorar sus conocimientos y habilidades. El aprendizaje evoluciona con el ser humano; por ejemplo, ha habido una evolución desde el conductismo (Watson, 1913), el cognitivismo (Merrill et al., 1981), el constructivismo (Bednar et al., 1991) hasta la teoría del socioconstructivismo (Vygotsky, 1962/1986), donde el aprendizaje tiene lugar dentro de un entorno sociocultural.

En pandemia todos estos aspectos han afectado los aprendizajes, lo que se ve reflejado en los resultados obtenidos. En el ámbito urbano, el nivel de desempeño satisfactorio en matemática fue alcanzado por el 33,1 % de los alumnos de cuarto grado de primaria y por el 18 % de los de segundo grado, mientras que en el ámbito rural solo llegó a este nivel el 13,0 % de los alumnos de cuarto grado de primaria, y el 11 % de los de segundo grado (MINEDU, 2020).

Desde hace mucho tiempo se ha estado investigando sobre esta problemática, que no es única del Perú, sino que se enfrenta en varios países latinoamericanos. El libro *Alternancias semióticas: estrategia didáctica en la enseñanza de las matemáticas* (Escotto Córdova, 2021) intenta analizar esta problemática, y su autor encontró en el método Tawa Pukllay, desde este enfoque, un recurso didáctico para enseñar matemáticas.

## 2. PRESENTACIÓN

El sistema SER0 Tawa Pukllay buscó apoyar el proceso de aprendizajes profundos significativos en la iniciación de las matemáticas, permitiendo a los estudiantes resolver problemas de manera lúdica, como la lectura de los números bajo el sistema inca y su traducción indo-arábica, y la resolución de problemas aritméticos básicos sin tener que memorizar operaciones. Creemos que es necesario desarrollar nuevos modelos mentales, tales como el pensamiento matemático, que fortalezcan su proceso de aprendizaje. Para ello, se tuvo en cuenta tres aspectos: la motivación, la atención y la dinámica para reforzar su aprendizaje, de manera que sea permanente. Esto se hizo a través de un juego serio en una tableta adecuada al contexto sociocultural de una comunidad andina en pandemia, para ayudar a cubrir la brecha digital.

El sistema *online/offline* SER0 Tawa Pukllay (SER0-TP) incluye el juego serio y un sistema de soporte pensado para contextos de epidemias. Contiene videotutoriales para el proceso de apoyo al autoaprendizaje, así como mecánicas, dinámicas y elementos estéticos propios de la comunidad. Dicho juego implementa el método Tawa Pukllay de la Yupana Inca (YITP) (Prem, 2016), que tiene una naturaleza semiótica (signos y significados); adicionalmente, disminuye la carga de la memoria de trabajo cognitiva requerida por el sistema aritmético indo-árabe, debido a que la notación y las operaciones de YITP dependen de las relaciones visoespaciales, el reconocimiento de patrones, los movimientos sencillos y la visualización a tiempo completo de las cantidades y las operaciones realizadas.

Los hallazgos de la investigación sobre SER0 Tawa Pukllay sugieren que los niños de una comunidad remota, con rezago educativo y tecnológico, en condiciones de aislamiento temporal por la pandemia, sin maestro y con solo dos meses de actividad lúdica autogenerada, no tienen dificultad para aprender dos nuevos contenidos: (1) utilizar la tecnología moderna y (2) aprender la lógica de las matemáticas incaicas instaladas en una tableta. YITP se mostró como una poderosa alternancia semiótica que, incrustada en una tableta electrónica (SER0-TP), fomenta un aprendizaje eficaz y rápido, basado en un esquema que estimula el autoaprendizaje lúdico.

La Asociación Yupanki busca recuperar las ciencias ancestrales andinas, con particular énfasis en la matemática inca, para desarrollar nuevas metodologías pedagógicas potentes y amigables, así como para el desarrollo de nuevas tecnologías basadas en este paradigma. La difusión de este conocimiento a través del método Tawa Pukllay tiene evidencia empírica en el ámbito educativo básico tanto en personas neurotípicas como en personas con condiciones

cognitivas especiales (discalculia, acalculia, autismo, TDAH, discapacidad visual, entre otros). Se han diseñado estrategias tanto individuales como grupales (*atipanakuy*), cuyas aplicaciones prometen también tener un impacto positivo en el ámbito de la salud mental (terapia de recuperación cognitiva). TP cuenta también con validación formal matemática (validación lógica y algebraica), así como con evidencias científicas en aplicaciones como el SERO-TP, desarrollo del pensamiento computacional en niños y la patente en certificación de un sumador para ALU en paralelo, entre otras investigaciones actuales que ofrecen alcances incluso dentro de la computación cuántica.

### 3. CONCLUSIONES

SERO-TP es una herramienta didáctica que apoya las necesidades individuales de cada alumno para aprender el sentido numérico y la noción de valor posicional. Es fundamental desarrollar estas habilidades iniciales que marcarán las trayectorias educativas a largo plazo de estos niños en matemáticas. Adicionalmente, las propiedades semióticas del YITP, tanto numéricas como didácticas, pueden ser una herramienta valiosa para la lectura de números en niños de primaria tanto de la ciudad como de lugares rurales, pero, sobre todo, para aquellos que por sus condiciones socioeconómicas, culturales y sociales no han tenido más apoyo académico que el que les brindan en sus escuelas sus maestros y amigos. Igualmente, se ha identificado a través de pruebas empíricas que el YITP tiene la representación simbólica del cero con valor de lugar en una matriz visoespacial, y que su representación visoespacial, utilizada como alternancia semiótica, facilita el dominio del aprendizaje de los números indo-árabes con ceros con valor de lugar.

### REFERENCIAS

- Aguerrondo, I. (2020, 5 de mayo). *Conferencia “Una mirada hacia el futuro de la educación”* [Video]. YouTube. [https://www.youtube.com/watch?v=KvbG2rO\\_Dnw](https://www.youtube.com/watch?v=KvbG2rO_Dnw)
- Bednar, A. K., Cunningham, D., Duffy, T. M., & Perry, J. D. (1991). Theory into practice: How do we link? En G. J. Anglin (Ed.), *Instructional technology: Past, present, and future* (pp. 88-101). Libraries Unlimited.
- Escotto Córdova, E. A. (Ed.). (2021). *Alternancias semióticas: estrategia didáctica en la enseñanza de las matemáticas*. Universidad Nacional Autónoma de México; Facultad de Estudios Superiores Zaragoza.
- Grupo de Análisis para el Desarrollo. (2020). *Cifras sobre la educación rural en el Perú*. Proyecto CREER. <http://www.grade.org.pe/creer/educacion-rural-en-el-peru/cifras/>

- Guzmán Jiménez, R., Prem, D., Saldívar, A., & Escotto Córdova, E. A. (en prensa). Semiotic alternations with the Yupana Inca Tawa Pukllay in the gamified learning of numbers at a rural Peruvian school. *Educational Technology & Society*, 26(1), 79-94.
- Ministerio de Educación del Perú. (2020). *Informe de resultados de la Evaluación Censal de Estudiantes 2018*. <http://umc.minedu.gob.pe/informe-de-resultados-de-la-evaluacion-censal-de-estudiantes-2018/>
- Ministerio de Desarrollo Agrario y Riego. (2019). *Comunidades campesinas y comunidades nativas*. <https://www.ana.gob.pe/2019/consejo-de-cuenca/mantaro/MA>
- Merrill, M. D., Kowalis, T., & Wilson, B. G. (1981). Instructional design in transition. En F. H. Farley & N. J. Gordon (Eds.), *Psychology and education: The state of the union* (pp. 298-348). McCutchan.
- Organización para la Cooperación y el Desarrollo Económicos. (2019). *OECD FUTURE OF EDUCATION AND SKILLS 2030*. [https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD\\_Learning\\_Compas\\_2030\\_Concept\\_Note\\_Series.pdf](https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD_Learning_Compas_2030_Concept_Note_Series.pdf)
- Piaget, J. (1983). Piaget's theory. En P. Mussen (Ed.), *Handbook of Child Psychology* (4.ª ed., vol. 1). Wiley.
- Prem, D. (2016). *Yupana Inka – Decodificando la Matemática Inka. Método Tawa Pukllay*. Asociación Yupanki.
- Vygotsky, L. (1986). *Language and thought*. MIT Press. (Obra original publicada en 1962).
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20(2), 158-177.

# Venciendo los desafíos para la inclusión de mujeres en STEM

Guillermo Dávila (moderador)

gdavila@ulima.edu.pe

<https://orcid.org/0000-0002-1479-2585>

Universidad de Lima, Perú

Indira Guzmán

irguzman@cpp.edu

<https://orcid.org/0000-0003-3670-7270>

California State Polytechnic University, Estados Unidos

Cristiano Maciel

cristiano.maciel@ufmt.br

<https://orcid.org/0000-0002-2431-8457>

Universidade Federal de Mato Grosso, Brasil

Claudia Quintanilla

claudia.quintanilla@rextie.com

Rextie Perú

Recibido: 3 de septiembre del 2022 / Aceptado: 5 de octubre del 2022

doi: <https://doi.org/10.26439/ciis2022.6067>

**RESUMEN.** Los cambios disruptivos en organizaciones y sociedades originados en áreas como la tecnología, economía y medioambiente desafían la flexibilidad de empresas e instituciones. En este contexto, se ha incrementado la demanda por profesionales e investigadores especialmente en áreas como STEM. La falta de investigadores y profesionales en STEM es un problema, en particular en los países latinoamericanos. En ese sentido, la inclusión de mujeres en STEM ha sido catalogada como un desafío clave y, en consecuencia, se ha convertido en un tópico que viene recibiendo creciente atención por parte de académicos y líderes políticos. Existen diversos proyectos en curso (como el Proyecto ELLAS, financiado por el International Development Research Centre, Canadá) que vienen trabajando en distintas

iniciativas para cerrar las brechas de género en STEM. En el presente panel, discutiremos las principales barreras existentes en Latinoamérica para la inclusión de mujeres en este campo. Además, abordaremos iniciativas y buenas prácticas que han servido para mitigar algunas de esas barreras; y, finalmente, analizaremos la aplicabilidad de estas medidas exitosas en contextos como el peruano.

PALABRAS CLAVE: mujeres, inclusión, liderazgo, STEM

## OVERCOMING THE CHALLENGES FOR THE INCLUSION OF WOMEN IN STEM

ABSTRACT. Disruptive changes in organizations and societies that emerged from technology, economy, and environmental areas, challenge the flexibility of companies and institutions. In this context, the demand for professionals and researchers has increased, especially in areas such as STEM. The lack of researchers and professionals in STEM is a problem, especially in contexts such as Latin America. In this sense, the inclusion of women in STEM is seen as a key challenge. Consequently, it has received increasing attention from scholars and policy-makers. Several ongoing projects (e.g., ELLAS Project, financed by the International Development Research Center – Canada) have been working on different initiatives to close gender gaps in STEM. In this panel, we will discuss the main barriers to the inclusion of women in STEM in Latin America. In addition, we will address initiatives and good practices that have contributed to mitigating some of these barriers. Then, we will discuss these successful practices' applicability in contexts such as the Peruvian.

KEYWORDS: women, inclusion, leadership, STEM

## 1. INTRODUCCIÓN

El contexto actual, caracterizado por su creciente dinamismo y turbulencia, viene desafiando no solo la competitividad, sino también la flexibilidad de las instituciones (North & Varvakis, 2016; Dávila & Dos Anjos, 2021). Organizaciones públicas y privadas empiezan a perseguir resultados sostenibles, más allá de solo reconocimiento o lucratividad, respectivamente (Alberti & Varon Garrido, 2017). El postulado de Schumpeter (1976), que caracteriza a la innovación como un componente intrínseco de las organizaciones que pretenden sobrevivir, sigue más vigente que nunca. Y, en el contexto actual, las metas relacionadas con la sostenibilidad e innovación pueden ser resueltas en gran parte con el uso inteligente de la tecnología (Pan et al., 2022).

El contexto descrito marca claramente una mayor necesidad de profesionales calificados en tecnología e ingeniería. De hecho, un estudio reciente del Foro Económico Mundial (2020) mostró que la demanda laboral más creciente hacia el 2025 corresponde a cinco especialidades que tienen que ver con tecnología: especialistas en *data analytics*, inteligencia artificial, *big data*, *marketing* digital y automatización de procesos. Sin embargo, el mismo reporte muestra que las mujeres fueron más vulnerables y propensas a perder sus posiciones de trabajo, en comparación con los hombres, durante la crisis del 2020.

Esta constatación es desafortunada y va contra una orientación global que busca más y mejores especialistas en ciencia y tecnología. Esta brecha de especialistas por ausencia o vulnerabilidad de mujeres en STEM se agrava más en países emergentes. Por ejemplo, el Reporte de Competitividad Global del Foro Económico Mundial (Schwab & Zahidi, 2020) presenta el indicador “Firmas buscando la diversidad, equidad e inclusión para potenciar la creatividad”, el cual incluye equidad de género en los lugares de trabajo. Schwab y Zahidi (2020) muestran que mientras países desarrollados como Australia, Finlandia o Estados Unidos tienen al indicador en valores altos (arriba de 70/100), países como Brasil, Chile o México solo alcanzan niveles medios (entre 57 y 52). Esta situación tiende a inhibir el ingreso de mujeres en ciertos campos, como STEM. Por ejemplo, en Brasil, mientras las estudiantes universitarias mujeres representan el 50 % del total, al analizar carreras STEM, las mujeres constituyen solo el 20 % del universo (Frigo & Maciel, 2019). Así, el cerrar la brecha de género en STEM, garantizando un mayor acceso de mujeres a estos campos, se torna un desafío para países emergentes que buscan aumentar su masa de especialistas calificados. Vencer estos desafíos tiene una relevancia no solo social, sino también económica, ya que permitirá mejorar la competitividad institucional, y de forma agregada, del país donde los actores socioeconómicos se encuentran.

Para vencer los desafíos en torno a la inclusión y liderazgo de mujeres en STEM, se han implementado diversas iniciativas en Latinoamérica con resultados alentadores. Una de ellas es el Proyecto Meninas Digitais (Frigo & Maciel, 2019), creado en el año 2011 con el apoyo de la Sociedad Brasileña de Computación. Este proyecto tiene como objetivo aproximar a chicas de educación escolar básica y media a las carreras de computación y otras relacionadas; para

ello, integra diversos proyectos de enseñanza, investigación y extensión. Al 2022, ya cuenta con 87 proyectos activos que promueven el acceso de chicas a la computación en todo Brasil. Otra iniciativa más reciente es el Proyecto ELLAS, financiado por el International Development Research Centre, Canadá; además de formar una importante red de investigadores latinoamericanos especialistas del tema, tiene como objetivo desarrollar una plataforma de datos abiertos con políticas e iniciativas orientadas a favorecer la inclusión y liderazgo de mujeres en STEM. El Proyecto ELLAS se encuentra analizando esta temática en diferentes contextos latinoamericanos, con especial atención en Brasil, Bolivia y Perú. Las iniciativas mencionadas, junto a otras, y su potencial aplicabilidad e impacto en nuevos contextos serán discutidas por el presente panel.

## 2. PRESENTACIÓN

El panel cuenta con la presencia de participantes con diferentes *backgrounds* en lo que respecta al ámbito laboral, profesional y de cultura nacional. Ello enriquece la discusión a través de la diversidad de puntos de vista discutidos en torno al mismo objetivo: vencer los desafíos para la inclusión y liderazgo de mujeres en STEM.

## 3. CONCLUSIONES

Garantizar la inclusión y permanencia de mujeres en las carreras STEM significa asegurar que los mejores talentos estén actuando en dichos campos. Consecuentemente, este desafío tiene una relevancia más allá de lo social, pues impacta también en el crecimiento de los ecosistemas productivos y, por tanto, en la economía. Las buenas prácticas existentes, como el Proyecto Meninas Digitais, se deben replicar en los diversos ámbitos en Latinoamérica, no solo a nivel macro, sino también en nuestras acciones diarias como gestores empresariales, líderes de opinión, docentes y colegas. La literatura identifica algunas barreras al acceso y liderazgo de mujeres STEM relacionadas con los estereotipos, la motivación, la autoeficacia y la orientación hacia la equidad de género en los diseños, estrategias y acciones institucionales, entre otros. Se deben identificar las barreras más influyentes en Latinoamérica y, a partir de allí, desplegar prácticas que permitan mitigarlas, facilitando el acceso y liderazgo de mujeres en campos STEM.

## REFERENCIAS

Alberti, F. G., & Varon Garrido, M. A. V. (2017). Can profit and sustainability goals co-exist? New business models for hybrid firms. *Journal of Business Strategy*, 38(1), 3-13. <https://doi.org/10.1108/JBS-12-2015-0124>

- Dávila, G. A., & Dos Anjos, E. C. (2021). Configurations of knowledge management practices, innovation, and performance: Exploring firms from Brazil. *International Journal of Innovation Management*, 25(06), 2150065. <https://doi.org/10.1142/S1363919621500651>
- Frigo, L., & Maciel, C. (2019). Programa Meninas Digitais: inspirando a Nova Geração. *Computação Brasil*, 41, 26-29. [http://www.sbc.org.br/images/flippingbook/computacaobrasil/computa\\_41/pdf/CompBrasil\\_41.pdf](http://www.sbc.org.br/images/flippingbook/computacaobrasil/computa_41/pdf/CompBrasil_41.pdf)
- North, K., & Varvakis, G. (Eds.). (2016). *Competitive strategies for small and medium enterprises. Increasing crisis resilience, agility and innovation in turbulent times*. Springer. <https://link.springer.com/book/10.1007/978-3-319-27303-7>
- Pan, W., Xie, T., Wang, Z., & Ma, L. (2022). Digital economy: An innovation driver for total factor productivity. *Journal of Business Research*, 139, 303-311. <https://doi.org/10.1016/j.jbusres.2021.09.061>
- Schumpeter, J. A. (1976). *Capitalism, socialism, and democracy*. George Allen & Unwin.
- Schwab, K., & Zahidi, S. (2020). *Global competitiveness report: special edition 2020*. World Economic Forum. [https://www3.weforum.org/docs/WEF\\_TheGlobalCompetitivenessReport2020.pdf](https://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2020.pdf)
- World Economic Forum. (2020). *The future of jobs report*. [https://www3.weforum.org/docs/WEF\\_Future\\_of\\_Jobs\\_2020.pdf](https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf)



# Desafíos del aprendizaje profundo en la visión por computador

Víctor Hugo Ayma Quirita (moderador)

vayma@ulima.edu.pe

<https://orcid.org/0000-0002-0284-2610>

Universidad de Lima, Perú

Pedro Marco Achanccaray Díaz

p.diaz@tu-braunschweig.de

<https://orcid.org/0000-0002-7324-9611>

Institute of Geodesy and Photogrammetry

Technical University of Braunschweig, Alemania

Smith Washington Arauco Canchumuni

saraucoc@puc-rio.br

<https://orcid.org/0000-0003-0812-0085>

Pontifical Catholic University of Rio de Janeiro, Brasil

Pedro Juan Soto Vega

pjsotove@ifremer.fr

<https://orcid.org/0000-0001-5396-8531>

Institut Français de Recherche pour l'Exploitation de la Mer, Francia

Recibido: 3 de septiembre del 2022 / Aceptado: 5 de octubre del 2022

doi: <https://doi.org/10.26439/ciis2022.6070>

**RESUMEN.** La visión por computador es un área de estudio en la inteligencia artificial que se enfoca en el desarrollo de técnicas computacionales para percibir el mundo a través de entradas visuales, como videos o imágenes. El aprendizaje profundo ha demostrado ser una técnica eficiente para el análisis e interpretación de datos visuales. Sin embargo, afronta innumerables desafíos según su aplicación en las diferentes tareas de la visión por computador. Este panel reúne un grupo de expertos en aprendizaje profundo, quienes ofrecerán información sobre su aplicación y los desafíos en sus respectivas áreas de investigación con relación a la visión por computador.

**PALABRAS CLAVE:** visión por computador, aprendizaje profundo

## CHALLENGES OF DEEP LEARNING IN COMPUTER VISION

**ABSTRACT.** Computer vision is a field of study within artificial intelligence that focuses on developing computational techniques to perceive the world through visual data, such as video or images. Deep learning has proven to be efficient in visual data analysis and interpretation. Nevertheless, it faces countless challenges, given its application in several computer vision tasks. This panel brings together deep learning experts, who will share information about deep learning applications and challenges to overcome in their research fields regarding computer vision.

**KEYWORDS:** computer vision, deep learning

## 1. INTRODUCCIÓN

La visión por computador es una subárea de la inteligencia artificial que se enfoca en el desarrollo de técnicas computacionales para el análisis e interpretación de datos visuales, como videos o imágenes (Prince, 2012). Se ha convertido en una tecnología fundamental para muchos campos de la industria, como la seguridad, el cuidado de la salud, la agricultura, el entretenimiento, así como la industria textil y automotriz.

El aprendizaje profundo es un conjunto de técnicas basadas en redes neuronales (RN) que han contribuido al desarrollo de la visión por computador. Desde el sorprendente rendimiento alcanzado por las redes neuronales convolucionales (CNN) en la competencia ImageNet<sup>1</sup>, estas se han convertido en los modelos de aprendizaje profundo de referencia para la clasificación de imágenes (Yu et al., 2022; Krizhevsky et al., 2012), detección de objetos (Long et al., 2020), segmentación semántica (Zoph et al., 2020) y estimación de la postura humana (Cao et al., 2021). No obstante la popularidad de las CNN, otros modelos de aprendizaje profundo, como los *autoencoders* (AE), las redes neuronales recurrentes (RNN) y las redes generativas adversariales (GAN), han demostrado ser eficientes en la realización de diferentes tareas de la visión por computador. Por ejemplo, las RNN se han aplicado en el reconocimiento de acciones (Singh et al., 2017) y en el de escritura (Carbune et al., 2020); los AE se han aprovechado de forma eficiente para eliminar el ruido de imágenes (Bajaj et al., 2020) y realizar búsquedas en la web con base en imágenes; las GAN se han aplicado en la generación de imágenes realistas a partir de texto y bocetos (Lu et al., 2018; Reed et al., 2016), así como en la generación de vistas frontales de rostros para sistemas de reconocimiento facial.

A pesar de los grandes avances que ha experimentado el aprendizaje profundo en la última década, y los impresionantes resultados alcanzados en la realización de tareas de visión por computador, los diferentes modelos de redes neuronales que componen la tecnología del aprendizaje profundo aún enfrentan desafíos que necesitan ser atendidos.

## 2. PRESENTACIÓN

Este panel reúne un grupo de tres expertos en aprendizaje profundo y visión por computador con reconocida trayectoria en investigación, proyectos de desarrollo e innovación, quienes ofrecerán sus perspectivas sobre los avances en este campo y los desafíos que esta tecnología enfrenta en sus áreas de investigación relacionadas con la visión por computador. En primer lugar, se presentará a los panelistas y se facilitará las preguntas y respuestas del público. Luego se hará un resumen de la sesión, así como una breve introducción al aprendizaje profundo y sus aplicaciones en la teledetección; se comentarán los desafíos que esta tecnología afronta

---

1 Imagenet: base de datos de imágenes. Disponible en <https://image-net.org/>

en el análisis de imágenes satelitales para el análisis de cultivos. Posteriormente, la discusión se enfocará en el uso del aprendizaje profundo para la generación de muestras, con especial interés en la generación de datos sintéticos de facies geológicas para el estudio de modelos de reservorios de petróleo. Finalmente, se abordarán los desafíos del aprendizaje profundo en la detección de deforestación en regiones amazónicas a partir del análisis de imágenes de satélite con dicha tecnología.

### 3. CONCLUSIONES

Este foro reunió a un panel conformado por tres expertos en visión por computador y aprendizaje profundo, quienes comentaron sobre los desafíos a los que se enfrentan las técnicas de aprendizaje profundo en el desarrollo de aplicaciones de visión por computador. El panel coincidió en que los mayores desafíos están relacionados con la disponibilidad de bases de datos con información suficiente para la generación de arquitecturas de aprendizaje profundo, así como con la fiabilidad de datos obtenidos a partir de técnicas de generación de muestras sintéticas.

### REFERENCIAS

- Bajaj, K., Singh, D. K., & Ansari, M. A. (2020). Autoencoders based deep learner for image denoising. *Procedia Computer Science*, 171, 1535-1541. <https://doi.org/10.1016/j.procs.2020.04.164>
- Cao, Z., Hidalgo, G., Simon, T., Wei, S., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(01), 172-286. <https://doi.org/10.1109/TPAMI.2019.2929257>
- Carbune, V., Gonnet, P., Deselaers, T., Rowley, H. A., Daryin, A., Calvo, M., Wang, L.-L., Keysers, D., Feuz, S., & Gervais, P. (2020). Fast multi-language LSTM-based online handwriting recognition. *International Journal on Document Analysis and Recognition*, 23, 89-102. <https://doi.org/10.1007/s10032-020-00350-4>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing*, 25, 1-9.
- Long, X., Deng, K., Wang, G., Zhang, Y., Dang, Q., Gao, Y., Shen, H., Ren, J., Han, S., Ding, E., & Wen, S. (2020). PP-YOLO: An effective and efficient implementation of object detector. *ArXiv e-prints*. <https://doi.org/10.48550/arXiv.2007.12099>
- Lu, Y., Wu, S., Tai, Y. W., & Tang, C. K. (2018). Image generation from sketch constraint using contextual GAN. En V. Ferrari, M. Hebert, C. Sminchisescu & Y. Weiss (Eds.),

- Computer Vision – ECCV 2018. ECCV 2018. Lecture notes in computer science* (vol. 11220, pp. 213-228). [https://doi.org/10.1007/978-3-030-01270-0\\_13](https://doi.org/10.1007/978-3-030-01270-0_13)
- Prince, S. (2012). *Computer vision: Models, learning, and inference*. Cambridge University Press.
- Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. En *Proceedings of the 33rd International Conference on Machine Learning*, 48, 1060-1069.
- Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M., & Holzinger, A. (2017). Human activity recognition using recurrent neural networks. En A. Holzinger, P. Kieseberg, A. Tjoa & E. Weippl (Eds.), *Machine Learning and Knowledge Extraction. CD-MAKE 2017. Lecture notes in computer science* (vol. 10410, pp. 267-274). [https://doi.org/10.1007/978-3-319-66808-6\\_18](https://doi.org/10.1007/978-3-319-66808-6_18)
- Yu, J., Wang, Z., Vasudevan, V., Yeung, L., Seyedhosseini, M., & Wu, Y. (2022). CoCa: Contrastive captioners are image-text foundation models. *ArXiv e-prints*. <https://doi.org/10.48550/arXiv.2205.01917>
- Zoph, B., Ghiasi, G., Lin, TY., Cui, Y., Liu, H., Cubuk, E. D., & Le, Q. (2020). Rethinking pre-training and self-training. *Advances in Neural Information Processing*, 33, 1-13.



# Gamit!: plataforma para la gamificación basada en un sistema de recompensas para estudiantes universitarios

Hernán Alejandro Quintana Cruz (moderador)

[hquintan@ulima.edu.pe](mailto:hquintan@ulima.edu.pe)

<https://orcid.org/0000-0002-7037-4302>

Universidad de Lima, Perú

Elvira Rincón-Flores

[elvira.rincon@tec.mx](mailto:elvira.rincon@tec.mx)

<https://orcid.org/0000-0001-5957-2335>

Tecnológico de Monterrey, México

Brenda Santos

[brendasantos@tec.mx](mailto:brendasantos@tec.mx)

<https://orcid.org/0000-0002-2370-0777>

Tecnológico de Monterrey

Centro de Vinculación y Desarrollo Profesional,  
Campus Monterrey, México

Alberto Matsuura

[amatsuur@ulima.edu.pe](mailto:amatsuur@ulima.edu.pe)

<https://orcid.org/0000-0003-1217-4166>

Oficina de Innovación y Calidad Educativa  
Universidad de Lima, Perú

Recibido: 8 de septiembre del 2022 / Recibido: 10 de octubre del 2022

doi: <https://doi.org/10.26439/ciis2022.6068>

**RESUMEN.** La gamificación es una técnica que permite utilizar en distintos contextos las que se presentan en los juegos. En la actualidad, existen varias iniciativas para aplicarla en contextos educativos, las cuales buscan impactar en la experiencia de los estudiantes mediante la mejora en la motivación (Surendeleg et al., 2014; Hamari et al., 2015; Rincón-Flores et al., 2018) y la atención de ellos (Hamari & Keronen, 2017). Para una correcta aplicación de la gamificación,

es necesario basarse en alguna metodología que ordene y facilite su adopción, así como una herramienta que dinamice su utilización por parte de los docentes y los estudiantes. Gamit! es una plataforma tecnológica de gamificación cuyo valor radica en facilitar la adopción de esta técnica mediante una metodología basada en narrativas.

PALABRAS CLAVE: gamificación, educación superior, innovación educativa

### GAMIT!: A GAMIFICATION PLATFORM FOR UNIVERSITY STUDENTS BASED ON A REWARD SYSTEM

ABSTRACT. Gamification is a technique that uses game concepts in domains different from games. There are many gamification initiatives in education because of its impact on student experience, especially in motivation (Surendelegh et al., 2014; Hamari et al., 2015; Rincón-Flores et al., 2018) and focus (Hamari & Keronen, 2017). A methodology for easy game-concept adoption and tools to help teachers and students in gamification activities are necessary for proper gamification. Gamit! uses a narrative-based methodology to make gamification easier to follow and implement in classes.

KEYWORDS: gamification, higher education, educational innovation

## 1. INTRODUCCIÓN

Algunos de los retos que se presentaron en la implementación de la gamificación, antes del uso de la plataforma Gamit!, se centran en la administración de la dinámica, desde la narrativa hasta la socialización de las insignias que se otorgaron. La gestión del tiempo y la información fue uno de los retos mayores y que mermó el resultado inicial de la gamificación al convertirse en una carga adicional para el docente. Por otro lado, el tablero de liderazgo se visualizaba como plano, sin diseño y con un acomodo poco atractivo para los alumnos. A su vez, la socialización del tablero tuvo diferentes áreas de oportunidad en la elección de la mejor manera de presentarlo en cada plataforma o en cada ocasión que el docente compartía la información, o que los alumnos consultaron las insignias obtenidas.

## 2. PRESENTACIÓN

La discusión realizada por el panel se centró en la utilización de la gamificación como herramienta para mejorar la motivación por parte de los estudiantes en instituciones de educación superior. Los resultados mostrados se basaron en los de la implementación y utilización de la plataforma web llamada Gamit! (Rincón-Flores et al., 2022) en cursos de educación superior en la Universidad de Lima (Perú), así como en el Tecnológico de Monterrey (México).

Sobre la base de la experiencia realizada por los expositores, se analizaron los puntos a favor y los problemas de la aplicación de gamificación. Además, se planteó una guía para la implementación de una metodología de gamificación basada en narrativas.

## 3. CONCLUSIONES

La gamificación ofrece beneficios pedagógicos que modifican el ambiente de aprendizaje en cuanto a la relación que se establece en diferentes direcciones: alumno-contenido, alumno-profesor, alumno-alumno. Sin embargo, para la obtención de dichos beneficios, se requiere un diseño apropiado y contextualizado para cada asignatura.

El uso de la tecnología, como lo es la plataforma Gamit!, facilita la implementación de la gamificación en un ambiente digital seguro y más atractivo, que, además, contribuye a la administración de la técnica para que el usuario (docente y alumnos) se centren en el aprendizaje y en el aprovechamiento de los recursos educativos.

## REFERENCIAS

- Hamari, J., Keronen, L., & Alha, K. (2015). Why do people play games? A review of studies on adoption and use. En *2015 48th Hawaii International Conference on System Sciences* (pp. 3559-3568). <https://doi.org/10.1109/HICSS.2015.428>
- Rincón-Flores, E. G., Gallardo, K., & De la Fuente, J. M. (2018). Strengthening an educational innovation strategy: Processes to improve gamification in calculus course through performance assessment and meta-evaluation. *International Electronic Journal of Mathematics Education*, *13*(1), 1-11. <https://doi.org/10.12973/iejme/2692>
- Rincón-Flores, E. G., Rodríguez-Rodríguez, N., Santos-Guevara, B. N., Matsuura Sonoda, A., & Quintana-Cruz, H. (2022). Gamit! Interactive platform for gamification. En *2022 IEEE Global Engineering Education Conference (EDUCON)* (pp. 10-13). <https://doi.org/10.1109/EDUCON52537.2022.9766380>
- Surendele, G., Murwa, V., Yun, H.-K., & Kim, Y. S. (2014). The role of gamification in education – A literature review. *Contemporary Engineering Sciences*, *7*(29), 1609-1616. <http://dx.doi.org/10.12988/ces.2014.411217>

# PONENCIAS

---



# Análisis global de la producción científica respecto del aprendizaje organizacional en la educación superior\*

Liliana Pedraja-Rejas  
lpedraja.rejas@gmail.com  
<https://orcid.org/0000-0001-7732-4183>  
Universidad de Tarapacá, Chile

Emilio Rodríguez-Ponce  
erodriguez@uta.cl  
<https://orcid.org/0000-0003-4861-002X>  
Universidad de Tarapacá, Chile

Camila Muñoz-Fritis  
camila.munoz.fritis@gmail.com  
<https://orcid.org/0000-0002-1247-9456>  
Universidad de Tarapacá, Chile

Recibido: 8 de septiembre del 2022 / Aceptado: 9 de octubre del 2022  
<https://doi.org/10.26439/ciis2022.6071>

RESUMEN. El objetivo de este artículo es estudiar, mediante un análisis bibliométrico y de contenido, la literatura disponible en Web of Science sobre aprendizaje organizacional en el contexto de la educación superior. Los resultados revelan que, desde el 2012, se observa una mayor producción de conocimiento en el campo; las publicaciones pertenecen en su mayoría a autores estadounidenses; hay un bajo nivel de colaboración científica internacional; gran parte de las publicaciones se realizan en revistas de educación; y existen dos corrientes principales de análisis: la primera discute el aprendizaje organizacional como foco central para el proceso de adaptación, mientras que la segunda se concentra en buscar relaciones entre este y otros

---

\* Los autores agradecen a la Agencia Nacional de Investigación y Desarrollo (ANID) por el apoyo recibido a través del proyecto Fondecyt 1220568, Chile.

resultados deseables. Se discuten los hallazgos y se presentan las implicaciones para el desarrollo del aprendizaje organizacional.

PALABRAS CLAVE: aprendizaje institucional, educación superior, análisis bibliométrico, análisis de contenido, análisis bibliográfico

## GLOBAL ANALYSIS OF SCIENTIFIC PRODUCTION REGARDING ORGANIZATIONAL LEARNING IN HIGHER EDUCATION

ABSTRACT. The aim of this article is to study through a bibliometric and content analysis, the literature available in Web of Science on organizational learning in the context of higher education. The results reveal that, since 2012, there is an increased production of knowledge in the field; publications belong mainly to US authors; a low level of international scientific collaboration is observed; a large part of the publications are in education journals; and there are two main streams of analysis, the first one discusses organizational learning as a central focus for the adaptation process, while the second one concentrates on looking for relationships between this and other desirable outcomes. The findings are discussed and implications for the development of organizational learning are presented.

KEYWORDS: institutional learning, higher education, bibliometric analysis, content analysis, bibliographic analysis

## 1. INTRODUCCIÓN

El aprendizaje organizacional es un fenómeno dinámico (Souza & Takahashi, 2019) que implica múltiples escenarios de interacción y análisis en los ámbitos individual, grupal, organizacional e interorganizacional (Mousa et al., 2020), y que depende del contexto en el que las organizaciones avanzan hacia el logro de los resultados deseados (Lyman et al., 2018). Este es un proceso que involucra el manejo y gestión de la información para reflejar nuevos conocimientos e ideas, y que puede ser utilizado como un marco para comprender cómo las universidades desarrollan estrategias, toman decisiones y fomentan el desempeño de roles adicionales entre sus colaboradores. El aprendizaje y el conocimiento están entrelazados, ya que el primero es el contenido creado como parte del proceso de aprendizaje, mientras que este último es el principal medio de renovación estratégica (Souza & Takahashi, 2019).

En el ámbito educativo, el aprendizaje organizacional se define como la “capacidad de la unidad académica, más allá de cualquier individuo específico, para aprender de los errores y escuchar las demandas del entorno, desencadenando la adaptación, el cambio organizacional, y generando consecuencias en términos de gestión” (Busco et al., 2018, p. 438), innovación (Aminbeidokhti et al., 2016; Qian & Huang, 2017) y desempeño (Hussein et al., 2016). Este tipo de aprendizaje también puede contribuir a crear y utilizar nuevos conocimientos sobre distintas prácticas para ayudar a los estudiantes a tener éxito académico, y luego compartir ese conocimiento en toda la organización (Mokher, Park-Gaghan et al., 2020).

Diversos autores se han propuesto determinar los factores que deben existir para respaldar el aprendizaje organizacional, y han encontrado que los sistemas integrados y conectados de información, las prácticas de aprendizaje en equipo, la cooperación interdisciplinaria, el liderazgo, la cultura y la estructura universitaria tienen el potencial de aportar al desarrollo de dicho aprendizaje (Hussein et al., 2016; Kezar & Holcombe, 2020; Voolaid & Ehrlich, 2017).

A pesar de que se reconoce que las universidades deben convertirse en organizaciones de aprendizaje para responder a los desafíos y presiones externos, lo cierto es que las investigaciones en torno a la temática son más bien escasas (Voolaid & Ehrlich, 2017). En este sentido, el presente trabajo busca analizar —mediante un análisis bibliométrico y de contenido— la literatura disponible en Web of Science sobre aprendizaje organizacional en el contexto de la educación superior, para identificar así tendencias y brechas en el conocimiento existente a fin de contribuir a la consolidación del campo de estudio.

## 2. METODOLOGÍA

Se lleva a cabo un estudio de naturaleza cuantitativa, frecuentemente efectuado en el campo de la ingeniería, referido a un análisis bibliométrico. La producción científica analizada fue recolectada desde la base de datos Web of Science Core Collection (WoS) en mayo del 2022.

Se utilizaron como palabras clave los términos *organizational learning or institutional learning* y *higher education or university or universities* como tema. Se filtró la información según el tipo de documento, seleccionando exclusivamente los artículos, y según las categorías de WoS, donde se eligieron “Education-Educational Research”, “Education Scientific Disciplines” y “Psychology Educational”.

WoS indexa las revistas donde se publican los trabajos científicos de mayor relevancia e influencia en la mayoría de los campos disciplinarios (Guzmán-Valenzuela et al., 2020). Esta base de datos internacional fue una de las primeras y es la más completa que ha salido; goza hasta la actualidad de un gran prestigio y es utilizada para múltiples propósitos (Pranckuté, 2021).

## 2.1 Procesamiento de los datos

Para cumplir con el objetivo propuesto, se examinaron los 105 artículos disponibles hasta la fecha de extracción de los datos, y se llevaron a cabo dos tipos de análisis: uno bibliométrico y otro de contenido.

La bibliometría comprende la aplicación de técnicas cuantitativas sobre datos bibliométricos (por ejemplo, unidades de publicación y cita) de los documentos existentes de un campo de investigación (Donthu et al., 2021). Los académicos han planteado que este es un método interdisciplinario que permite el mapeo efectivo de las direcciones y los temas abordados durante el desarrollo de un área específica de conocimiento (Tandon et al., 2021). Si bien existen diferentes métodos de análisis (Donthu et al., 2021), este trabajo se concentra en la producción por año, en los países y revistas más productivos y/o influyentes de la muestra, y en las palabras clave más utilizadas. Para el procesamiento de la información, se usó el *software* Bibliometrix, un programa orientado a la cienciometría y bibliometría (Pedraja-Rejas et al., 2021) que proporciona un conjunto de herramientas que apoya el flujo de trabajo de estos estudios (recopilación, análisis y visualización de datos) (Aria & Cuccurullo, 2017). También se empleó VOSviewer (Van Eck & Waltman, 2010) para generar el mapa de coocurrencia de palabras clave, que sirve para identificar relaciones en los usos de estos conceptos.

El análisis de contenido, por su parte, permite identificar temas relevantes de los artículos analizados según las preguntas de investigación formuladas (Guzmán-Valenzuela et al., 2020). Para hacerlo, se revisó en detalle la totalidad de las publicaciones de la muestra, de manera se puedan identificar metodologías de trabajo y temáticas principales abordadas.

## 3. RESULTADOS

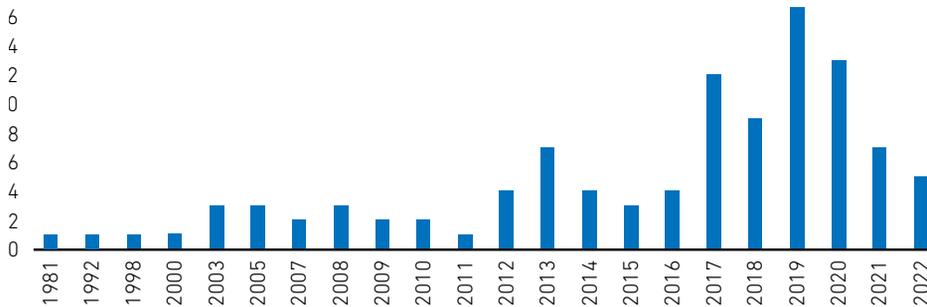
El análisis bibliométrico llevado a cabo requiere ilustrar los hallazgos desde una mirada deductiva. En efecto, la Figura 1 permite observar que, en general, existe una tendencia al alza en el

número de artículos publicados, aunque con ritmos diferentes si se analizan por etapas. Por ejemplo, hasta el 2000, se puede distinguir una producción más bien marginal (4 artículos; 3,8 % del total), luego se identifica una etapa de crecimiento moderado (16 artículos; 15,2 % del total) durante el periodo 2001-2011; mientras que, si se analiza la producción desde el 2012 en adelante, se distingue una etapa de mayor crecimiento que concentra casi el 81,0 % de la producción total (85 artículos).

En cuanto al idioma en el que están escritos los trabajos, se tiene que 102 se encuentran en inglés (97,1 %), y tres en ruso, español y turco, respectivamente.

**Figura 1**

*Número de publicaciones por año*

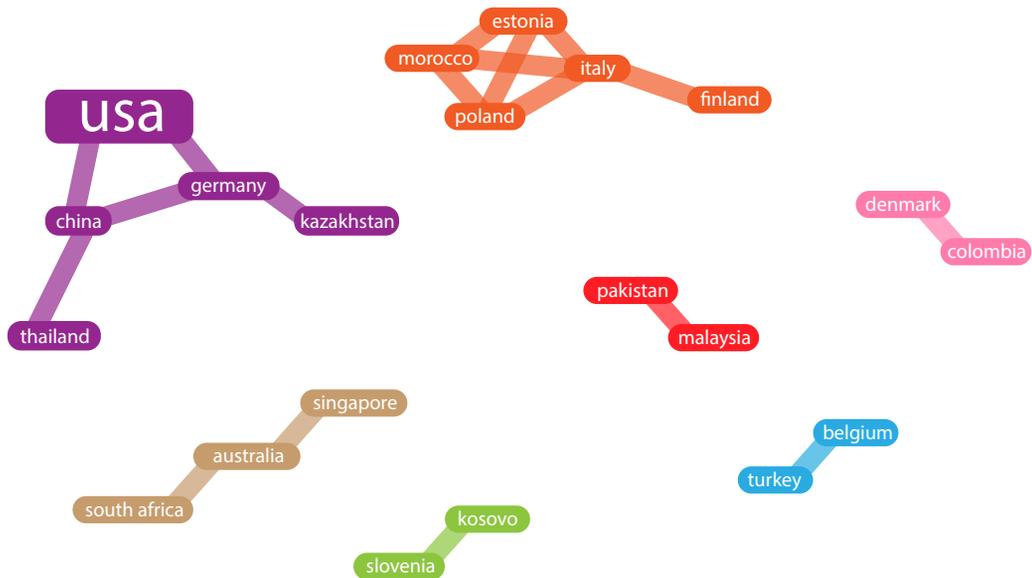


En cuanto a los países con producción en el campo, cuando solo se considera la nacionalidad del autor de correspondencia, se observa que Estados Unidos publicó el mayor número de documentos, seguido por Reino Unido y Australia. En relación con la publicación simple del país (SCP, por sus siglas en inglés) y la publicación con cooperación entre países (MCP, por sus siglas en inglés), se puede decir que, en general, pocos artículos (11; 10,5 %) se escribieron con colaboración internacional. Otro punto importante es que existe una baja representación latinoamericana, sobre todo en el ámbito de la colaboración científica, ya que solo Colombia registra cooperación con un autor extranjero (danés).

La Figura 2 detalla la red de colaboración entre países considerando las nacionalidades de los autores y coautores de los artículos. En ella también se evidencia la escasa interacción entre países.

Figura 2

Red de colaboración entre países



Nota. Generado con la herramienta Bibliometrix.

La ley de Bradford, descrita en 1934, se basa en la premisa de que las revistas que publican sobre un campo de estudio se pueden organizar en zonas según su relevancia, al ser mayormente citadas (Desai et al., 2018). De esta forma, se tiene que las primeras 11 revistas mencionadas en la Tabla 1 constituyen la zona 1 y son las principales fuentes de divulgación de conocimiento en el tema (35 artículos; 33,3 %). La zona 2 incluye 26 revistas que aportan con 36 trabajos (34,3 %), mientras que la zona 3 está conformada por 34 revistas que contribuyen con 34 artículos en total (32,4 %).

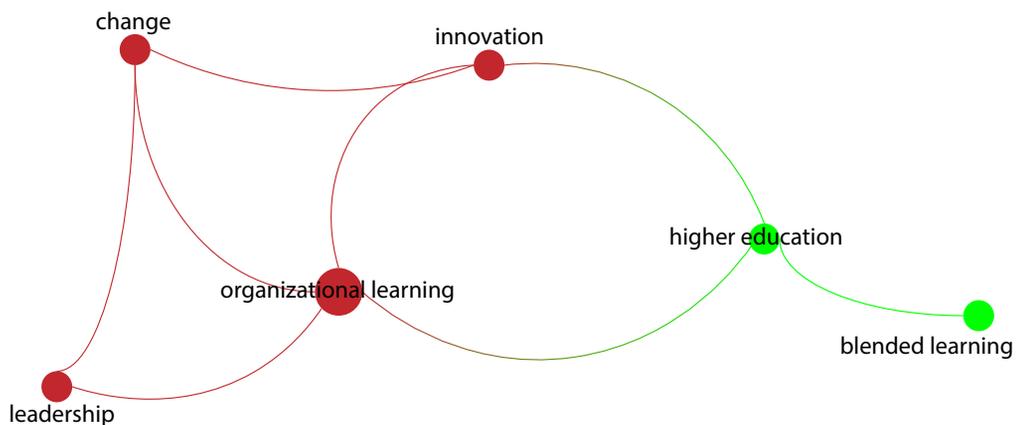
De manera global, se tiene que el 87,3 % de las publicaciones se hallan en revistas de educación, de las cuales el 19,7 % son revistas especializadas en educación superior. Los demás artículos fueron publicados en revistas de salud (9,9 %) y de otra índole (2,8 %). En cuanto a la indexación principal de las revistas, se aprecia que el 56,3 % se halla en la colección de Social Sciences Citation Index (SSCI), el 8,5 % en Science Citation Index Expanded (SCI-E) y el 35,2 % en Emerging Sources Citation Index (ESCI). SSCI y SCI-E son bases de datos de WoS desde 1975, las cuales publican investigaciones de gran relevancia en la mayoría de los campos disciplinarios, mientras que ESCI se lanzó en el 2015 y contiene publicaciones de importancia regional en campos científicos más bien emergentes (Guzmán-Valenzuela et al., 2020).

**Tabla 1***Revistas más productivas en el campo*

N.º	Revista	Colección	Total de artículos	SJR (2021)	Zona
1	Higher Education	SSCI	8	1,73	Zona 1
2	Eurasia Journal of Mathematics, Science and Technology Education	SSCI	4	0,57	Zona 1
3	Academy of Management Learning & Education	SSCI	3	1,76	Zona 1
4	AERA Open	SSCI	3	0,86	Zona 1
5	International Review of Research in Open and Distributed Learning	SSCI	3	1,32	Zona 1
6	Nurse Education Today	SSCI	3	0,99	Zona 1
7	Studies in Higher Education	SSCI	3	1,57	Zona 1
8	Academic Medicine	SCI-E	2	1,66	Zona 1
9	Educational Evaluation and Policy Analysis	SSCI	2	2,45	Zona 1
10	Eurasian Journal of Educational Research	ESCI	2	0,28	Zona 1
11	Hacettepe University Journal of Education	ESCI	2	0,19	Zona 1
12	Higher Education Research & Development	SSCI	2	1,44	Zona 2
13	Higher Education, Skills and Work-based Learning	ESCI	2	0,33	Zona 2
14	Interactive Learning Environments	SSCI	2	1,17	Zona 2
15	International Journal of Emerging Technologies in Learning	ESCI	2	0,63	Zona 2
16	International Journal of Sustainability in Higher Education	SSCI	2	0,86	Zona 2
17	Journal of Educational Change	SSCI	2	1,52	Zona 2
18	Journal of Studies in International Education	SSCI	2	1,13	Zona 2
19	Quality Assurance in Education	ESCI	2	0,32	Zona 2
20	Teaching and Teacher Education	SSCI	2	1,95	Zona 2
21	Teaching in Higher Education	SSCI	2	0,97	Zona 2

Al analizar las palabras clave, sin contar los términos de búsqueda, las más utilizadas por los autores en sus trabajos son *innovación* (7 veces), *liderazgo* (6), *cambio* (4), *aprendizaje* (4) y *aprendizaje semipresencial* (4). Dado que ninguna tiene una representación superior al 10 % en los artículos de la muestra, se puede afirmar, en una primera aproximación, que los subtemas abordados son más bien variados. La Figura 3 muestra la coocurrencia de las palabras clave que aparecen cuatro o más veces. Se identificaron dos conglomerados principales; el primero (color rojo) aborda los conceptos de cambio, innovación y liderazgo, los cuales son más bien relacionados con aspectos institucionales, mientras que el segundo (color verde) abarca el aprendizaje semipresencial, que está más centrado en los estudiantes y en los procesos de enseñanza-aprendizaje.

**Figura 3**  
*Coocurrencia de palabras clave*



*Nota.* Generado con la herramienta VOSviewer.

En cuanto a las metodologías utilizadas, y tal como se puede observar en la Tabla 2, los métodos de investigación cualitativos son mayormente empleados en estudios relacionados con la temática (46,7 %), seguidos por los cuantitativos (28,6 %). Más atrás se encuentran los artículos abordados a partir de técnicas mixtas (19,0 %) y discusiones teóricas o revisiones de literatura (5,7 %).

Tabla 2

*Metodologías de investigación utilizadas*

Método	Técnicas	Total de artículos
Cualitativo	Estudio de casos, entrevistas, análisis documental, etnográfico, etcétera	49
Cuantitativo	Estadística descriptiva, modelos de regresión, etcétera	30
Mixto	Técnicas cuantitativas y cualitativas	20
Revisión de literatura	Discusión teórica/conceptual	6

Respecto a las perspectivas abordadas en los trabajos de la muestra, se hallan dos corrientes principales de análisis. La primera trata sobre el potencial del aprendizaje organizacional para facilitar las respuestas a los entornos cambiantes de la educación superior, mientras que la segunda se concentra en encontrar relaciones (estadísticas y desde la perspectiva de los modelos de regresión) del aprendizaje organizacional con otras variables deseables para las instituciones educativas. A continuación, se procede a describir cada una de ellas.

### 3.1 Adaptabilidad para enfrentar los cambios

Las universidades están en un proceso de cambios: la modificación de la legislación y las políticas en el sector promueven cada vez más el espíritu empresarial, la internacionalización y la competencia entre instituciones por el financiamiento y por los investigadores y estudiantes de excelencia (Schnurbus & Edvardsson, 2022). En este marco, los efectos de las reformas han sido abordados en múltiples trabajos y diversos contextos, como el europeo (Capano & Pritoni, 2020; Kaçaniku, 2020; O Shea & O Hara, 2020), el americano (Bernasconi & Celis, 2017; Mokher, Spencer et al., 2020) y el asiático (Harun et al., 2020; Liu, 2022).

Ante los cambios gubernamentales, los trabajos de la muestra plantean que el aprendizaje institucional se considera un punto fuerte del proceso de adaptación (O Shea & O Hara, 2020), ya que, como explican Mokher, Park-Gaghan et al. (2020), a través de este proceso los miembros de la organización crean mecanismos para probar nuevos enfoques en la resolución de problemas, aprenden de los errores y luego modifican aún más las prácticas en aras de mejorarlas.

Sin embargo, Graham y Donaldson (2020) advierten que las lógicas e identidades institucionales condicionan el pensamiento y el aprendizaje dentro de las organizaciones, y es por eso que algunos enfatizan en la necesidad de establecer ciertas prácticas (como el desarrollo

de culturas de aprendizaje y de autoevaluación) que permitan consolidar una visión, mejorar el carácter y la calidad de los miembros, aprender de los errores y escuchar las demandas del entorno, y mantener así una perspectiva a largo plazo (Busco et al., 2018; Lyman et al., 2018).

Si bien el aprendizaje organizacional es reconocido como una herramienta importante para facilitar el cambio, este no está exento de obstáculos (Kezar & Holcombe, 2020), y es aquí donde radica el gran desafío por cultivarlo en las instituciones de manera que puedan sortear las transformaciones pasadas y venideras en un campo tan volátil y exigente como es la educación.

### 3.2 Aprendizaje organizacional y sus efectos desde la perspectiva estadística

Se encontró un grupo de artículos que trataba de identificar la relación del aprendizaje organizacional con otras variables a través de modelos de regresión. La Tabla 3 resume los principales resultados obtenidos en estas investigaciones. Ninguno de estos estudios encontró efectos adversos del aprendizaje organizacional, sino más bien que este favorecía variables tan significativas como la preparación de los docentes, la innovación y el desempeño institucional.

**Tabla 3**

*Resumen de artículos*

Autor(es) (año)	Principales resultados
Ansmann y Seyfried (2022)	El isomorfismo mimético es compatible con los procesos de aprendizaje organizacional y, por lo tanto, no compromete el desarrollo organizacional.
Zhou y Tu (2021)	La cultura de aprendizaje organizacional influye positivamente en el empoderamiento psicológico y en el aprendizaje permanente de los docentes universitarios desde el desarrollo profesional.
Ishak y Mansor (2020)	El aprendizaje organizacional influye en la preparación del personal académico para la educación 4.0.
Mousa et al. (2020)	El aprendizaje organizacional muestra una relación positiva con el nivel de resiliencia organizacional de los académicos.
Tabaghdehi et al. (2018)	El aprendizaje organizacional tiene un efecto directo sobre el capital intelectual y sobre el emprendimiento académico.
Qian y Huang (2017)	Una mayor capacidad de aprendizaje organizacional podría acumular mejor el capital humano en la industria médica y favorecer la capacidad de innovación.
Aminbeidokhti et al. (2016)	TQM afecta positiva y significativamente al aprendizaje organizacional, y este último tiene un efecto significativo en la innovación organizacional.
Abbasi y Zamani-Miandashti (2013)	Relación positiva y significativa entre el liderazgo transformacional y la cultura organizacional con el aprendizaje organizacional, e influencia positiva del aprendizaje en el desempeño.

*(continúa)*

(continuación)

Cheng et al. (2012)	Influencia positiva de los factores del entorno de aprendizaje organizacional con la motivación de los empleados para usar un sistema de aprendizaje electrónico en el lugar de trabajo.
Mebane y Galassi (2003)	Una cantidad significativa de variación en el aprendizaje organizacional percibido fue explicada por las variables de aprendizaje de grupo, tarea y equipo en una asociación de escuela de desarrollo profesional.

#### 4. DISCUSIÓN DE LOS RESULTADOS

En este estudio, se ha caracterizado la producción de conocimiento sobre el aprendizaje organizacional en el campo de la educación superior. Los resultados revelaron que el número de artículos en WoS ha aumentado con el tiempo, sobre todo si se considera desde el 2012 hasta la actualidad. El mayor interés de los académicos y profesionales por este tema puede ser explicado porque, en la era actual del conocimiento, el aprendizaje organizacional es considerado clave para añadir valor a las instituciones al facilitar el logro de sus objetivos. Aun así, pese al incremento en la producción, esta es todavía marginal si se compara con otros tópicos en la educación superior.

El idioma inglés es ampliamente dominante. Esto se debe a la existencia de una sobrerrepresentación en WoS de naciones y revistas anglosajonas (Vera-Baceta et al., 2019), la mayoría de las cuales están gestionadas por compañías de países ricos del Norte Global (Guzmán-Valenzuela et al., 2020).

De las 71 revistas que publicaron sobre el tema, *Higher Education* sobresale como la más productiva con 8 artículos. Es una revista internacional, prestigiosa e influyente, que destaca por su contribución al campo (Tight, 2018) al publicar artículos que evalúan el desarrollo educativo en todo tipo de instituciones de educación terciaria y en los sectores público y privado. En cuanto a la totalidad de revistas de la muestra, se encontró que el 87,3 % de las publicaciones se realizan en revistas de educación, de las cuales una fracción corresponde a revistas especializadas en educación superior (19,7 %). El resto se publica en revistas de áreas disciplinarias asociadas a la salud (9,9 %) y al ámbito social (2,8 %).

Si bien se evidencia un crecimiento en la producción científica a lo largo de los años, hay que ser cuidadosos, ya que la mayoría de los aportes provienen de autores estadounidenses (33,3 %). Este evidente estado de concentración científica genera oportunidades para los investigadores de otras latitudes, quienes pueden contribuir al estado del arte desde sus propios contextos culturales, sociales y tecnológicos, que muchas veces difieren de la gran nación americana. Este es el caso de los países de Latinoamérica, que tienen una representación más bien marginal en la muestra.

La colaboración científica entre países es muy limitada, pues apenas el 10,5 % de los artículos fueron escritos con autores de otras nacionalidades. Sin embargo, en la literatura se sostiene que la cooperación permite acceder a más puntos de vista y lleva a resultados más interesantes (Lyu et al., 2022). Por eso, se recomienda que los investigadores fortalezcan sus redes académicas, desarrollen oportunidades de colaboración internacional y participen en proyectos con investigadores influyentes en la disciplina (Cancino et al., 2020).

Al analizar las metodologías que utilizaron los autores en sus trabajos, se descubrió que casi un 47,0 % prefería las técnicas cualitativas. Esto puede ser explicado porque este campo de estudio es aún emergente, por lo que, al utilizar este método, se busca explorar y formular nuevas teorías y modelos para comprender mejor esta temática (Pedraja-Rejas et al., 2022).

En lo que respecta a los principales ejes temáticos, se hallan dos corrientes principales de análisis. La primera discute el aprendizaje organizacional como foco central para el proceso de adaptación, mientras que la segunda se concentra en buscar relaciones entre este y otros resultados deseables. En cuanto a lo primero, los autores piensan que no es de extrañar, ya que el aprendizaje organizacional es frecuentemente relacionado con la adaptabilidad (Kezar & Holcombe, 2020; O Shea & O Hara, 2020), lo cual resulta imperativo en la educación superior, donde la globalización, la expansión de matrículas y la cada vez mayor adopción de prácticas comerciales del sector privado (Bernasconi & Celis, 2017; Mahony & Weiner, 2019) traen constantes presiones a las instituciones educativas por adaptar sus estructuras y modelos de gestión. En torno a lo segundo, los autores coinciden en que este tipo de investigaciones fortalece el discurso de la necesidad de mejorar el aprendizaje organizacional en la institución educativa de educación superior, porque demuestra a través de datos empíricos los beneficios que traería el cultivo de esta capacidad en todos los niveles organizacionales.

## 5. CONCLUSIONES

La principal conclusión de este trabajo es que el aprendizaje organizacional es estudiado como un foco central en el proceso de adaptación de las instituciones. La revisión bibliométrica efectuada permite indicar, además, que falta discutir más acerca de que el conocimiento se pueda transformar de una óptica individual a una de tipo colectivo, de manera que traspase todos los niveles y ayude al cumplimiento de los objetivos. En la misma línea y a pesar de que hay algunos artículos que abordan este subtema (Callahan & Martin, 2007; Dai, 2019; Ding, 2019; Kezar & Holcombe, 2020), aún se puede profundizar más en cómo las asociaciones con entidades externas pueden contribuir al fortalecimiento de sistemas de aprendizaje internos.

Finalmente, advertimos que los resultados aquí presentados deben ser tomados con precaución, ya que el estudio solo analizó los artículos disponibles en una sola base de datos, y pueden no capturar el real estado del campo de estudio. Aun así, este trabajo presenta una primera aproximación, la cual puede ser complementada por futuras investigaciones, ya sea en

comparaciones en el ámbito latinoamericano, o bien en subáreas de la investigación, como la construcción de conocimiento colectivo.

## REFERENCIAS

- Abbasi, E., & Zamani-Miandashti, N. (2013). The role of transformational leadership, organizational culture and organizational learning in improving the performance of Iranian agricultural faculties. *Higher Education*, 66(4), 505-519. <https://doi.org/10.1007/s10734013-9618-8>
- Aminbeidokhti, A., Jamshidi, L., & Hoseini, A. M. (2016). The effect of the total quality management on organizational innovation in higher education mediated by organizational learning. *Studies in Higher Education*, 41(7), 1153-1166. <https://doi.org/10.1080/03075079.2014.966667>
- Ansmann, M., & Seyfried, M. (2022). Isomorphism and organizational performance: Evidence from quality management in higher education. *Quality Assurance in Education*, 30(1), 135-149. <https://doi.org/10.1108/QAE-07-2021-0114>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Bernasconi, A., & Celis, S. (2017). Higher education reforms: Latin America in comparative perspective. *Education Policy Analysis Archives*, 25(67), 1-15. <http://dx.doi.org/10.14507/epaa.25.3240>
- Busco, C., Dooner, C., & d'Alencon, A. (2018). Universidad de Chile: Self-assessment and its effects on university's management. *Higher Education*, 75, 431-447. <https://doi.org/10.1007/s10734-017-0148-7>
- Callahan, J. L., & Martin, D. (2007). The spectrum of school–university partnerships: A typology of organizational learning systems. *Teaching and Teacher Education*, 23(2), 136-145. <https://doi.org/10.1016/j.tate.2006.04.038>
- Cancino, C. A., Merigó, J. M., Urbano, D., & Amorós, J. E. (2020). Evolution of the entrepreneurship and innovation research in Ibero-America between 1986 and 2015. *Journal of Small Business Management*, 1-31. <https://doi.org/10.1080/00472778.2020.1776578>
- Capano, G., & Pritoni, A. (2020). What really happens in higher education governance? Trajectories of adopted policy instruments in higher education over time in 16 European countries. *Higher Education*, 80, 989-1010. <https://doi.org/10.1007/s10734-020-00529-y>

- Cheng, B., Wang, M., Moormann, J., Olaniran, B. A., & Chen, N. S. (2012). The effects of organizational learning environment factors on e-learning acceptance. *Computers & Education*, 58(3), 885-899. <https://doi.org/10.1016/j.compedu.2011.10.014>
- Dai, Y. (2019). Aligning for transformational cross-border partnership: A case of Sino-US university partnership. *Journal of Studies in International Education*, 23(3), 326-345. <https://doi.org/10.1177/1028315318797160>
- Desai, N., Veras, L., & Gosain, A. (2018). Using Bradford's law of scattering to identify the core journals of pediatric surgery. *Journal of Surgical Research*, 229, 90-95. <https://doi.org/10.1016/j.jss.2018.03.062>
- Ding, X. (2019). Marginal revolution: The impact of transnational education on higher education in host countries: A case study of China. *Higher Education Policy*, 32(3), 419-440. <https://doi.org/10.1057/s41307-018-0089-5>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Graham, S. W., & Donaldson, J. F. (2020). Academic leaders' response to the volatility of higher education: The influence of institutional logics. *Studies in Higher Education*, 45(9), 1864-1877. <https://doi.org/10.1080/03075079.2019.1586867>
- Guzmán-Valenzuela, C., Rojas-Murphy Tagle, A., & Gómez-González, C. (2020). Polifonía epistémica de la investigación sobre las experiencias estudiantiles: el caso latinoamericano. *Archivos Analíticos de Políticas Educativas*, 28(96), 1-36. <https://doi.org/10.14507/epaa.28.4919>
- Harun, H., Wardhaningtyas, S., Khan, H. Z., An, Y., & Masdar, R. (2020). Understanding the institutional challenges and impacts of higher education reforms in Indonesia. *Public Money & Management*, 40(4), 307-315. <https://doi.org/10.1080/09540962.2019.1627063>
- Hussein, N., Omar, S., Noordin, F., & Ishak, N. A. (2016). Learning organization culture, organizational performance and organizational innovativeness in a public institution of higher education in Malaysia: A preliminary study. *Procedia Economics and Finance*, 37, 512-519. [https://doi.org/10.1016/S2212-5671\(16\)30159-9](https://doi.org/10.1016/S2212-5671(16)30159-9)
- Ishak, R., & Mansor, M. (2020). The relationship between knowledge management and organizational learning with academic staff readiness for education 4.0. *Eurasian Journal of Educational Research*, 20(85), 169-184.
- Kaçaniku, F. (2020). Towards quality assurance and enhancement: The influence of the Bologna Process in Kosovo's higher education. *Quality in Higher Education*, 26(1), 32-47. <https://doi.org/10.1080/13538322.2020.1737400>

- Kezar, A. J., & Holcombe, E. M. (2020). Barriers to organizational learning in a multiinstitutional initiative. *Higher Education*, 79(6), 1119-1138. <https://doi.org/10.1007/s10734-019-00459-4>
- Liu, J. (2022). Neoliberal trends of higher education reforms in China, Japan, and Korea: Catchup and self-reorientation. En J. Zajda & W. J. Jacob (Eds.), *Discourses of globalisation and higher education reforms. Globalisation, comparative education and policy research* (pp. 133-147). Springer.
- Lyman, B., Cowan, L. A., & Hoyt, H. C. (2018). Organizational learning in a college of nursing: A learning history. *Nurse Education Today*, 61, 134-139. <https://doi.org/10.1016/j.nedt.2017.11.004>
- Lyu, P. H., Zhang, M. Z., Liu, C. J., & Ngai, E. W. (2022). Global scientific production, international cooperation and knowledge evolution of public administration. *Public Administration*, 1-29. <https://doi.org/10.1111/padm.12853>
- Mahony, P., & Weiner, G. (2019). Neo-liberalism and the state of higher education in the UK. *Journal of Further and Higher Education*, 43(4), 560-572. <https://doi.org/10.1080/0309877X.2017.1378314>
- Mebane, D. J., & Galassi, J. P. (2003). Variables affecting collaborative research and learning in a professional development school partnership. *The Journal of Educational Research*, 96(5), 259-268. <https://doi.org/10.1080/00220670309597638>
- Mokher, C. G., Park-Gaghan, T. J., Spencer, H., Hu, X., & Hu, S. (2020). Institutional transformation reflected: Engagement in sensemaking and organizational learning in Florida's developmental education reform. *Innovative Higher Education*, 45(1), 81-97. <https://doi.org/10.1007/s10755-019-09487-5>
- Mokher, C. G., Spencer, H., Park, T. J., & Hu, S. (2020). Exploring institutional change in the context of a statewide developmental education reform in Florida. *Community College Journal of Research and Practice*, 44(5), 377-390. <https://doi.org/10.1080/10668926.2019.1610672>
- Mousa, M., Abdelgaffar, H. A., Chaouali, W., & Aboramadan, M. (2020). Organizational learning, organizational resilience and the mediating role of multi-stakeholder networks: A study of Egyptian academics. *Journal of Workplace Learning*, 32(3), 161-181. <https://doi.org/10.1108/JWL-05-2019-0057>
- O Shea, S., & O Hara, J. (2020). The impact of Ireland's new higher education system performance framework on institutional planning towards the related policy objectives. *Higher Education*, 80(2), 335-351. <https://doi.org/10.1007/s10734-019-00482-5>
- Pedraja-Rejas, L., Rodríguez-Ponce, E., & Muñoz-Fritis, C. (2022). Human resource management and performance in Ibero-America: Bibliometric analysis of scientific

- production. *Management Letters / Cuadernos de Gestión*, 22(2), 123-137. <https://doi.org/10.5295/cdg.211569lp>
- Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, 9(1), 12. <https://doi.org/10.3390/publications9010012>
- Qian, Z. W., & Huang, G. (2017). Human capital and innovation ability in medical education: An empirical study. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(8), 5395-5403. <https://doi.org/10.12973/eurasia.2017.01008a>
- Schnurbus, V., & Edvardsson, I. R. (2022). The third mission among Nordic universities: A systematic literature review. *Scandinavian Journal of Educational Research*, 66(2), 238-260. <https://doi.org/10.1080/00313831.2020.1816577>
- Souza, C. P. S., & Takahashi, A. R. W. (2019). Dynamic capabilities, organizational learning and ambidexterity in a higher education institution. *The Learning Organization*, 26(4), 397-411. <https://doi.org/10.1108/TLO-03-2018-0047>
- Tabaghdehi, S. K. H., Safariyan, S., & Enaayati, T. (2018). The effect of organizational on intellectual capital considering the meditating role of academic entrepreneurship in Mazandaran Islamic Azad University. *Revista Publicando*, 5(16), 483-504.
- Tandon, A., Kaur, P., Mäntymäki, M., & Dhir, A. (2021). Blockchain applications in management: A bibliometric analysis and literature review. *Technological Forecasting and Social Change*, 166, 1-19. <https://doi.org/10.1016/j.techfore.2021.120649>
- Tight, M. (2018). Higher education journals: Their characteristics and contribution. *Higher Education Research & Development*, 27(3), 607-619. <https://doi.org/10.1080/07294360.2017.1389858>
- Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538. <https://doi.org/10.1007/s11192-0090146-3>
- Vera-Baceta, M. A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, 121(3), 1803-1813. <https://doi.org/10.1007/s11192-019-03264-z>
- Voolaid, K., & Ehrlich, Ü. (2017). Organizational learning of higher education institutions: The case of Estonia. *The Learning Organization*, 24(5), 340-354. <https://doi.org/10.1108/TLO02-2017-0013>
- Zhou, H., & Tu, C. C. (2021). Influential factors of university teachers' lifelong learning in professional development. *Australian Journal of Adult Learning*, 61(2), 267-297.

# Dynamic Information and Analytical System for Assessing and Managing Acquired Knowledge

Yana Goncharova

yana.goncharova@unicampania.it

<https://orcid.org/0000-0002-3245-2285>

University of Campania “Luigi Vanvitelli”, Italy

Olga Sheveleva

shoe.asp19@uni-dubna.ru

<https://orcid.org/0000-0002-5178-3573>

Dubna State University, Russia

Vladimir Dobrynin

i@vdobrynin.ru

<https://orcid.org/0000-0002-9770-2695>

Dubna State University, Russia

Recibido: 25 de julio del 2022 / Aceptado: 7 de septiembre del 2022

doi: <https://doi.org/10.26439/ciis2022.6071>

**ABSTRACT.** The paper presents the concept of a system for the intellectual digital e-university of the future (IDEUF) based on an integrated database that ensures the correct function of the whole university system. Particular attention is paid to the process of checking students' required and current level of acquired knowledge, which is considered one of the main parts of the educational process. The article provides a detailed comparative analysis of actual testing systems and presents a proper testing system model within the IDEUF framework. The proposed system considers the testing principles, scales, and technologies for test building and preparing educational materials

**KEYWORDS:** testing systems, digital education, educational materials, database

## SISTEMA DINÁMICO DE INFORMACIÓN Y ANÁLISIS PARA LA EVALUACIÓN Y GESTIÓN DEL CONOCIMIENTO ADQUIRIDO

RESUMEN. El artículo presenta el concepto del sistema de e-universidad digital intelectual del futuro (IDEUF, por sus siglas en inglés) fundado en una base de datos integrada que asegure el correcto funcionamiento de todo el sistema universitario. Se presta especial atención al proceso de verificación del nivel requerido y real del conocimiento adquirido por los estudiantes, que se considera una de las partes principales del proceso educativo. El artículo proporciona un análisis comparativo detallado de los sistemas de prueba reales y presenta un modelo de sistema de prueba adecuado dentro del marco de la IDEUF. El sistema propuesto considera los principios de evaluación, escalas y tecnologías para la formación de pruebas y para la preparación de materiales educativos.

PALABRAS CLAVE: sistemas de prueba, educación digital, materiales educativos, base de datos

## 1. INTRODUCTION

Nowadays, we live in the age of digitalization, when many business processes, and not only, are being moved into online environments, and new processes are also being formed. Digital education has been developing exponentially in the last few years, especially due to contemporary realities, such as the COVID-19 pandemic. All types of digital tools for education, including test systems, now constitute an integral part of our everyday routine and should be perceived not as a tribute to fashion but as a necessity for younger generations that are often called “digital natives” (Małkosa, 2013).

The term “digital technologies” is used in this research for the widest variety of software and hardware tools, including information communication technologies. The term “digital education” describes the wide range of educational processes that involve digital technologies (Blundell et al., 2016). Implementing digital technologies in education is a mechanism of educational reform introduced by means of a transformation of teaching practices. Although it is a mighty intermediary instrument, the Internet should not be perceived as the ruling master of modern education; instead, schools and universities set goals and identify the ways to achieve them with help from a variety of digital tools, like computer testing or e-course books (Małkosa, 2013). Multimedia study environments are more flexible, get students more involved in the educational process, give access to a great variety of informational sources, including teaching materials and opportunities to develop professionally, and so improve the effectiveness of the process, help minimize the pressure on teachers and raise the overall level of education. The researchers supporting online education worldwide underline that it helps to prepare students for future professional activities and working realities as well as to provide opportunities to bring high-quality education to the most remote areas (educational equity) (Best & Dunlap, 2012; Małkosa, 2013).

Online education requires the following conditions: high-speed Internet connection, a variety of devices (laptops, tablets, one-to-one computing environments, etcetera), online and/or blended classes, software for personalized learning, learning management systems to gather and track data, provide a platform for online institutions, personalize content, maintain administrative records, etcetera (Best & Dunlap, 2012).

Unfortunately, the full potential of digital education has yet to be realized, which presents an attractive field for a broad spectrum of researchers. Students do not demonstrate progress just using e-books instead of their paper copies. That is why it is necessary to improve the systems of online education. The requirements for specialists are changing, new professions are emerging, and the established standards of the educational process as a whole are constantly being revised. As a result, digital instruments should be improved, and there is also a vital need for a “modernized” university capable of solving a number of new tasks that remain unsolved by the classical university.

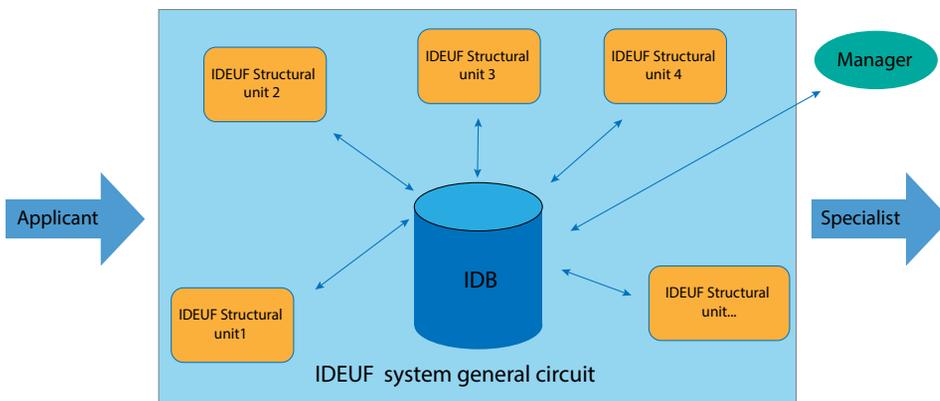
A university is an organization for the acquisition (transfer) of knowledge. A modern university should be an integral system, with an applicant as input and a specialist as output. Applicants' knowledge at the start of the process and specialists' at its end correspond to standards projected within the system and may change due to current requirements and realities. The system is an integral unit, which includes interacting, closely interconnected structures that ensure the correct working of the system.

The educational process, university infrastructure, scientific and methodological activities, housing stock (dormitories), student leisure, etcetera are university structures that are functionally and structurally different but ensure the unity of the university system and its ability to fulfill its educational task. The unity of these structures should be ensured by an integrated environment and an integrated knowledge base. The latter provides a timely, rapid exchange of knowledge, which is updated and accumulated in a prompt manner and creates a basis by which to manage the system, considering the needs of all its structural components.

## 2. MANAGEMENT OF THE INTELLECTUAL DIGITAL E-UNIVERSITY OF THE FUTURE (IDEUF) SYSTEM

The intellectual digital e-university of the future system includes an integrated database (IDB) and all other directly interconnected components of the university system. This approach allows to manage the entire system more effectively and to set up the main process of the system, the students' education, and other vital processes for the system in an efficient manner. The work of all components, including feedback, continuous and well-timed interaction of components, and timely data exchange, is provided by an IDB. It acts not only as a data exchange center for the system but also as a center for knowledge/experience accumulation. Accumulated knowledge can be used to make further management decisions regarding the university system.

**Figure 1**  
*IDEUF system*



### 3. COMPARATIVE ANALYSIS OF TESTING SYSTEMS

This part of our research focuses on computer testing systems. Evaluative testing of students' knowledge is one of the fundamental elements of the educational process. The effectiveness of this process is closely related to the development of methods for knowledge control. Achievements in education sciences highlight the successful use of tests with the help of computer systems, for example, during the experimental implementation of the Unified State Exam in Russia. The variety of systems we will analyze can be applied to different levels of education, from elementary school to university. Checking the compliance of knowledge with specified standards with the help of a testing system is one of the most important processes in the university system, which is being designed in response to the great demand for innovations in control systems for students' knowledge and skills. Computer testing is considered the universal method of control and evaluation in the educational process. It also helps to individualize learning and provides students with opportunities to organize self-training, as well as to increase students' and teachers' productivity by reducing the time spent on checking drastically. With the help of testing systems, it is possible to create a large bank of tasks with high evaluation objectivity. Providing equal objective conditions for all students requires the correct control questions that guarantee the right answer is unambiguously distinguished from the wrong one. In addition, the testing results should be easily identifiable.

A comparative analysis of the most common systems in the educational field is necessary because there is a great variety of testing systems and software complexes on the market. All of them have the common function of providing control both locally and remotely.

Table 1 lists 10 testing systems with their main characteristics.

**Table 1**

*Testing system comparative tables*

N.º	Name of the system	Type	Characteristics
1.	Apolo	Cloud-hosted version	There are no built-in analytics, only the number of correct answers is visible. The user is displayed as an email address in the system. There is no error checking (it is possible to create a question where all the answers are incorrect, and none of them has a tick mark for the correct answer). Test notifications arrive by mail from the Apolo team, and not from the structure, that uses this system for testing. There is a timer for the whole test, but no timer for single questions. It is possible to create questions with single and multiple-choice answers (Tirthankar, 2021).

*(continúa)*

*(continuación)*

2.	SunRav Web Class	Box version	Platform for storing ready-made tests. To create tasks, tMaker constructor is also needed. In the system, it is possible to create two types of tasks: "true – false", "put a score from 1 to 10". tMaker program helps to expand this list to seven: single and multiple choice questions, match, ordered lists, questions with an open line to write an answer. There are four report types: Answer Matrix, User Outcomes, Group Reports, and Topic Reports. The report format is csv. (Vereshchagina & Shaker, 2009; LmsList.ru Distance Learning Systems, 2019).
3.	ISpring	Cloud-hosted version	Platform for online learning and testing. There are 14 types of tasks: matching, single or multiple choice, area selection, drag-and-drop, sequence etcetera. It is possible to change the design of each question and set the rules for the test: set points and penalties, mix tasks automatically before testing, specify the number of attempts and limit the response time for each question, so that students do not cheat. You can carry out tests using computer or mobile phone. There are 15 report types available. The system collects all information (answers, time) into reports that can be downloaded in excel format (Zaripova N. & Zaripova M., 2014).
4.	StartExam	Cloud-hosted version	There are analysts that check the answers and collect reports. In the test, you can set time limits for tasks and for the number of attempts and mix the questions automatically before starting the test. You can create polls and quizzes using 9 types of questions: single and multiple choice, sorting, matching, text input, essay, Likert scale, video interview and 360-degree assessment. In addition to questions, information slides can be added to the test. If a student makes a mistake, StartExam will automatically send him to this slide. The StartExam report has 29 fields by default, it is possible to select certain options and turn off the extra ones (LmsList.ru Distance Learning Systems, 2019).
5.	Indigo	Both cloud-hosted and box versions	Professional tool for testing and processing results, designed to solve different ranges of tasks: knowledge testing and control of knowledge level of students, psychological testing, surveys and competitions. The demo version is available on the official website, where a licensed copy can be purchased. There are three types of tests: a survey, a knowledge test and a training test. The latter shows the correct answers if the student makes a mistake. Test Builder offers to create five types of tasks: single and multiple choice, answer entry, ordering and matching. The number of questions is unlimited. It is possible to set time limits for each task and add explanations and tips. It is also possible to import questions of all types from text files (.txt), work simultaneously in the editor with several tests and copy questions and groups of questions from one test to another. The program will automatically send a link to the finished test by mail to every specified student. Statistics: the system shows the points or percentages the students scored and the mistakes they made in the test. Reports in XLS format. The system stores information about users and can provide information about the tests carried out by users. The program has an intuitive and modern interface. It is possible to divide users into groups and send them e-mails, as well as import and export user lists to transfer information to another device. Among the disadvantages of the system are the price and the size of the program that reaches 125 MB (without tests) (Glazov & Emelyanova, 2016; LmsList.ru Distance Learning Systems, 2019).

*(continúa)*

*(continuación)*

6.	TestMaker	Box version	Local software system for testing with an unlimited number of questions and answers. It allows the creation of sets of tests for monitoring students' knowledge, and it is possible to change previously created tests, display test results in percent, save test results in text/encrypted format, save newly created or modified sets of tests in a self-executable file. There is no possibility to build reports on the results of testing. also It is possible to create different types of questions, add graphics, images, formulas, time limits, and generate questions randomly. During the test, students see the time spent and the number of the tasks completed. There is no possibility to view the theoretical part before passing the test. The software works on Windows 2000/XP/Vista/Server 2003/Server 2008/7, and the interface is intuitive and modern (Matveeva & Frolova, 2016; Ilyina et al., 2017).
7.	Examiner	Cloud-hosted version	Program for exams, digital and "paper" testing, for educational institutions and any subject studied. It enables the use of text or graphics for questions. The number of tests and questions in tests is unlimited. The system provides separate roles for teachers and students. Both parts of the program can be installed on the same device or on different devices on the local network. Users have to log in to run the teachers' program. Within the program it is possible to create courses and link various tests, so students can choose courses and disciplines before going to the page of tests available for each selected course. There are only two types of questions, single and multiple choice. The software package provides an opportunity to display the time spent on the test and a preliminary assessment. The system allows to set a minimum percentage for successful test pass, skipping questions and moving on to the next blocks. Tasks for tests may be issued out of order (Matveeva & Frolova, 2016).
8.	RichTest	Box version	Software package for creating test tasks and conduct testing. It consists of three modules: Admin (designed for creating and editing test tasks), Customer (designed for testing), and Reports (designed for collecting and analyzing test results in the form of reports). The system complex allows to pass to the theory before the test and to the training mode, where the correct answer is displayed in case of an error. The test settings are very flexible, which allows the creation of various types of questions, tips and answers, but not questions of extended answer type or more complex tests with different correct values. It is also possible to set the level of difficulty for each question, to build reports on test results with detailed information about mistakes, hints, etc. Technical support is available as well (Alyautdinova et al., 2017).
9.	UniTest System	Both cloud-hosted and box versions	Powerful product for automation of the testing process. Using UniTest System, you can create your tests using text, graphic editor or specialized programs (for example, AutoCAD). Functionality of the UniTest System software package includes the ability to use any OLE objects in your questions. In each question, you can set a number of correct and/or partially correct answers. There are six types of tasks: multiple choice, unordered choice, ordered choice, match, direct input. Subtexts allow you to highlight sections in the test and build a database of questions. The UniTest System has a rather complex interface, so basic skills are required to work with it (Shakhov & Demkin, 2017).

(continuación)

10.	VeralTestEditor	Both cloud-hosted and box versions	Software package that provides testing or certification for students. The main advantage is the interface simplicity that allows for the creation of single .exe tests, which can be run on any personal computer on Windows without installing additional software. There are different types of questions available, such as single and multiple-choice questions, text and number input questions, match (order elements or group any elements according to certain characteristics), as well as opportunities to prepare "paper tests" to use during offline or blended classes. It is possible to import questions from a file and to insert pictures and resize them without help of other programs. The program has an intuitive and modern interface, and technical support is available. All necessary documentation is available in the program (menu item "Help – Contents") (Ilyina et al., 2017).
-----	-----------------	------------------------------------	--

1. Judging from Table 1, none of the presented systems can be used to implement the IDEUF system, since this system requires a combination of the following solutions:
  1. Direct binding of the test question to the materials used
  2. Possibility to sort users into groups and to assign tests immediately to user groups
  3. Possibility to quote a period when every particular test is available (for example, access from 00:00 March 1 to 00:00 March 7)
  4. Possibility to set time limits not only for the tests, but for individual questions as well
  5. Possibility to create tasks (questions) of various types
  6. Automatic task checking for questions with open answers
  7. Statistics for a certain period both for user groups as well as for single users
  8. Binding questions to educational materials and the possibility to create recommendations automatically
  9. Possibility to create different variations of tests by mixing lists of questions

Due to the fact that none of the present systems meets the entire range of requirements, a decision was taken to create a new system for knowledge assessment.

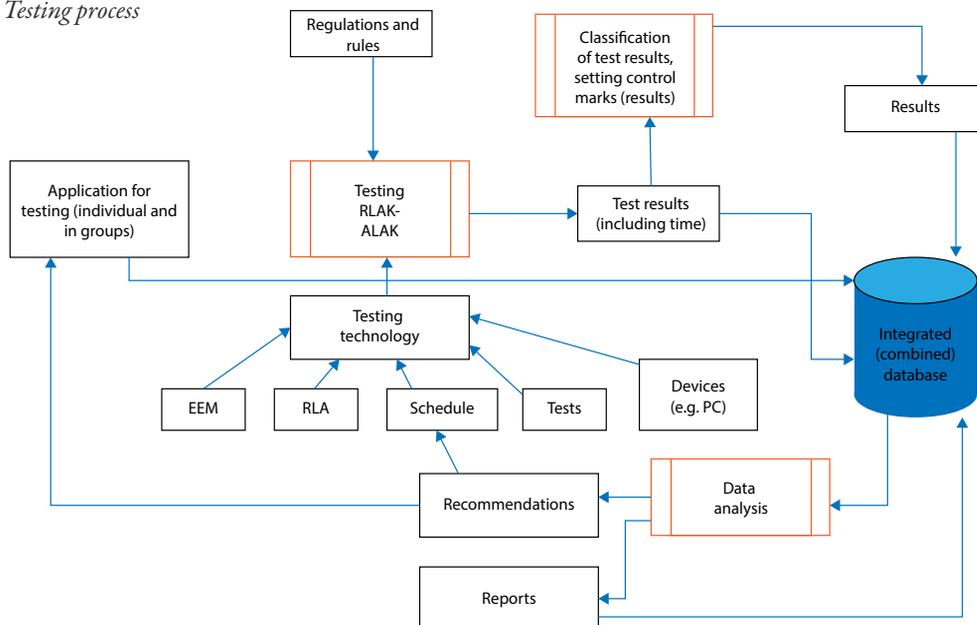
#### 4. TESTING PRINCIPLE

Our work is based on the testing process adapted from the principle of 3-component testing, which is described in Dobrynin et al. (2022a, 2022b). To quantify the final knowledge (FK), the function (F) of understanding (U), skill (S) and proficiency (P) is used.

$$FK=F(U, S, P) \quad (1)$$

The testing process starts with a testing request, which enters the input of the RLAK-ALAK testing process. The test results are sent to the integrated database to be analyzed (see Figure 2). Based on the analysis, reports and recommendations are generated, which are also recorded in the Database and can become the basis for adjustments to other processes and components of the IDEUF (for example, EEM, schedule).

**Figure 2**  
*Testing process*



*Note.* Required level of acquired knowledge (RLAK); actual level of acquired knowledge (ALAK); electronic educational materials (EEM).

## 5. THE PROCESS OF PREPARING AND EDITING EDUCATIONAL MATERIALS

The testing process is based on educational materials widely used in the learning process, that is why, in this part of our research, the process of preparing electronic educational materials (EEM) and the relationship between the processes of EEM preparing and testing are considered. In the process of preparing and editing educational materials, an electronic version of the EEM is formed, which is formatted according to a certain technology during the loading of the EEM into the system. The formatting of educational materials is based on the idea of

scientific texts analysis (Filozova & Dobrynin, 2012; Dobrynin & Filozova, 2014, 2015). Educational materials are divided into the following sections:

- Basic attributes (separate parts) of a scientific text (title, information about authors, abstract, keywords, main text, bibliographic references, bibliographic list)
- Key words (used to compile a glossary of terms)
- Questions to various parts of the main EEM (questions are allocated to each semantic part, the answer to which is contained in the corresponding passage of EEM)

### Questions to EEM

A logical-semantic network “question-answer” is being formed. Educational materials are prepared in a specific way for further use in the system. EM are divided into topics, a list of questions is compiled for each topic, direct links are formed between the questions and the paragraphs of each topic considering this or that issue (see Table 2). This division allows the building of individual learning paths: to choose questions of interest and study only those parts of the topic that correspond to the selected questions. In the educational materials, the main terms are highlighted, a list of keywords and phrases of the topic is compiled, which can be used to compile tests and check the test results.

**Table 2**

*Example of EEM division*

Topic text divided into semantic parts	Question numbers corresponding to the topic of the paragraph(s)
Digital energy, including smart instrumentation systems...	1,3,7
Digital products: including cloud and infrastructure products based on the development of proper processing centers...	2,4,5,9
Digital NPP, including: Digital NPP project; Digital operating template...	6,8,10

### Schedule

Schedule is another important component in EEM preparation. The schedule is directly related to EEM, at least in parts. The schedule indicates the dates and times when a certain part of EEM is opened (including tests), the deadlines for completing the tasks and sending them for control, as well as a timetable of online meetings for a single student (or group of students) with a teacher or more (see Table 3).

**Table 3***EM schedule for a discipline*

EEM	Dates and times (when the access to EEM is granted and blocked)
Topic 1: Studying experience of successful management of global corporations allowed the author to conclude that to improve the efficiency of...	01.09.2022 (00:00) – 31.07.2022 (00:00)
Topic 2: After studying the world experience, the author identified the following main models of knowledge management in corporations...	
Task 1: Draw a diagram...	01.09.2022 (10:00) – 10.09.2022 (00:00)
Task 2: Make a plan...	
Lecture 1	01.09.2022 (09:00) – 01.09.2022 (10:40)
Seminar 1	01.09.2022 (10:40) – 01.09.2022 (12:20)
Test 1	:00) – 30.09.2022 (00:00)

**6. CONCLUSION**

University should be a well-coordinated unified system. One of the main processes in university is the learning process and the related testing process. To achieve greater transparency while testing knowledge acquired by students for a certain period, we proposed a system which linked the educational material directly to the test questions. One of the main advantages of this system is that this approach alleviates teacher stress. With this system, routine work disappears. The material to be studied doesn't have to be specified in order to answer the test questions correctly; recommendations are formed automatically and require only one-time filling. Moreover, this approach is suitable for use in the IDEUF system.

Despite all the advantages of this approach, there are some limitations. This system is designed for computer testing, i.e. in electronic form. Another important condition and limitation is that the material must be submitted in electronic form and processed (classified) into certain groups before the educational process starts for further linking to the test questions.

In the following works, we are planning to develop the system that forms and links educational materials to test questions. The main feature of the testing process is the use of 3-component testing of the final knowledge. What is more, in the following works we are planning to discuss the relationship and possible modification of other components of the university system.

## REFERENCES

- Alyautdinova, G. R., Ignarina, E. S., & Raugina, L. R. (2017). Сравнительный анализ тестирующих программ TestMaker и RichTest [Comparative analysis of testing programs TestMaker and RichTest]. *Educational Technologies and Society*, 20(2), 377-384. <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-testiruyuschih-programm-testmaker-i-richtest>
- Best, J., & Dunlap, A. (2012). *Beyond access: Effective digital learning for a globalized world*. Mid-continent Research for Education and Learning. <https://files.eric.ed.gov/fulltext/ED544254.pdf>
- Blundell, C., Lee, K.-T., & Nykvist, S. (2016). Digital learning in schools: Conceptualizing the challenges and influences on teacher practice. *Journal of Information Technology Education: Research*, 15, 535–560. <https://doi.org/10.28945/3578>
- Dobrynin V. N., Filozova I. A. (2014). Семантический поиск в научных электронных библиотеках [Semantic search in scientific electronic libraries]. *Informatization of Education and Science*, 2(22), 111-127. <https://elibrary.ru/item.asp?id=21404653>
- Dobrynin V. N., Filozova I. A. (2015). Технология формирования каталога информационного фонда [Technology of information fund catalog generation]. *Computer Research and Modeling*, 7(3), 661-673. <https://doi.org/10.20537/2076-7633-2015-7-3-661-673>
- Dobrynin, V., Mastroianni, M. & Sheveleva, O. (2022a). A data warehousing system for ICT competencies assessment. En Abraham, A., Madureira, A. M., Kaklauskas, A., Gandhi, N., Bajaj, A., Muda, A. K., Kriksciuniene, D., Ferreira, J. C. (Eds.) *Innovations in bio-inspired computing and applications* (pp. 435-446). Springer. [https://doi.org/10.1007/978-3-030-96299-9\\_42](https://doi.org/10.1007/978-3-030-96299-9_42).
- Dobrynin, V., Mastroianni, M., & Sheveleva, O. (2022b). A New Structured Model for ICT Competencies Assessment Through Data Warehousing Software. In Abraham, A., Madureira, A. M., Kaklauskas, A., Gandhi, N., Bajaj, A., Muda, A. K., Kriksciuniene, D., & Ferreira, J. C. (Eds.), *Innovations in bio-inspired computing and applications* (pp. 435–446). Springer International Publishing. [https://doi.org/10.1007/978-3-030-96299-9\\_42](https://doi.org/10.1007/978-3-030-96299-9_42)
- Filozova, I., Dobrynin, V. (2012). Задача извлечения знаний из Электронных фондов на основе логико-семантической сети вопрос-ответ-реакция [Problems of knowledge extraction from digital funds on the base of the logic-semantic network question-answer-reaction]. *System Analysis in Science and Education*, 4, 92-106. <https://elibrary.ru/item.asp?id=21249826>

- Glazov A. V., Emelyanova V. N. (2016). Сравнительный анализ программных комплексов TestMaker и Indigo [Comparative analysis of TestMaker and Indigo software packages]. *Educational Technologies and Society*, 19(2), 519-527. <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-programmnyh-kompleksov-testmaker-i-indigo>
- Ilyina, T. V., Kuzmin, A. I., Mamonova, V. S. (2017). Обзор возможностей программных комплексов TestMaker и VeralTestEditor [Overview of the capabilities of the TestMaker and VeralTestEditor software packages]. *Educational Technology and Society*, 20 (2), 346-354. <https://cyberleninka.ru/article/n/obzor-vozmozhnostey-programmnyh-kompleksov-testmaker-i-veraltesteditor>
- LmsList.ru Distance Learning Systems (2019). Обзор 3 популярных систем для тестирования, аттестации и оценки персонала [Overview of 3 popular systems for testing, certification and evaluation of personnel]. Retrieved June 29, 2022, from June 17, 2022, from <https://lmslist.ru/sistema-testirovaniya/>
- Małkosa, P. (2013). Advantages and disadvantages of digital education. *Biuletyn Edukacji Medialnej*, 2, 21–31. <https://www.cceol.com/search/article-detail?id=113896>
- Matveeva, S. V., Frolova, K. A. (2016). Обзор возможностей программных комплексов TestMaker и Экзаменатор [Overview of the capabilities of TestMaker and Examiner software packages]. *Educational Technologies and Society*, 19(2), 549-557. <https://cyberleninka.ru/article/n/obzor-vozmozhnostey-programmnyh-kompleksov-testmaker-i-ekzamenator>
- Shakhov, Y. V., Demkin, D. V. (2017). Сравнительная оценка возможностей систем компьютерного тестирования TestMaker и UniTest System [Comparative evaluation of the capabilities of computer testing systems TestMaker and UniTest System]. *Educational Technologies and Society*, 20(2), 385-396. <https://cyberleninka.ru/article/n/sravnitel'naya-otsenka-vozmozhnostey-sistem-kompyuternogo-testirovaniya-testmaker-i-unitest-system>
- Tirthankar (2021). Система тестирования “Apolo” [Apolo testing system]. Retrieved June 29, 2022, from <http://tirthankar.ru/apolo-learning-system/apolo-landing-1/>
- Vereshchagina, E. A., & Shaker, T. D. (2009). Подготовка электронных учебных пособий в среде SunRav [Preparation of e-tutorials in SunRav environment]. Вологодские чтения [Vologda readings], (73), 59-61. <https://cyberleninka.ru/article/n/podgotovka-elektronnyh-uchebnyh-posobiy-v-srede-sunrav>
- Zaripova, N. Zh., & Zaripova, M. Zh. (2014). Использование программы iSpring Suite при создании мультимедийных электронных учебных пособий [Using iSpring Suite to create multimedia e-tutorials]. *Gaudeamus*, 2(24). <https://cyberleninka.ru/article/n/ispolzovanie-programmy-ispring-suite-pri-sozdanii-multimediynyh-elektronnyh-uchebnyh-posobiy>



# Estimación de la cantidad de heridos en accidentes de tránsito dentro de la provincia de Lima utilizando modelos de regresión lineal múltiple y PCA

Alisson Georgeth Avalos Tapia

ag.avalost@alum.up.edu.pe

<https://orcid.org/0000-0003-0985-0290>

Juan David Cárdenas Zúñiga

jd.cardenasz@alum.up.edu.pe

<https://orcid.org/0000-0002-3801-4227>

Rodrigo Fernando Caballero Chocano

rf.caballeroc@alum.up.edu.pe

<https://orcid.org/0000-0002-3090-1008>

Victor Andres Ayma Quirita

va.aymaq@up.edu.pe

<https://orcid.org/0000-0003-2987-2761>

Universidad del Pacífico, Perú

Recibido: 6 de agosto del 2022 / Aceptado: 23 de septiembre del 2022

<https://doi.org/10.26439/ciis2022.6073>

**RESUMEN.** Existe una alta probabilidad de que un accidente de tránsito deje víctimas, muchas de las cuales podrían encontrarse en un estado crítico que requiera de atención médica inmediata para su sobrevivencia. Sin embargo, en el Perú, donde los recursos de atención en salud en estas emergencias son bastante limitados, se tiene la necesidad de priorizar la atención inmediata de ciertos accidentes sobre otros. Por ello, el presente trabajo tiene como objetivo construir un modelo que estime la cantidad de heridos en accidentes de tránsito dentro de la provincia de Lima con base en registros de accidentes ocurridos entre los años 2017 y 2016, según el Censo Nacional de Comisarias del 2017. Para este fin, se desarrollará un análisis de

regresión lineal múltiple tomando ciertas variables seleccionadas a partir de los registros de accidentes; asimismo, se realizará un análisis de componentes principales en busca de representar mejor algunas de las variables dentro del modelo y extraer la información más relevante de estos datos. De esta manera, se pretende brindar información valiosa al personal de salud que atiende estos eventos, de forma que les permita priorizar la atención de aquellos accidentes cuyas condiciones permitan maximizar el número de sobrevivientes sobre el total de heridos.

**PALABRAS CLAVE:** accidentes de tránsito, número de heridos, modelo predictivo, regresión lineal múltiple, PCA

### ESTIMATION OF THE NUMBER OF INJURIES IN TRAFFIC ACCIDENTS WITHIN THE PROVINCE OF LIMA USING MULTIPLE LINEAR REGRESSION MODELS AND PCA

**ABSTRACT.** There is a high probability that traffic accidents will leave several victims in such a critical condition that they will require immediate medical care to survive. In Peru, where healthcare resources for emergencies are quite limited, it is essential to prioritize some cases over others. This work aims to build a model that estimates the number of people injured in traffic accidents in the province of Lima based on 2016 and 2017 accident records from the Police Station National Census. It develops a multiple linear regression analysis based on variables taken from the accident records and principal component analysis to represent better and simplify some of the original model variables and work with the most relevant information. The model seeks to provide valuable information to medical professionals to prioritize attention and maximize the number of survivors.

**KEYWORDS:** traffic accidents, people injured, predictive model, multiple linear regression, PCA

## 1. INTRODUCCIÓN

En el Perú, los accidentes de tránsito han sido definidos como prioridad en investigación debido a su alta frecuencia, su elevado nivel de mortalidad y la gran cantidad de afectados que quedan con lesiones irreversibles en estos eventos (Peden et al., 2004). Según la Defensoría del Pueblo (2021), en los últimos cinco años ocurrieron más de 420 000 accidentes de tránsito, los cuales dejaron cerca de 14 000 muertos y 272 000 personas heridas o discapacitadas. Además, los accidentes de tránsito representan entre el 1,5 % al 2 % del PBI nacional, lo que significa 1000 millones de dólares en costos de atenciones, según la Estrategia Sanitaria Nacional de Accidentes de Tránsito (ESNAT, 2009) del Ministerio de Salud. En este contexto, cabe mencionar que, entre todas las provincias peruanas, Lima es la que concentra la mayor cantidad de accidentes registrados, con cerca del 50 % de los casos sobre el total de accidentes a nivel nacional (Instituto Nacional de Estadística e Informática [INEI], 2018).

Frente a un accidente de tránsito, los principales responsables de acudir ante la emergencia son el Sistema de Atención Móvil de Urgencias (SAMU), el Cuerpo General de Bomberos Voluntarios, la red de atención de salud pública y privada, entre otros. Estos entes desempeñan un papel fundamental, puesto que de su respuesta oportuna depende la vida de las personas accidentadas; además, en un accidente de este tipo, solo cuentan con “una hora de oro” para asistir a las personas accidentadas que presenten heridas graves y aumentar su probabilidad de sobrevivir. Según Maldonado (2016), el 75 % de las posibles defunciones en un accidente podrían suceder si no se presta asistencia inmediata a la persona. Sin embargo, en ocasiones, varios accidentes se registran al mismo tiempo y, tomando en cuenta la cantidad limitada de los recursos de atención, es necesario tomar la difícil decisión sobre cuál de los casos debe recibir una atención prioritaria respecto del otro.

Así pues, este trabajo tiene por objetivo desarrollar un modelo que estime la cantidad posible de heridos en un accidente de tránsito en la provincia de Lima. Este modelo pretende convertirse en una primera fuente de consulta e información capaz de ayudar en la priorización de la atención de accidentes de tránsito simultáneos para optimizar la respuesta de atenciones e intentar maximizar la cantidad de sobrevivientes del total de víctimas en un accidente de este tipo. Para tal fin, se utilizó la información del registro de accidentes de tránsito obtenida del Censo Nacional de Comisarías del 2017 (INEI, 2018), a partir de la cual se armaron diferentes modelos de estimación en busca de aquel que se aproxime mejor a los datos reales. Para el desarrollo de tales modelos se emplearon herramientas como la regresión lineal múltiple y el análisis de componentes principales (PCA).

Este artículo se encuentra organizado de la siguiente manera: la segunda sección presenta una revisión general acerca de los conceptos básicos de los métodos estadísticos y algebraicos utilizados; el conjunto de datos empleado y la metodología son descritos en la tercera sección; los resultados se presentan en la cuarta sección; y, finalmente, en la quinta sección, se discuten las conclusiones y recomendaciones a futuro de este trabajo.

## 2. MARCO TEÓRICO

### 2.1 Análisis de componentes principales

El análisis de componentes principales (PCA) es un método estadístico que permite simplificar la complejidad de los espacios de muestreo multidimensional, al mismo tiempo que conserva su información (Amat, 2017). En este sentido, mediante PCA se calculan nuevas variables ortogonales, denominadas *componentes principales*, con el fin de extraer la información más importante de una distribución de datos, además de comprimir y simplificar su tamaño, y analizar la estructura de observaciones y variables (Abdi & Williams, 2010). Puesto que los primeros componentes principales son los que llevan consigo la mayor variabilidad del modelo, cuanto mayor sea el número de componentes, mayor es la pérdida de su valor, dada la menor variabilidad con la que representan el sistema. Bajo este concepto, en general, los últimos componentes son los que representan el ruido del proceso y, por consiguiente, no son considerados (García & Fuente, 2011).

### 2.2 Regresión lineal múltiple

El análisis de regresión lineal múltiple permite establecer la relación entre una variable dependiente ( $Y$ ) y un conjunto de variables independientes. En este trabajo, la variable independiente es un vector con observaciones, es decir,  $(X_1, X_2, \dots, X_i)$ . Asimismo, cada una de las variables es un vector con observaciones, en otras palabras,  $X_i = [X_{i1}, X_{i2}, \dots, X_{ik}]$ .

Para la aplicación de esta técnica, se busca que tanto la variable dependiente como las variables independientes sean continuas. Sin embargo, también es posible utilizar esta técnica cuando se relaciona una variable dependiente continua con un conjunto de variables categóricas; o en el caso de que se relacione una variable dependiente nominal con un conjunto de variables continuas (Montero Granados, 2016).

De acuerdo con Montero Granados (2016), en el modelo de regresión lineal múltiple se asume que más de una variable influye o está correlacionada con el valor de una tercera variable; por esta razón, se espera que los sucesos tengan una forma funcional, como se muestra en la ecuación 1:

$$Y_j = b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_kx_{kj} + u_j \quad (1)$$

Donde  $Y$  es la variable endógena,  $X$  las variables exógenas,  $u$  los residuos y  $b$  los coeficientes estimados del efecto marginal entre cada  $X$  y  $Y$ . De acuerdo con el método diseñado por Rubio y Toma (2019), se busca minimizar la suma de cuadrados del error del modelo planteado, función señalada en la ecuación 2:

$$Q(b) = \sum^n error_i^2 \tag{2}$$

Donde el error es entendido como la diferencia entre el valor real y el estimado, como se muestra en la ecuación 3:

$$error = Y_{estimado} - Y_{real} \tag{3}$$

Adicionalmente, en busca de minimizar el error, se calculan las derivadas parciales respecto a cada variable de la regresión (ecuación 4):

$$\frac{\partial Q(b)}{\partial b} = 0 \tag{4}$$

Cada derivada parcial ofrece una ecuación, las cuales, en conjunto, forman el sistema de ecuaciones normales presentado en la Figura 1. La solución a este sistema consiste en un vector que tiene por elementos los coeficientes del modelo de regresión lineal múltiple  $[b_0, b_1, b_2 \dots b_k]$ . Usualmente, se utilizan métodos de solución como el cálculo de la pseudoinversa de la matriz A o la factorización QR. Esta última se emplea en el presente trabajo y se explica con más detalle en el ítem 2.3.

**Figura 1**

*Sistema de ecuaciones  $A \times b = B$*

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \dots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \dots & \sum_{i=1}^n x_{i1}x_{ik} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}x_{i1} & \sum_{i=1}^n x_{i2}^2 & \dots & \sum_{i=1}^n x_{i2}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \dots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_{i1} \\ \sum_{i=1}^n y_{i1}x_{i1} \\ \sum_{i=1}^n y_{i1}x_{i2} \\ \vdots \\ \sum_{i=1}^n y_{i1}x_{ik} \end{bmatrix}$$

### *Variables categóricas*

Dado que la regresión lineal que se quiere generar cuenta con variables categóricas entre las que se requiere analizar, se debe realizar un tratamiento adecuado respecto de la naturaleza de este tipo de información. Estas variables permiten darle una interpretación a las  $n$  categorías que se tienen en una variable a través de la asignación de  $n - 1$  coeficientes, los cuales indican ausencia o presencia de cada categoría. Se considera  $n - 1$  porque la ausencia de todas las variables igualmente se traduce en la presencia de otra variable, ya que si estuviera representada como otro coeficiente, se trataría de una colinealidad de variables (Montero Granados, 2016).

### 2.3 Factorización QR

La factorización QR es un método clásico que permite resolver sistemas sobredeterminados (Vadillo, 2018). Es ampliamente utilizado en algoritmos computacionales para diversos cálculos, como la resolución de ecuaciones o la determinación de valores propios (Lay, 2012). Básicamente, este método permite factorizar una matriz con columnas linealmente independientes ( $A$ ) en función de una matriz ortonormal ( $Q$ ) y una matriz diagonal superior ( $R$ ), tal que  $A = QR$ . En el presente trabajo, la factorización QR se emplea para resolver un sistema de ecuaciones lineales de forma más simple.

### 2.4 Coeficiente de regresión ajustado

El coeficiente de determinación  $R$  cuadrado ajustado mide la proporción de la variación total de la variable explicada por la línea de regresión estimada (Rubio & Toma, 2019). Esta medida estadística tiene una interpretación sencilla, pues un valor de 0 significa que el modelo explica el 0 % de la variación total, mientras que un valor de 1 significa que el 100 % de la variación de la variable es explicada por el modelo. Para el cálculo de este estadístico, se emplea la fórmula presentada en la ecuación 5.

Se plantea el uso del coeficiente  $R$  cuadrado ajustado, que presenta una interpretación similar a la del estimador  $R$  cuadrado, debido a que el coeficiente no ajustado suele sobrestimar la proporción de la variación explicada (Rubio & Toma, 2019). En términos sencillos, el coeficiente determina la cercanía de la línea de regresión a los datos.

$$r^2 \text{ ajustado} = 1 - \frac{(n - 1) \sum_i^n (Y_{real_i} - Y_{estimada_i})^2}{(n - k - 1) \sum_i^n (Y_{real_i} - \bar{Y})^2} \quad (5)$$

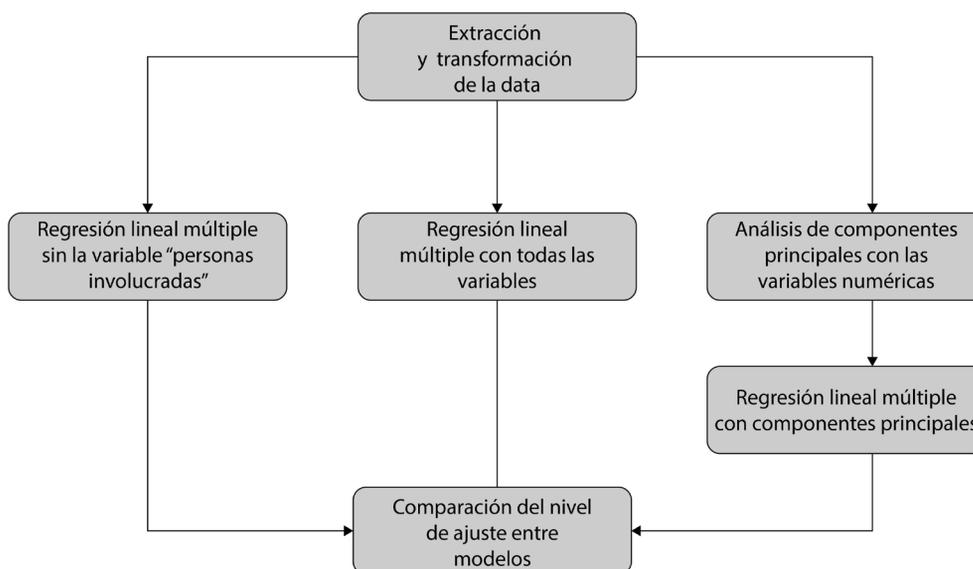
### 3. METODOLOGÍA

A continuación, se detalla el proceso realizado. Conforme se presenta en la Figura 2, el proceso comienza con la extracción, filtrado y transformación de datos del Censo Nacional de Comisarías del 2017 (INEI, 2018). Posteriormente, se construyeron dos modelos de regresión lineal múltiple con los datos obtenidos. El primero se armó tomando en cuenta todas las variables; para el segundo, dada la gran influencia de la variable “personas involucradas”, se propuso un modelo que prescindía de esta última variable. También se generó un análisis de componentes principales entre cuatro variables cuantitativas para luego realizar una nueva regresión lineal múltiple con menos variables.

Finalmente, se compararon ambos modelos para identificar cuál de ellos se aproximaba mejor a la variable dependiente “cantidad de heridos”.

**Figura 2**

*Flujograma del procedimiento metodológico*



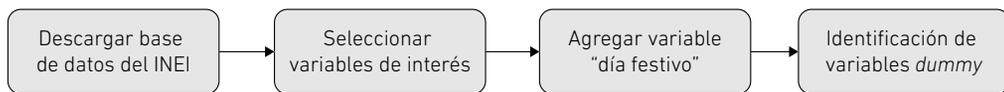
En adelante, la presente sección describirá en detalle cómo se procesaron los datos, así como la aplicación del PCA sobre las variables cuantitativas y también el análisis de regresión múltiple.

### 3.1 Datasets y preprocesamiento

Para la realización del trabajo, la data fue extraída del Censo Nacional de Comisarías del 2017 (INEI, 2018), de la sección de accidentes de tránsito. Luego se procedió a filtrar todos aquellos registros de accidentes de tránsito en la provincia de Lima. De estos datos, se seleccionaron las variables de interés para trabajar en el modelo, y se agregó la variable “día festivo”, que permite identificar si la fecha del accidente corresponde o se acerca a una fecha de ese tipo y analizar su relevancia en el modelo. La inclusión de esta variable se debe a que, por lo general, hay más muertes por accidentes de tránsito en días festivos (NHTSA, 2019). Asimismo, para la inclusión de las variables categóricas, se crearon variables “categóricas”. Este proceso se presenta en la Figura 3.

**Figura 3**

*Flujograma de la transformación del dataset*



De este modo, se obtuvieron las variables que se incluyen en el modelo de regresión lineal múltiple, como se muestra en la Tabla 1.

**Tabla 1**

*Clasificación de las variables seleccionadas*

Variables independientes				
Categoría	Variable	Nombre de la variable	Descripción	Tipo de variable
Fecha	$X_1$	Hora	Hora del día en que se registró la ocurrencia del accidente.	Cuantitativa
	$X_2$	Día festivo	Cuantitativa, se asignan probabilidades que indican cuando un día es festivo.	Cuantitativa
Días de la semana (lunes es la referencia)	$X_3 - X_8$	Martes a domingo	Nombre del día de la semana en que se registró la ocurrencia del accidente.	Categórica
Trimestre (trimestre 1 es la referencia)	$X_9 - X_{11}$	Trimestre 2 a trimestre 4	Número del trimestre del año en que se registró la ocurrencia del accidente.	Categórica

*(continúa)*

(continuación)

Tipo de vía (interprovincial es la referencia)	$X_{12} - X_{14}$	Urbana, rural, otras	Clasificación del tipo de vía en que se registró la ocurrencia del accidente.	Categoría
Tramo de la vía (intersección es la referencia)	$X_{15} - X_{18}$	Recta, curva, rotonda, bifurcada	Clasificación del tramo de la vía en que se registró la ocurrencia del accidente.	Categoría
Tipo de accidente (atropello es la referencia)	$X_{19} - X_{21}$	Otros, colisión, despiste o volcadura	Clasificación del tipo de accidente registrado.	Categoría
Vehículos	$X_{22} - X_{24}$	Vehículos mayores, vehículos mayores masivos, vehículos menores	Cantidad de vehículos involucrados en el accidente registrado.	Cuantitativa
Personas	$X_{25}$	Personas involucradas	Cantidad de personas involucradas registradas en el accidente (heridos y muertos).	Cuantitativa
Variable dependiente				
Categoría	Variable	Nombre de la variable	Descripción	Tipo de variable
Personas	Y	Heridos	Cantidad de personas heridas registradas en el accidente.	Cuantitativa

Cabe aclarar que la variable  $X_2$  (día festivo) asigna valores entre 1 y 0 a cada registro de acuerdo con la distancia de la fecha de accidente con la fecha festiva más próxima. Donde 1 corresponde a una fecha festiva, 0,7 a registros con un día de distancia, 0,3 a accidentes con dos días de distancia y 0 para una distancia mayor a dos días.

### 3.2 Regresión lineal múltiple con todas las variables

Para el desarrollo, implementación y aplicación de la técnica de regresión lineal múltiple, se usó el *software* MATLAB. Primero, se cargaron los datos transformados. Luego, se calcularon la matriz A y el vector B como parte del sistema de ecuaciones normales de la regresión lineal múltiple. Con este fin, se llenaron los datos de forma iterativa, como se indica en la Figura 4.

Figura 4

*Pseudocódigo del cálculo de la matriz A y el vector B*

Crear matriz A y vector B llenos de ceros

$A(1,1)$  = número de registros

$B(1,1)$  = sumatoria de columna heridos

Para i desde 1 hasta la última columna de los datos:

$A(1,i+1)$  = suma de elementos de la columna i de los datos

Para j desde 1 hasta la última columna de los datos:

$A(j+1,1)$  = suma de elementos de la columna j de los datos

$B(j+1,1)$  = producto punto de la columna j de los datos y la columna de heridos

Para i desde 1 hasta la última columna de los datos:

$A(j+1,i)$  = producto punto de la columna j de los datos y la columna i de los datos

Después, se utilizó el método de factorización QR para resolver este sistema de ecuaciones. La factorización se realizó mediante la función predefinida de MATLAB y, posteriormente, se empleó la sustitución hacia atrás del método de eliminación Gauss-Jordan para hallar la solución del sistema (vector con los coeficientes del modelo). La resolución del sistema se desarrolló de esta manera dadas las dimensiones de la matriz R (diagonal superior), las cuales no permitían hallar su inversa debido a limitaciones de la precisión de cálculo del *software*.

Para medir el grado de ajuste del modelo a los datos, se implementó el cálculo del coeficiente de regresión R cuadrado ajustado. Esta medida fue calculada para cada modelo y, luego, se empleó para interpretar los resultados del trabajo.

El modelo de regresión lineal múltiple sin la variable “número de personas” siguió el mismo procedimiento que lo expuesto anteriormente, solo que para este modelo esta variable no fue tomada en consideración.

### 3.3 Regresión lineal múltiple con PCA

Por otro lado, se realizó un análisis de componentes principales (PCA) mediante la función predefinida en MATLAB sobre las variables cuantitativas  $X_{22}$ ,  $X_{23}$ ,  $X_{24}$  y  $X_{25}$  (relacionadas con la cantidad de vehículos y la cantidad de personas involucradas), por la posibilidad de darse una correlación entre el número de vehículos y el número de personas. El método de PCA

permite, a través de los autovectores (después de hallar los autovalores de las componentes que representan sus respectivas varianzas), reflejar las proyecciones de los datos sobre los respectivos componentes.

Luego de realizar esta transformación, se implementó un modelo de regresión que incluía el resto de las variables que no fueron analizadas mediante PCA, y los componentes principales más representativos de la transformación, los cuales correspondían a los dos primeros componentes principales. Esta regresión se hizo siguiendo los pasos descritos anteriormente: tanto el cálculo de coeficientes como la medida del grado de ajuste.

#### 4. RESULTADOS

En esta sección, se presentan, primero, los coeficientes calculados para cada modelo de estimación desarrollado y su  $R$  cuadrado ajustado respectivo. Luego, se muestran los resultados obtenidos a partir de la evaluación de tres registros de accidentes para los tres modelos con mejor ajuste.

La Tabla 2 presenta el valor de los coeficientes calculados mediante el programa de MATLAB para cada variable de acuerdo con cada uno de los modelos de estimación desarrollados. Además, muestra el valor del  $R$  cuadrado ajustado, el cual nos da una idea del nivel de precisión de los modelos.

En la Tabla 2, se puede observar que el modelo que explica mejor esta regresión lineal múltiple es el que se basa en todas las variables, ya que tiene un  $R$  cuadrado ajustado superior al de los demás modelos. Igualmente, se aprecia que el modelo sin la variable “personas involucradas” es el que realiza de manera menos adecuada las predicciones del número de heridos en los accidentes. Mientras que el  $R$  cuadrado del modelo de componentes 1 y 2 se asemeja al que se basa en todas las variables, con la consideración de que este modelo optimiza el proceso de alguna manera, ya que usa dos variables menos respecto al modelo inicial, que incluye todas las variables.

**Tabla 2**

*Coefficientes de los distintos modelos elaborados para la estimación de heridos*

Variable	Nombre de la variable	Coefficientes	Modelo basado en todas las variables	Modelo sin la variable “personas involucradas”	Modelo con componentes 1 y 2
$X_0$	Constante	b0	-0,3085	1,1326	1,0953
$X_1$	Hora	b1	-0,0004	0,0009	-0,0004
$X_2$	Día festivo	b2	-0,0253	0,0175	-0,0231
$X_3$	Martes	b3	-0,0049	-0,0424	-0,0145

(continúa)

(continuación)

X <sub>4</sub>	Miércoles	b4	-0,0131	-0,0528	-0,0189
X <sub>5</sub>	Jueves	b5	-0,0044	-0,0366	-0,0060
X <sub>6</sub>	Viernes	b6	-0,0179	-0,0575	-0,0199
X <sub>7</sub>	Sábado	b7	-0,0094	0,0052	-0,0077
X <sub>8</sub>	Domingo	b8	-0,0133	0,0543	0,0101
X <sub>9</sub>	Trimestre 2	b9	0,0118	0,0186	0,0143
X <sub>10</sub>	Trimestre 3	b10	0,0484	0,0299	0,0408
X <sub>11</sub>	Trimestre 4	b11	0,0377	0,0369	0,0371
X <sub>12</sub>	Vía urbana	b12	0,0007	-0,1326	0,0239
X <sub>13</sub>	Vía rural	b13	-0,1454	0,0215	-0,0551
X <sub>14</sub>	Otro tipo de vía	b14	-0,0114	-0,1633	0,0313
X <sub>15</sub>	Recta	b15	-0,0019	-0,1268	-0,0105
X <sub>16</sub>	Curva	b16	-0,0074	-0,1506	-0,0407
X <sub>17</sub>	Rotonda	b17	-0,0556	-0,3323	-0,1051
X <sub>18</sub>	Bifurcación	b18	0,0261	-0,3268	-0,0245
X <sub>19</sub>	Otro tipo de accidente	b19	-0,0414	-0,1905	-0,1828
X <sub>20</sub>	Colisión y choque	b20	-0,2245	-0,3905	-0,4882
X <sub>21</sub>	Despiste y/o volcadura	b21	0,2533	-0,1351	0,1999
X <sub>22</sub>	Cantidad de vehículos mayores	b22	-0,5632	-0,0404	
X <sub>23</sub>	Cantidad de vehículos mayores masivos	b23	-0,4681	0,3366	
X <sub>24</sub>	Cantidad de vehículos menores	b24	-0,1747	0,4232	
X <sub>25</sub>	Personas involucradas	b25	0,8592		
X <sub>26</sub>	Componente principal 1	b26			0,7562
X <sub>27</sub>	Componente principal 2	b27			-0,4060
R cuadrado ajustado	0,7673	0,0863	0,7375		

*Nota.* Resultados obtenidos mediante el procesamiento de datos en MATLAB.

La Tabla 3 muestra tres casos de accidentes tomados de la base de datos, junto con los valores de sus variables. En la parte inferior de la tabla, se observa el total de heridos estimados como resultado de la evaluación de cada uno de los tres accidentes en cada uno de los tres modelos de estimación con mejor ajuste seleccionados: modelo basado en todas las variables, modelo sin variable “personas involucradas”, y modelo con componentes 1 y 2.

**Tabla 3**

*Resultados de los tres mejores modelos de estimación*

Variable	Nombre de la variable	Accidente 1	Accidente 2	Accidente 3
X <sub>0</sub>	Constante	1	1	1
X <sub>1</sub>	Hora	6	19	22
X <sub>2</sub>	Día festivo	0	0	0,7
X <sub>3</sub>	Martes	0	0	0
X <sub>4</sub>	Miércoles	1	0	0
X <sub>5</sub>	Jueves	0	0	0
X <sub>6</sub>	Viernes	0	0	0
X <sub>7</sub>	Sábado	0	1	0
X <sub>8</sub>	Domingo	0	0	0
X <sub>9</sub>	Trimestre 2	0	1	0
X <sub>10</sub>	Trimestre 3	0	0	0
X <sub>11</sub>	Trimestre 4	0	0	1
X <sub>12</sub>	Vía urbana	0	1	1
X <sub>13</sub>	Vía rural	0	0	0
X <sub>14</sub>	Otro tipo de vía	0	0	0
X <sub>15</sub>	Recta	1	0	1
X <sub>16</sub>	Curva	0	0	0
X <sub>17</sub>	Rotonda	0	0	0
X <sub>18</sub>	Bifurcación	0	0	0
X <sub>19</sub>	Otro tipo de accidente	0	0	0
X <sub>20</sub>	Colisión y choque	0	1	0
X <sub>21</sub>	Despiste y/o volcadura	1	0	0
X <sub>22</sub>	Cantidad de vehículos mayores	0	1	2
X <sub>23</sub>	Cantidad de vehículos mayores masivos	1	1	0

(continúa)

(continuación)

X <sub>24</sub>	Cantidad de vehículos menores	0	0	0
X <sub>25</sub>	Personas involucradas	27	11	6
X <sub>26</sub>	Componente principal 1	24,21	8,6076	3,8
X <sub>27</sub>	Componente principal 2	-4,1	-1,1647	0,49
Total de heridos estimado según modelo basado en todas las variables	22,6575	7,8818	3,7296	
Total de heridos estimado según modelo sin variable "personas involucradas"	1,1600	0,9469	0,8618	
Total de heridos estimado según modelo con componentes 1 y 2	21,2346	7,6114	3,7950	
Total de heridos reales	25	10	2	

*Nota.* Datos de prueba tomados del Censo Nacional de Comisarías del 2017 (INEI, 2018).

## 5. DISCUSIÓN DE LOS RESULTADOS

### 5.1 Modelo basado en todas las variables

En la Tabla 2, en el análisis del modelo basado en todas las variables, se puede identificar que la principal variable que explica la cantidad de heridos en un accidente de tránsito es la de “personas involucradas” ( $X_{25}$ ); así, cuando se aumenta una persona involucrada en el accidente, la cantidad de heridos aumenta en 0,86 personas (manteniendo las demás variables como constantes). Vale notar, además, que la inclusión de esta variable explica el modelo de manera bastante potente (con un  $R$  cuadrado de 0,77). Esto quiere decir que el modelo explica satisfactoriamente la variación total que presentan los datos. Por ello, se realizó la aplicación del modelo sin esta variable explicativa para ver únicamente la influencia de las demás variables sobre el modelo.

En la Tabla 2, a partir del valor de los coeficientes hallados, se puede apreciar que, contrario a las expectativas, la hora ( $X_1$ ) parece no influir mucho, al igual que el día festivo ( $X_2$ ). Por otra parte, el día lunes (variable categórica referencial) es el día que más influye en el modelo (aunque sea mínimo). El trimestre 4 ( $X_{11}$ ) es el que influye más entre todos los trimestres. La bifurcación ( $X_{18}$ ) es el tramo de vía que afecta más, seguido por la intersección. Por otro lado, la variable despiste y/o volcadura ( $X_{21}$ ) presenta el segundo coeficiente más alto en el modelo. Se puede interpretar que es la categoría de tipo de accidentes asociada a un mayor número de heridos.

Finalmente, se introdujeron tres registros de accidentes de tránsito tomados de la base de datos y se obtuvieron los resultados mostrados en la Tabla 3. De estos registros, se obtuvo una predicción bastante cercana a la cantidad de heridos reales. Esto concuerda con los coeficientes de determinación, que indican una cercanía considerable entre las predicciones estimadas por el modelo y los datos reales producidos en dichos accidentes.

### 5.2 Modelo sin la variable “personas involucradas”

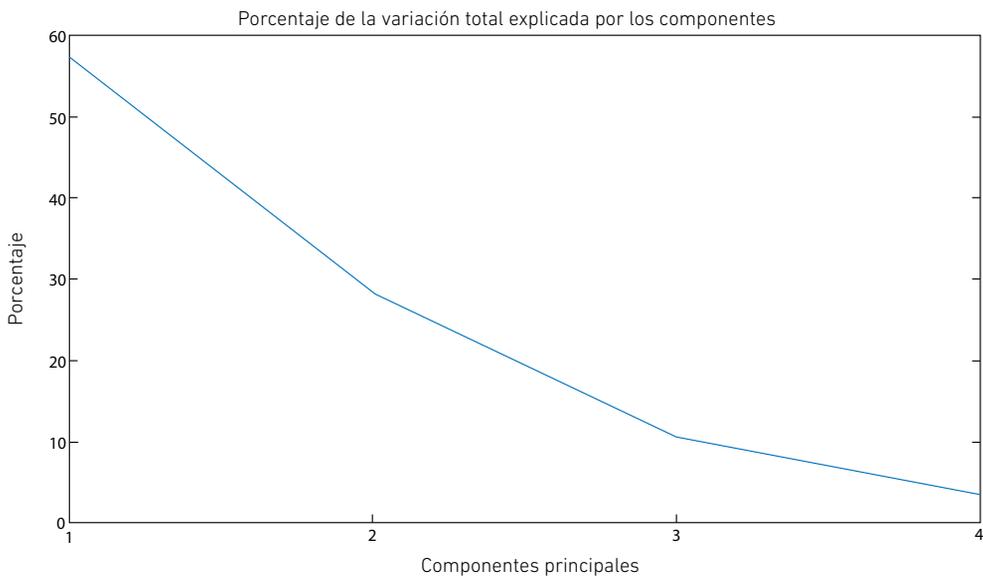
Como se ya se ha mencionado, el primer modelo implementado, que se construye con base en todas las variables, explica bastante bien la cantidad de heridos a partir de la cantidad de personas involucradas ( $X_{25}$ ). En vista de ello, se plantea una situación en la que no se pueda disponer del número de personas, excluyendo esta variable del modelo. En la Tabla 2 se aprecian los resultados de este nuevo modelo, donde el valor de su métrica  $R$  cuadrado disminuye a 0,08. Esto nos llevaría a desestimar las predicciones realizadas por este modelo, ya que, como se observa en la Tabla 3, donde se evaluaron registros de accidentes de tránsito extraídos de la base de datos, los resultados obtenidos con este modelo muestran una notoria imprecisión en la estimación de la cantidad de heridos.

### 5.3 Análisis de componentes principales (PCA)

Para iniciar el análisis de los componentes utilizados respecto de los cuatro atributos originales, en la Figura 5 se presenta la influencia de cada componente. Para ello, se generan cuatro componentes principales que explican la variación de los datos en diferente magnitud. Se puede notar en la Figura 5 que el primer componente principal explica alrededor del 57 % de la variación total del conjunto de datos de las variables tomadas. De manera similar, el segundo componente explica cerca del 30 % de la variación de los datos. Por esta razón, se emplearon ambos componentes para los siguientes modelos de regresión lineal planteados.

**Figura 5**

*Variación explicada por los componentes principales generados*



*Nota.* Resultados obtenidos mediante el procesamiento de datos en MATLAB.

Los modelos planteados en la Tabla 2 logran reducir el número de variables de forma limitada, pues solo aminoran el modelo a 22 variables en el más resumido. De acuerdo con la Tabla 2, el modelo que emplea las dos primeras componentes principales antes mencionadas genera un coeficiente de determinación similar al del modelo más preciso ( $R$  cuadrado =  $0,74 \approx 0,77$ ).

Así como en el caso previo, sobre la base de este modelo se evaluaron los mismos tres registros de accidentes de tránsito extraídos de la base de datos y se obtuvieron los resultados presentados en la Tabla 3, que muestran una estimación cercana a la cantidad de heridos reales; sin embargo, es más imprecisa que el modelo con todas las variables. Como en los demás modelos, la precisión de las estimaciones es congruente con los coeficientes de determinación.

## 6. CONCLUSIONES

Este trabajo presenta un modelo de estimación de la cantidad de heridos que se podrían producir en un accidente de tránsito en la provincia de Lima a través de la implementación de un modelo de regresión lineal múltiple, tomando como base los registros de accidentes ocurridos entre los años 2017 y 2016, de forma que se brinde información importante para el personal de salud que atiende estos eventos, quienes deben priorizar la atención de aquellos accidentes cuyas condiciones permitan maximizar el número de sobrevivientes sobre el total de heridos como producto de la atención oportuna. Así, del análisis desarrollado, se puede concluir que la regresión lineal múltiple permite realizar predicciones adecuadas mediante un modelo con variables explicativas, el cual se aplicó en este trabajo al tratar de predecir el número de heridos mediante las variables propuestas.

Para este fin, se crearon tres distintos modelos de estimación: el primero usó todas las variables independientes; el segundo prescindió de la información sobre el número de personas involucradas en el accidente; y el tercero se desarrolló a través de un análisis de componentes principales de sus variables cuantitativas.

El modelo que mejor explica la cantidad de heridos involucrados en un accidente sería el que se basa en todas las variables. Este hecho se refleja en que la variable independiente “personas involucradas” fue determinante para realizar la estimación de la cantidad de personas heridas de manera adecuada; por esta razón, se recomienda priorizar la atención de accidentes en los que se cuente con el registro de este atributo. Sin embargo, se rescata que, si no se contara con la información del número de personas involucradas, se debería priorizar a los accidentes en función del lugar donde ocurren, como las intersecciones, así como también a aquellos eventos donde se tenga una mayor participación de vehículos menores; aunque vale la pena mencionar que las estimaciones de este modelo (sin el número de personas) poseen una precisión muy inferior con respecto al modelo que incluye la cantidad de personas involucradas.

El análisis de componentes principales permitió reducir de 4 a 2 el número de variables mediante la combinación lineal de otras variables, lo que se toma en cuenta para poder volver más simple un modelo sin perder tanta variabilidad.

Para trabajos futuros, se propone emplear distintos métodos para reducir el número de variables sin pérdida de información relevante, especialmente con referencia al tratamiento de variables categóricas como el análisis de correspondencia. De esta forma se busca generar un modelo más simple, pero considerablemente más preciso.

Asimismo, se puede incluir información relacionada con el número de fallecidos que provoca el accidente como indicador de su gravedad, ya que agregar esta información a los modelos de estimación permitiría lograr una mejor toma de decisiones respecto de la priorización de las atenciones.

## REFERENCIAS

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433-459. <https://doi.org/10.1002/wics.101>
- Amat, J. (2017). *Análisis de componentes principales (Principal Component Analysis, PCA) y t-SNE*. Cienciadedatos.net. [https://www.cienciadedatos.net/documentos/35\\_principal\\_component\\_analysis](https://www.cienciadedatos.net/documentos/35_principal_component_analysis)
- Defensoría del Pueblo. (2021, 20 de diciembre). *Defensoría del Pueblo: más de 14 000 personas fallecieron en accidentes de tránsito en los últimos cinco años*. <https://www.defensoria.gob.pe/defensoria-del-pueblo-mas-de-14-000-personas-fallecieron-en-accidentes-de-transito-en-ultimos-cinco-anos/>
- Estrategia Sanitaria Nacional de Accidentes de Tránsito. (2009). *Accidentes de tránsito: problema de salud pública. Informe nacional*. [http://bvs.minsa.gob.pe/local/MINSA/829\\_MINSA1412.pdf](http://bvs.minsa.gob.pe/local/MINSA/829_MINSA1412.pdf)
- García, D., & Fuente, M. J. (2011). Estudio comparativo de técnicas de detección de fallos basadas en el análisis de componentes principales (PCA). *Revista Iberoamericana de Automática e Informática Industrial*, 8(3), 182-195. <https://doi.org/10.1016/j.riai.2011.06.006>
- Instituto Nacional de Estadística e Informática. (2018). Análisis de los accidentes de tránsito ocurridos en el año 2016. En *Perú: VI Censo Nacional de Comisarías 2017. Resultados definitivos* (pp. 123-147). [https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1528/cap03.pdf](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1528/cap03.pdf)
- Lay, D. C. (2012). *Álgebra lineal y sus aplicaciones* (4.ª ed.). Pearson Educación.
- Maldonado, M. (2016, 25 de noviembre). La hora de oro: 60 minutos que pueden salvar vidas. *Revista Digital INESEM*. <https://www.inesem.es/revistadigital/biosanitario/hora-de-oro/>
- Montero Granados, R. (2016). *Modelos de regresión lineal múltiple. Documentos de trabajo en economía aplicada*. Universidad de Granada. [https://www.ugr.es/~montero/matematicas/regresion\\_lineal.pdf](https://www.ugr.es/~montero/matematicas/regresion_lineal.pdf)
- NHTSA. (2019). *Estimate of motor vehicle traffic crash fatalities for the holiday periods of 2019* [Traffic Safety Facts Research Note. Report N.º DOT HS 812 823]. National Highway Traffic Safety Administration. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812823>
- Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., & Mathers, C. (Eds.). (2004). *Informe mundial sobre prevención de los traumatismos causados por el tránsito*. Organización Mundial de la Salud.

Rubio, J., & Toma, J. (2019). *Estadística aplicada. Segunda parte* (2.ª ed.). Universidad del Pacífico.

Vadillo, F. (2018). *Problemas de mínimos cuadrados*. Universidad del País Vasco. [https://www.ehu.eus/~mepvaarf/ficheros/minimos\\_cuadrados.pdf](https://www.ehu.eus/~mepvaarf/ficheros/minimos_cuadrados.pdf)



# Algoritmos supervisados de *machine learning* para determinar la ubicación de dispositivos wifi

Javier More Sánchez

javier.more@unmsm.edu.pe

<https://orcid.org/0000-0003-1313-6751>

Universidad Nacional Mayor de San Marcos, Perú

Recibido: 6 de agosto del 2022 / Aceptado: 23 de septiembre del 2022

<https://doi.org/10.26439/ciis2022.6074>

**RESUMEN.** El presente artículo tiene como objetivo elegir, bajo un determinado escenario, el mejor algoritmo supervisado de *machine learning* para localizar un terminal que soporte wifi. Se usa un dataset que cuenta con 2000 registros de Received Signal Strength Indicator (RSSI), obtenidos de 7 puntos de acceso (AP), los cuales se cargan en 8 algoritmos supervisados de machine learning. Luego se elige el algoritmo que realiza la predicción más precisa, incluso cuando se cuenta con un menor número de AP. La mayor precisión se logra con el algoritmo *naive Bayes*, tanto para el caso de 7 AP (99 % de precisión) como para cuando se cuenta con un número menor de AP. Asimismo, se observa que los algoritmos basados en redes neuronales presentan el peor rendimiento. Finalmente, se proponen trabajos futuros para continuar con la investigación sobre el tema de localización de dispositivos wifi en interiores.

**PALABRAS CLAVE:** aprendizaje automático, localización en interiores, RSSI, wifi

## SUPERVISED MACHINE LEARNING ALGORITHMS TO DETERMINE THE LOCATION OF WI-FI DEVICES

**ABSTRACT.** This article aims to choose the best-supervised machine learning algorithm for locating a terminal that supports Wi-Fi in a specific scenario. It uses a dataset of 2000 received signal strength indicator (RSSI) records obtained from 7 access points (AP), loaded into eight supervised machine learning algorithms. The algorithm that produces the most accurate prediction is then chosen, even when fewer APs are available. The Naive Bayes algorithm achieved the highest accuracy for the 7AP (99 % accuracy) scenario and for a smaller number of APs. The algorithms based on neural networks had the worst performance. The article proposes future research on the location of Wi-Fi devices indoors.

**KEYWORDS:** indoor location, machine learning, RSSI, Wi-Fi

## 1. INTRODUCCIÓN

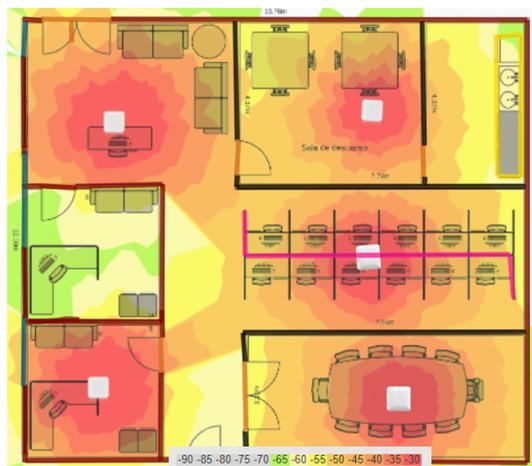
Determinar la ubicación de un objeto o un terminal en una zona de interiores (o exteriores) tiene múltiples aplicaciones, tales como monitoreo de productos, monitoreo de personal, monitoreo de personas en situaciones de emergencia, entre muchas otras. En los últimos años han surgido otras aplicaciones, diferentes del uso de sistemas de posicionamiento satelital, para definir la ubicación de un objeto en interiores usando redes inalámbricas del tipo WPAN (*bluetooth*), WLAN (*wifi*), WMAN (redes móviles 2G/3G/4G/5G, WiMAX) o LPWAN (LoRa WAN, SIGFOX). Dado el amplio uso de las redes wifi, el presente artículo se enfocará en esta tecnología.

Un aspecto clave es el parámetro que se utilizará para realizar la predicción en un escenario basado en redes wifi. La literatura muestra la presencia de dos parámetros relevantes: el RSSI y el CSI (*Channel State Information*). No obstante, como indican Xun et al. (2020), para trabajar con el CSI, se requiere de *hardware* específico y, además, realizar modificaciones en el *firmware* de este. Por tanto, en el presente trabajo se usará el RSSI, el cual puede ser obtenido fácilmente del *smartphone* y de cualquier otro dispositivo que se conecte a una red wifi.

Cuando un terminal (*smartphone*, tableta, *laptop*, *smart TV* o similar) se conecta a una red wifi, estandarizada mediante IEEE 802.11 y sus evoluciones (George & Rao, 2017), específicamente a un punto de acceso (AP), recibe un determinado nivel de intensidad de señal (medido en dBm), denominado RSSI. De acuerdo con los principios de radiopropagación, cuando el terminal se aleja del AP, el RSSI disminuye, y aumenta cuando el terminal se acerca al AP. Si bien el nivel de RSSI en un determinado punto puede ser estimado usando diversas herramientas de predicción (por ejemplo, véase la Figura 1), en la vida real, el valor de RSSI diferirá de la predicción, en la medida en que existen diversos obstáculos que no pueden ser modelados.

Figura 1

Ejemplo de simulación de cobertura de cinco puntos de acceso usando el aplicativo Wi-Fi Designer



Nota. Simulación en la herramienta Wi-Fi Designer, desarrollada por Cambium Networks (<https://wfd.cloud.xirrus.com/wfdc/>).

Para el desarrollo del presente trabajo, se utilizó la herramienta KNIME (*Konstanz Information Miner*), la cual cuenta con licencia Open Source GPLv3, que facilita su libre uso (KNIME, s. f.). Es preciso señalar que esta herramienta ha sido empleada para la elaboración de diversos artículos orientados a *machine learning*, tales como Muenzberg et al. (2019), Çelik y Çinar (2021), Feltrin (2015), y Chauhan y Sehgal (2018). Sin embargo, en la revisión bibliográfica realizada, no se ha encontrado el uso de KNIME para determinar la ubicación de un *smartphone* conectado a la red wifi. Por tanto, dada la facilidad de uso de KNIME, los interesados en aprender sobre algoritmos de *machine learning* y que, además, cuenten con algunos conocimientos de telecomunicaciones podrán tomar el presente artículo como un punto de partida para una mejor comprensión del comportamiento de los algoritmos.

Para tal efecto, aprovechando la versatilidad de KNIME, se propone usar y comparar la precisión de diversos algoritmos de *machine learning* supervisados, para determinar la ubicación de un terminal conectado a la red wifi. En específico, se usarán ocho algoritmos: *decisión tree* (DT), *naive Bayes* (NB), *k-nearest neighbors* (k-NN), *random forest* (RF), *support vector machine* (SVM), *gradient boosted* (GB), así como algoritmos basados en redes neuronales: RProp MLP (*multilayer perceptron*) y PNN (*probabilistic neural network*). Con el fin de profundizar en los algoritmos utilizados, se sugiere revisar KNIME (2020).

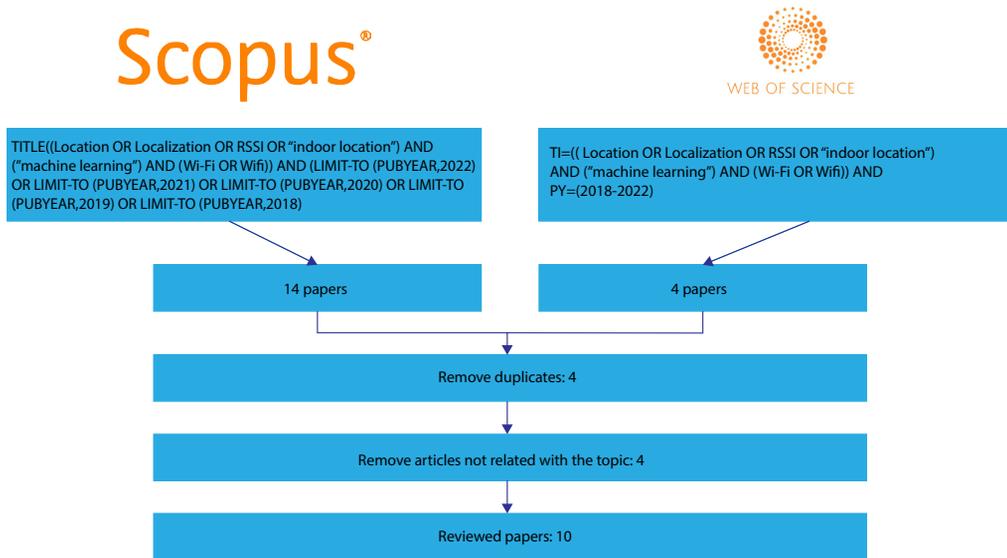
El presente artículo parte de describir los trabajos realizados sobre el tema en los últimos cinco años. Luego se presenta la metodología utilizada, partiendo de la obtención de los datos, seguido por el uso de los algoritmos, hasta implementar el modelo en KNIME. Finalmente, se describen los resultados, así como las conclusiones y las recomendaciones para los trabajos futuros.

## 2. ESTADO DEL ARTE

Se han encontrado diversos artículos que tienen como objetivo determinar la ubicación de un objeto (que sea compatible con wifi) usando la información de diversos puntos de acceso (AP), a los que dicho objeto tiene alcance. En efecto, Bellavista-Parent et al. (2021) contabilizaron y revisaron un total de 64 artículos entre el 2019 y el 2021, que usan diversos datasets y que de una u otra forma están orientados a determinar el posicionamiento *indoor* utilizando las redes wifi y que a la vez hacen uso de algoritmos de *machine learning*.

No obstante, para el presente artículo se realizó una nueva búsqueda en los portales especializados Scopus y Web of Science (WoS), tomando como punto de partida los títulos de los trabajos (dejando de lado el contenido del *abstract* o de las palabras clave). Se encontró un total de 18 artículos cuyos títulos guardan relación con la búsqueda; sin embargo, finalmente se seleccionaron 10 artículos que guardan estricta relación con posicionamiento vía wifi usando algoritmos de *machine learning* y que han sido publicados en los últimos cinco años (véase la Figura 2).

Figura 2  
 Artículos revisados en Scopus y Web of Science



Para determinar la ubicación de un terminal en cuatro habitaciones, Sabanci et al. (2018) compararon seis algoritmos utilizando redes neuronales artificiales (ANN), k-NN, DT, NB, *extreme learning machine* (ELM) y SVM, y encontraron que el algoritmo k-NN presenta el mejor rendimiento. Por su lado, Tahat et al. (2021) usaron un dataset que contiene información en las bandas de 2,4 GHz y 5 GHz. Emplearon los algoritmos k-NN, SVM, DT y ANN, y hallaron que el algoritmo basado en redes neuronales presenta el mejor resultado (no obstante, señalan que es necesario un mayor tiempo de entrenamiento).

Se han encontrado investigaciones que usan el dataset UJIndoorLOc, el cual contiene mediciones en tres edificios con diversos pisos (3 a 5). En el estudio de Wadhwa et al. (2019), para la detección del edificio se tiene una precisión del 100 % usando k-NN, mientras que para la predicción del piso se logran precisiones de hasta 97,98 % usando el algoritmo RF. Asimismo, en el estudio de Cao et al. (2021), se emplean los algoritmos *eXtreme gradient boosting* (XGBoost) y *gradient boosting decision tree* (GBDT), y se encontró que para la detección del piso se tienen precisiones de entre 99,54 % a 99,22 %. Por su parte, Maaloul et al. (2022) usan los algoritmos NB, SVM, k-NN, RF y GB, entre los cuales el algoritmo GB presenta la mayor precisión.

Un nuevo enfoque se presenta en el estudio de Xue et al. (2020), al que denominan *localización indoor de alta adaptabilidad* (HAIL), para lo cual se diseña una red neuronal de

retropropagación (BPNN). Los autores indican que queda como trabajo pendiente continuar investigando nuevos algoritmos para mejorar la precisión.

Finalmente, Singh et al. (2021) realizaron una revisión de los artículos publicados entre los años 2007 al 2020 sobre el uso de algoritmos de *machine learning* para determinar la ubicación *indoor*. Asimismo, considerando que se vive en un mundo conectado, se plantean cuestiones sobre la privacidad en la recolección de información.

**Tabla 1**

*Aspectos clave de los artículos revisados*

Título	Año	Puntos clave
A Universal Wi-Fi Fingerprint Localization Method Based on Machine Learning and Sample Differences	2021	Uso de algoritmos XGBoost y GBDT. En la detección del piso se logra una precisión de entre 99,54 % a 99,22 %.
An Empirical Evaluation of Machine Learning Algorithms for Indoor Localization Using Dual-Band WiFi	2021	Uso de banda de 2,4 y 5 GHz. Los algoritmos basados en ANN presentan los mejores resultados.
Experimental Machine Learning for RSSI Fingerprint in Indoor WiFi Localization	2021	Uso de tres <i>routers</i> en una habitación.
Using Machine Learning Techniques and Wi-Fi Signal Strength for Determining Indoor User Location	2021	Uso de los algoritmos KNN y SVM. Para el caso de KNN se logra una precisión de 93 %.
Machine Learning for Location Prediction Using RSSI on Wi-Fi 2.4 GHz Frequency Band	2021	Se logra una precisión de 96 %.
Machine Learning Based Indoor Localization Using Wi-Fi RSSI Fingerprints: An Overview	2021	Realiza una revisión de la literatura desde el 2007 al 2020. Plantea cuestiones sobre la privacidad en la recolección de información.
A WiFi Fingerprint Based High-adaptability Indoor Localization Via Machine Learning	2020	Uso de algoritmo HAIL basado en redes neuronales.
Machine Learning Based Indoor Localization Using Wi-Fi Fingerprinting	2019	Base de datos UJIndoorLOc: tres edificios. Uso de dos algoritmos: k-NN y RF.
WiFi Based Indoor Localization: Application and Comparison of Machine Learning Algorithms	2018	Base de datos de UCI: cuatro habitaciones. Uso de seis algoritmos. Mejor precisión: k-NN (98,75 %).

Así, con base en las investigaciones previas, el aporte del presente trabajo de investigación es evaluar el comportamiento de los distintos algoritmos ante cambios en las condiciones del

experimento (reducción del número de AP utilizados). Así se determinará que, pese a que algunos AP dejen de funcionar, el algoritmo elegido siempre será el que presente la mayor precisión. Además, se plantean trabajos futuros para continuar investigando sobre el tema.

### 3. METODOLOGÍA

Para la comparación de los algoritmos, se usó el dataset denominado *Wireless Indoor Localization*, disponible en la página web de la Universidad de California en Irvine (UCI, 2017). El dataset contiene 2000 mediciones de intensidad de señal de recepción en un *smartphone*, distribuidas en 4 habitaciones con 500 muestras cada una. Para ello, el *smartphone* se desplazó por 4 habitaciones, recogiendo las mediciones del RSSI de 7 AP. Cabe señalar que no se tiene mayor detalle del dataset, por ejemplo, banda de frecuencia, modelo del terminal móvil, canales utilizados, entre otros. Para más información, un extracto del dataset se presenta en la Tabla 2.

**Tabla 2**

*Extracto del dataset utilizado*

AP_01	AP_02	AP_03	AP_04	AP_05	AP_06	AP_07	Room
-64	-56	-61	-66	-71	-82	-81	Room_1
-68	-57	-61	-65	-71	-85	-85	Room_1
-63	-60	-60	-67	-76	-85	-84	Room_1
-34	-53	-52	-19	-65	-75	-69	Room_2

*Nota.* Reproducido de *Wireless Indoor Localization*, by Rajen Bhatt [Dataset], por Universidad de California en Irvine, 2017, Machine Learning Repository (<https://archive-beta.ics.uci.edu/ml/datasets/wireless+indoor+localization>).

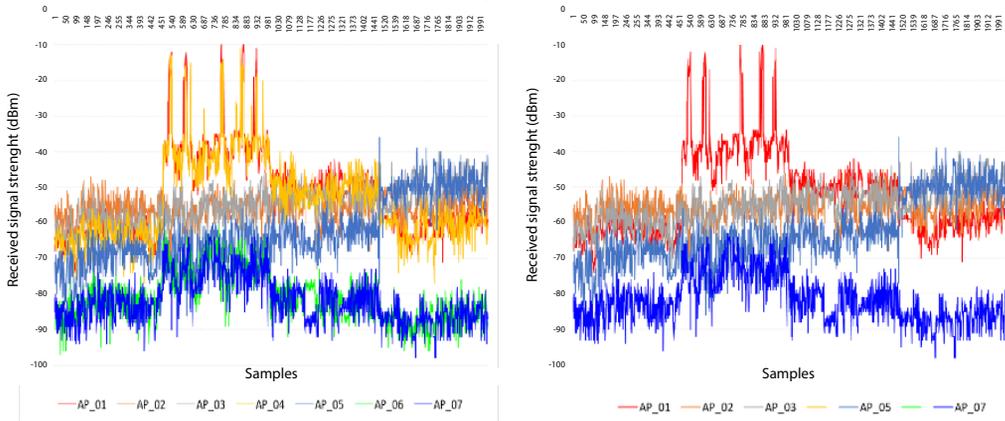
Es preciso señalar que, según el autor del dataset, la información se utilizó en Jayant et al. (2017) aplicando *Fuzzy Hybrid of Particle Swarm Optimization & Gravitational Search Algorithm* (FPSOGSA); no obstante, al revisar dicho artículo, se observa que solo usa tres habitaciones. Asimismo, tal como se señaló previamente, Sabanci et al. (2018) también usan este dataset, entrenando la data con seis algoritmos de *machine learning*.

El dataset, que contiene mediciones de diversos AP, se dividirá en tres tipos. En primer lugar, se usará el dataset original (con 7 AP), luego con 6 AP y, finalmente, solo considerando 5 AP. Esto se realiza con el fin de validar el comportamiento del algoritmo ante fallas de un determinado AP. En la Figura 3, se observa que los niveles de RSSI oscilan entre  $-10$  dBm (nivel de señal muy bueno) y  $-98$  dBm (nivel de señal muy malo). Luego la información de cada uno de ellos se normalizará en el rango de 0 a 100, con el fin de poder utilizar todos los algoritmos indicados en la sección previa. Después, se particionará en una proporción de 70 %

a 30 %; el 70 % servirá para realizar el entrenamiento del algoritmo y el 30 % para calcular la precisión del algoritmo. Esta información se ingresará a los bloques que contienen los diversos algoritmos y se elegirá el que arroje como resultado el menor error (mayor precisión).

**Figura 3**

*Niveles de intensidad de señal recibidos en el smartphone usando datos de 7 y 5 AP*

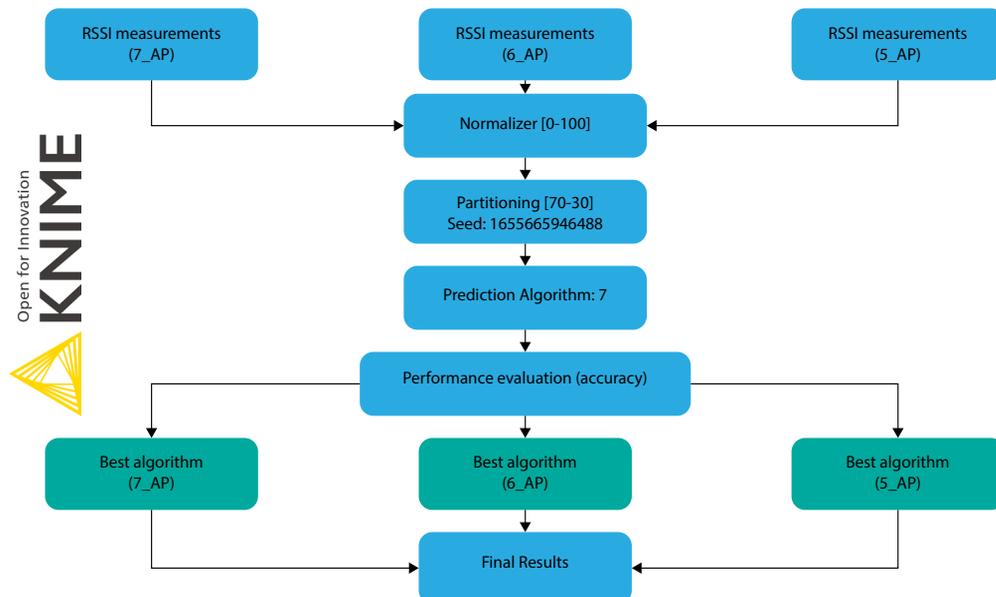


*Nota.* Elaboración propia usando el dataset de UCI (2017).

En la Figura 4, se muestra en forma de esquema el modelo utilizado en el presente artículo, partiendo desde la formación de tres grupos de mediciones (datos de ingreso del modelo) hasta llegar al algoritmo que presente el menor error de predicción (datos de salida del modelo).

**Figura 4**

*Modelo utilizado para evaluar la precisión de los algoritmos*

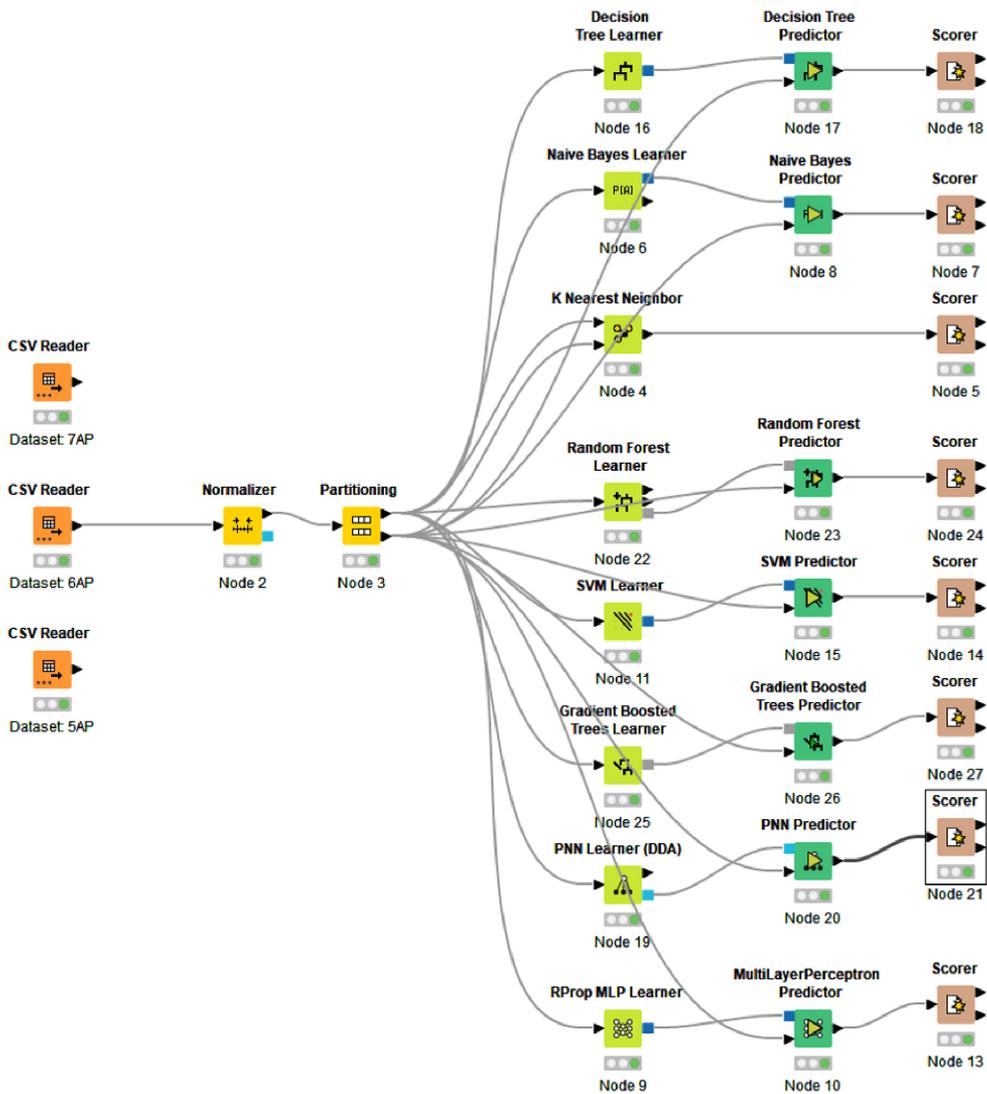


#### 4. IMPLEMENTACIÓN DE LOS ALGORITMOS DE MACHINE LEARNING EN KNIME

Los tres datasets se cargan de manera independiente en el flujo de trabajo (*workflow*) elaborado en KNIME (véase la Figura 5). KNIME trabaja con bloques, en los cuales cada uno cumple una determinada función. El dataset ingresa al flujo de trabajo por medio del módulo “CSV Reader” y los resultados se obtienen en el bloque “Scorer”.

Figura 5

*Flujo de trabajo implementado en KNIME con ocho algoritmos de machine learning*



Es importante resaltar que, previo a realizar el particionamiento, el dataset debe ser normalizado (en este caso, de 0 a 100). Luego se realiza el particionado (70/30) de tal forma que cada habitación tenga la misma proporción del número de muestras. Asimismo, es preciso resaltar que, con el fin de que los resultados puedan ser replicados, en el bloque de particionado de KNIME se usó la siguiente semilla: 1655665946488.

El *workflow* contempla el uso de ocho algoritmos no supervisados aplicados a un problema de clasificación, en específico, para detectar la habitación en la que se encuentra un terminal móvil. Así se implementan los algoritmos DT, NB, k-NN, RF, SVM, GB, así como algoritmos basados en redes neuronales artificiales (ANN), en particular RProp MLP y PNN.

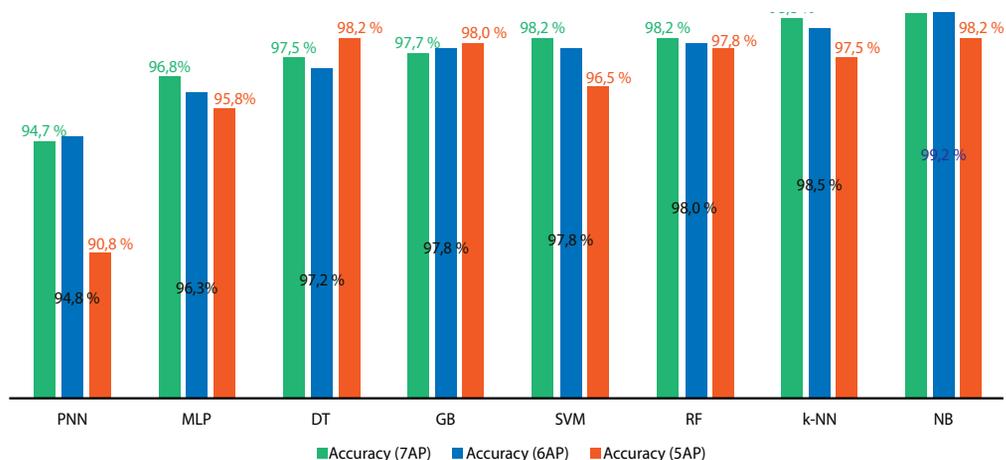
Dada la facilidad de uso de KNIME, cada algoritmo de *machine learning* se carga como un bloque independiente. Posteriormente, la precisión de cada algoritmo se obtiene en el módulo “Scorer” de KNIME, con base en la matriz de confusión. Para el caso del algoritmo MLP, se usan los siguientes parámetros: (i) número máximo de iteraciones: 1000; (ii) número de capas ocultas: 2; y (iii) número de neuronas ocultas por capa: 10. Asimismo, para el algoritmo PNN se usan 2000 épocas. En el caso de los demás algoritmos, se emplea la configuración predeterminada de cada módulo.

## 5. RESULTADOS

Se observa que, usando la herramienta KNIME, el algoritmo *naive Bayes* presenta la mejor precisión: 99,00 % usando 7 AP y 98,167 % usando 5 AP. En segundo lugar, queda el algoritmo *k-nearest neighbors*. También se observa que los algoritmos basados en ANN, tales como el MLP y el PNN, presentan la precisión más baja. En efecto, con MPL se logra una precisión de 96,833 % y para PNN tan solo se llega a 94,667 % (para el caso de 7 AP). Todos los resultados de las diversas simulaciones se visualizan en la Figura 6.

Figura 6

Resultados de la simulación en KNIME



Es preciso indicar que el presente análisis de clasificación tiene la particularidad de ser multiclase, es decir que la variable que se predice no es binaria (por ejemplo, SÍ/NO o 1/0), sino que más bien se trata de cuatro valores (Room\_1, Room\_2, Room\_3 y Room\_4).

## 6. CONCLUSIONES, TRABAJOS FUTUROS Y LIMITACIONES

El algoritmo *naive Bayes* presenta la precisión más alta, pues llega hasta un 99 % de precisión, seguido por el algoritmo k-NN con 98,833 % (para el caso de 7 AP). Asimismo, cuando el sistema es forzado a trabajar con un número menor de AP, se tiene una precisión del 98,167 % para NB. Además, se observa que los algoritmos basados en ANN (en específico, MLP y PNN) presentan el peor rendimiento. En efecto, para el caso de 7 AP, con MLP se tiene una precisión del 96,833 % y con PNN solo se llega a 94,667 %.

Diversos trabajos futuros se desprenden de la presente investigación. Por el lado de las redes wifi, resulta necesario crear un dataset propio que contenga mediciones de todas las bandas de espectro utilizadas para redes wifi: 2,4 GHz, 5 GHz y 6 GHz, de tal forma que se pueda determinar cuál es la banda que permite tener una mejor predicción. Asimismo, por el lado de las redes móviles, también se puede hacer pruebas georreferenciadas (con información de latitud y longitud) de *drive test* o *walk test* para identificar los Cell\_ID vecinos, y sobre esa base predecir la ubicación de un terminal móvil usando algoritmos de *machine learning*.

Finalmente, los resultados se deben tomar con mucha precaución en la medida que cada ambiente interior es único y, por tanto, para que el sistema de localización funcione correctamente, se requiere que previamente se haya realizado una adecuada recolección de datos. Es decir, para otro escenario, los resultados podrían variar, lo cual evidentemente es una restricción de los trabajos que usan algoritmos para localizar dispositivos sin usar sistemas de posicionamiento satelital.

## REFERENCIAS

- AlQahtani, A. A. S., & Choudhury, N. (2021). Machine learning for location prediction using rssi on Wi-Fi 2.4 GHz frequency band. En *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 0336-0342). DOI: 10.1109/IEMCON53756.2021.9623104
- Bellavista-Parent, V., Torres-Sospedra, J., & Perez-Navarro, A. (2021). New trends in indoor positioning based on WiFi and machine learning: A systematic review. En *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (pp. 1-8). <https://doi.org/10.1109/IPIN51156.2021.9662521>

- Cao, X., Zhuang, Y., Yang, X., Sun, X., & Wang, X. (2021). A universal Wi-Fi fingerprint localization method based on machine learning and sample differences. *Satellite Navigation*, 2, 27. <https://doi.org/10.1186/s43020-021-00058-8>
- Çelik, H., & Çınar, A. (2021). An application on ensemble learning using KNIME. En *2021 International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 400-403). <https://doi.org/10.1109/ICDABI53623.2021.9655815>
- Chauhan, C., & Sehgal, S. (2018). Sentiment classification for mobile reviews using KNIME. En *2018 International Conference on Computing, Power and Communication Technologies (GUCON)* (pp. 548-553). <https://doi.org/10.1109/GUCON.2018.8674946>
- Feltrin, L. (2015). KNIME an open source solution for predictive analytics in the geosciences [Software and datasets]. *IEEE Geoscience and Remote Sensing Magazine*, 3(4), 28-38. <https://doi.org/10.1109/MGRS.2015.2496160>
- George, D., & Rao, S. (2017). Enabling rural connectivity: long range Wi-Fi versus super Wi-Fi. En *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)* (pp. 1-4). <https://doi.org/10.1109/ICIC.2017.8524142>
- Insany, G. P., Ayu, M. A., & Mantoro, T. (2021). Using machine learning techniques and Wi-Fi signal strength for determining indoor user location. En *2021 IEEE 7th International Conference on Computing, Engineering and Design (ICCED)*. doi: 10.1109/ICCED53389.2021.9664859
- Jayant G., R., Perumal, B., Narayanan, S., Thakur, P., & Bhatt, R. (2017). User localization in an indoor environment using Fuzzy Hybrid of Particle Swarm Optimization & Gravitational Search Algorithm with Neural Networks. En *Proceedings of Sixth International Conference on Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing* (vol. 546). [https://doi.org/10.1007/978-981-10-3322-3\\_27](https://doi.org/10.1007/978-981-10-3322-3_27)
- KNIME. (s. f.). KNIME Open Source Story. <https://www.knime.com/knime-open-source-story>
- KNIME. (2020, 25 de junio). Classification\_and\_Predictive\_Modelling. [https://hub.knime.com/knime/spaces/Examples/latest/04\\_Analytics/04\\_Classification\\_and\\_Predictive\\_Modelling](https://hub.knime.com/knime/spaces/Examples/latest/04_Analytics/04_Classification_and_Predictive_Modelling)
- Koovimol, P., & Pattaramalai, S. (2021). Experimental machine learning for RSSI fingerprint in indoor WiFi localization. En *2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 1018-1021). DOI: 10.1109/ECTI-CON51831.2021.9454865
- Maaloul, K., Abdelhamid, N. M., & Lejdel, B. (2022). Machine learning based indoor localization using Wi-Fi and smartphone in a shopping malls. En B. Lejdel, E.

- Clementini & L. Alarabi (Eds.), *Artificial Intelligence and Its Applications. AIAP 2021. Lecture Notes in Networks and Systems* (vol. 413). [https://doi.org/10.1007/978-3-030-96311-8\\_1](https://doi.org/10.1007/978-3-030-96311-8_1)
- Muenzberg, A., Sauer, J., Hein, A., & Roesch, N. (2019). Checking the plausibility of nutrient data in food datasets using KNIME and big data. En *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. <https://doi.org/10.1109/WiMOB.2019.8923233>
- Sabancı, K., Yigit, E., Ustun, D., Toktas, A., & Aslan, M. (2018). WiFi based indoor localization: Application and comparison of machine learning algorithms. En *2018 XXIIIrd International Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory (DIPED)* (pp. 246-251). <https://doi.org/10.1109/DIPED.2018.8543125>
- Singh, N., Choe, S., & Punmiya, R. (2021). Machine learning based indoor localization using Wi-Fi RSSI fingerprints: An overview. *IEEE Access*, 9, 127150-127174. <https://doi.org/10.1109/ACCESS.2021.3111083>
- Tahat, A., Awwad, R., Baydoun, N., Al-Nabih, S., & Edwan, T. A. (2021). An empirical evaluation of machine learning algorithms for indoor localization using dual-band WiFi. En *2021 2nd European Symposium on Software Engineering* (pp. 106-111). <https://doi.org/10.1145/3501774.3501790>
- Universidad de California en Irvine. (2017, 4 de diciembre). *Wireless Indoor Localization, by Rajen Bhatt* [Dataset]. Machine Learning Repository. <https://archive-beta.ics.uci.edu/ml/datasets/wireless+indoor+localization>
- Wadhwa, S., Rai, P., & Kaushik, R. (2019). Machine learning based indoor localization using wi-fi fingerprinting. *International Journal of Recent Technology and Engineering*, 8(3), 502-506. <https://doi.10.35940/ijrte.A2133.098319>
- Xue, J., Liu, J., Sheng, M., Shi, Y., & Li, J. (2020). A WiFi fingerprint based high-adaptability indoor localization via machine learning. *China Communications*, 17(7), 247-259. <https://doi.org/10.23919/J.CC.2020.07.018>
- Xun, W., Sun, L., Han, C., Lin, Z., & Guo, J. (2020). Depthwise separable convolution based passive indoor localization using CSI Fingerprint. En *2020 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1-6). <https://doi.org/10.1109/WCNC45663.2020.9120638>

# PÓSTERES

# MOOC para la educación continua en sistemas de información durante la pandemia

Diana Concepción Mex Álvarez  
diancmex@uacam.mx / <http://orcid.org/0000-0001-9419-7868>

Margarita Castillo Téllez  
mcastill@uacam.mx / <https://orcid.org/0000-0001-9639-1736>

Francisco Emmanuel Alfaro Juárez  
al059109@uacam.mx / <https://orcid.org/0000-0003-0253-2249>  
Universidad Autónoma de Campeche, México

Recibido: 28 de julio del 2022 / Aceptado: 9 de octubre del 2022  
doi: <https://doi.org/10.26439/ciis2022.6075>

**RESUMEN.** La educación continua de la comunidad de la Universidad Autónoma de Campeche durante la pandemia del COVID-19 se realizó a través del programa Coursera for Campus. En este trabajo se presenta la metodología y el análisis de los resultados obtenidos de la comunidad universitaria que participó en el programa de la plataforma en el área de sistemas de información.

**PALABRAS CLAVE:** educación, computación, MOOC

## MOOCS FOR CONTINUING EDUCATION IN INFORMATION SYSTEMS DURING THE PANDEMIC

**ABSTRACT.** The continuing education of the Universidad Autónoma de Campeche community during the COVID-19 pandemic was carried out through the Coursera for Campus program. This paper presents the methodology and analysis of the results obtained from the university community that participated in the platform program in the area of Information Systems.

**KEYWORDS:** education, computing, MOOCs

## 1. INTRODUCCIÓN

La transformación digital era una necesidad y un requerimiento con respecto al avance del conocimiento y de la sociedad (García-Peñalvo, 2020).

Todas las instituciones educativas, a raíz de la pandemia del COVID-19, implementaron estrategias de uso de las herramientas digitales. Tal es el caso de los cursos abiertos en línea, y entre ellos, el empleo de los cursos abiertos masivos en línea, mejor conocidos como *massive open online courses* (MOOC, por sus siglas en inglés). Este tipo de cursos ofrece un aprendizaje de carácter no lineal y asíncrono; es decir, es posible para el estudiante aprender a su propio ritmo, debido a que no se ubica en un espacio físico (Universidad Autónoma de Campeche [UAC], 2020; Martin et al., 2018). Para la UAC, el desarrollo de contenidos propios para crear MOOC de educación continua era imposible, ya que no cuenta con un departamento para este fin.

Afortunadamente, existen organizaciones que ofrecen MOOC en diferentes áreas del conocimiento. La UAC optó por Coursera, una plataforma educativa establecida en el año 2011 que está asociada con prestigiosas universidades y organizaciones de todo el mundo (Suárez, 2012). Entre sus principales ventajas, están los certificados avalados por las universidades asociadas y profesionales con credenciales prestigiadas, así como el nivel avanzado que ofrece en las diversas temáticas. Sin embargo, su principal desventaja ante otras plataformas es su elevado precio; por ello, la UAC buscó el beneficio de cursos y certificados gratuitos para su comunidad, al incorporarse en el programa Coursera for Campus, derivado de la contingencia por el COVID-19.

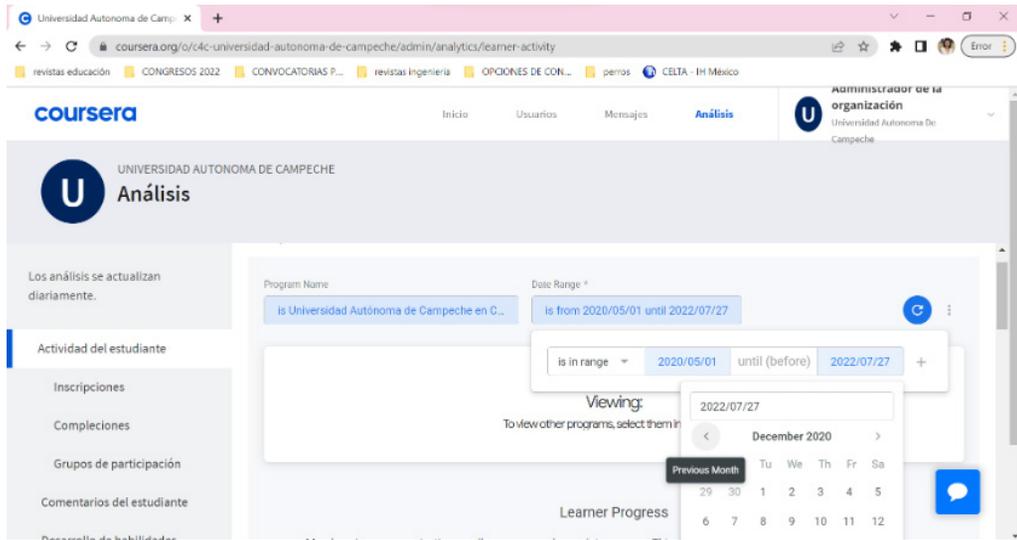
Coursera basa el aprendizaje en cuatro ideas clave: la eficacia del aprendizaje en línea, el aprendizaje para el dominio, la evaluación entre compañeros y el aprendizaje mixto (Polanco Martínez, 2018).

## 2. DESCRIPCIÓN GENERAL DEL TRABAJO

En el presente trabajo nos enfocaremos en los resultados obtenidos de la incorporación de la Universidad Autónoma de Campeche, México, en el programa Coursera for Campus. Para ello se presenta la metodología empleada.

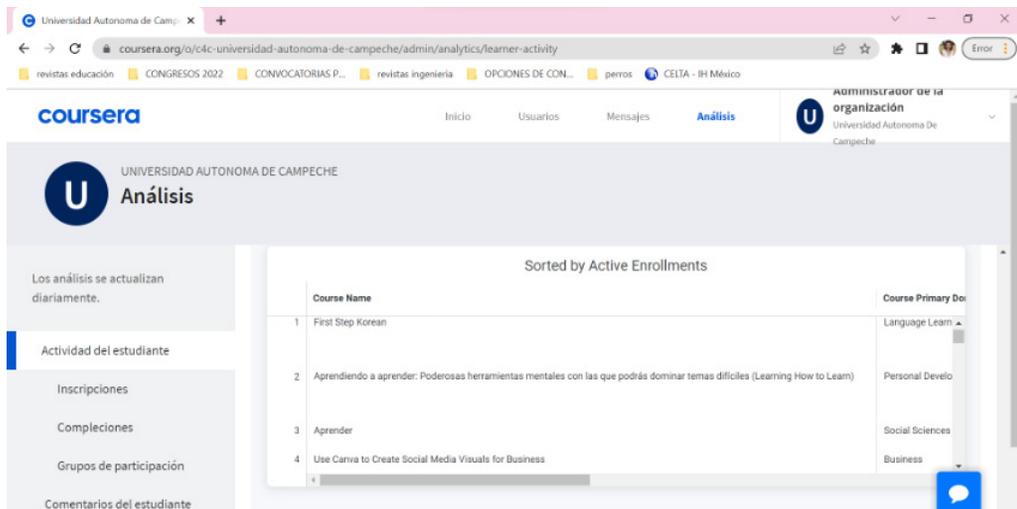
Posterior al ingreso de la plataforma, se elige la sección de Análisis y se genera el filtro para obtener la actividad de los estudiantes del 1 mayo al 31 de diciembre del 2020, tal como se muestra en la Figura 1.

Figura 1  
Sección de Análisis de la plataforma



Una vez filtrado por fecha, se muestra una tabla de récords con el listado de cursos, su área temática, la fecha de la primera inscripción, la fecha de la última inscripción, el número de inscripciones, el número de inscripciones activas, el número de participantes que completaron cursos y el porcentaje de participantes que completaron cursos (véase la Figura 2).

Figura 2  
Tabla de récords



En la tabla de récords, se da clic sobre el título de la columna de área temática para que sea ordenada de manera alfabética y, posteriormente, se descargan los resultados oprimiendo el ícono de tres puntos y seleccionando descargar en formato Excel.

Una vez descargado el formato Excel, abrimos el archivo y se nos mostrará una tabla con toda la información de los cursos. Podemos utilizar el filtro para seleccionar cursos de alguna de las áreas de conocimiento, como se muestra en la Figura 3.

Figura 3

*Filtrado en Excel*

First enrolment	Most recent enrolment	Enrolment	Active enrolment	Completion	Completion%
13/05/2020	31/06/2020	18	18	0	33.30%
12/05/2020	21/12/2020	21	18	5	20.80%
13/05/2020	22/09/2020	18	14	8	44.40%
12/05/2020	27/09/2020	15	10	7	46.70%
07/05/2020	30/09/2020	13	10	1	7.70%
05/05/2020	10/09/2020	12	10	1	8.30%
13/05/2020	15/08/2020	12	10	6	50.00%
12/05/2020	15/09/2020	11	9	4	36.40%
14/05/2020	19/09/2020	10	9	1	10.00%
05/05/2020	17/09/2020	9	9	7	77.80%

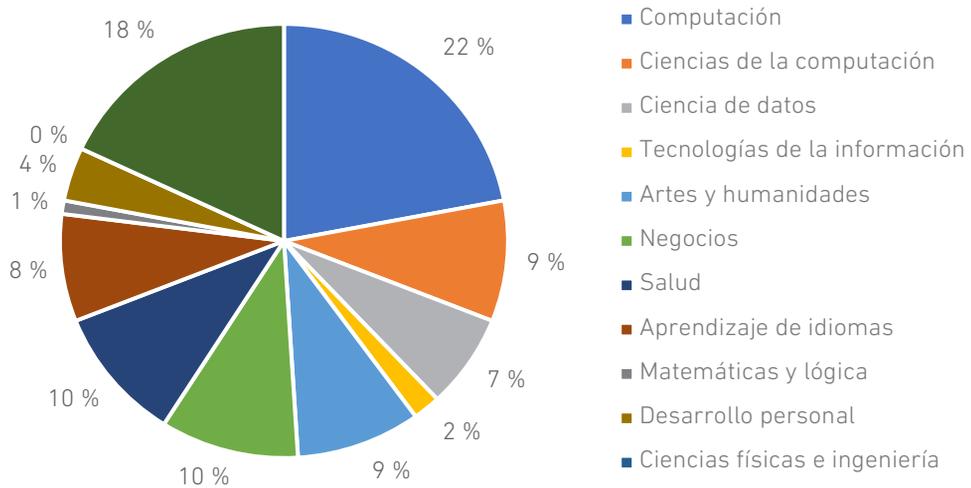
El procesamiento estadístico de los datos fue realizado a través de Microsoft Excel, con la ayuda de las herramientas que proporciona para los estudios de tipo descriptivo.

### 3. CONTRIBUCIONES Y TRABAJO FUTURO

Del total de 404 cursos impartidos, 74 fueron sobre ciencia de los datos, tecnologías de la información, ciencias de la computación y computación. Lo anterior significa que más del 18 % de los cursos de interés fueron de temas relativos a la computación y las tecnologías. En la Figura 4 se puede apreciar la distribución de los cursos por área.

**Figura 4**

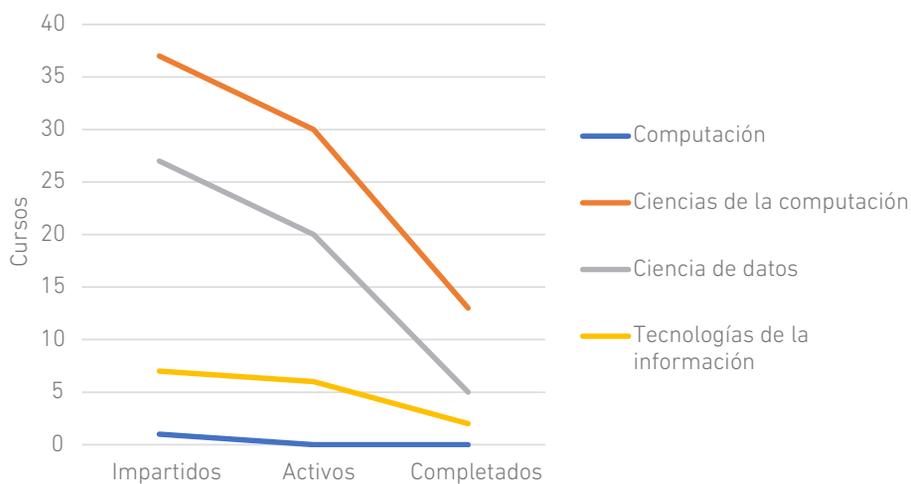
*Porcentaje de cursos impartidos en total*



Sin embargo, es de gran relevancia que solo 20 de los 74 cursos fueron culminados en su totalidad. En la Figura 5, se observa 1 curso solicitado sobre computación, 37 sobre ciencias de la computación, 27 sobre ciencia de los datos y 7 de tecnologías de la información. La gráfica decrece respecto a la actividad de los cursos hasta llegar a 0 completados en el área de computación, 13 en ciencias de la computación, 5 en ciencia de los datos y 2 en tecnologías de la información. De ahí que se manifieste que los MOOC en las ciencias de la computación y tecnologías de la información requieren de estrategias complementarias que aseguren su culminación con éxito.

**Figura 5**

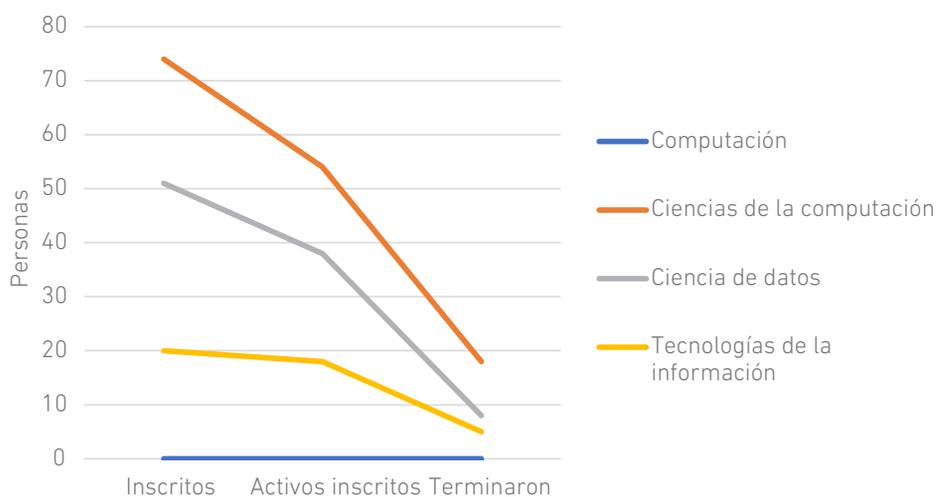
*Estado de los cursos*



En la Figura 6, se presenta el comportamiento de las 146 personas que estuvieron interesadas en los 74 cursos durante el periodo evaluado. Solo 32 terminaron exitosamente.

**Figura 6**

*Comportamiento de las personas participantes*



Para mejorar la eficiencia terminal de los MOOC referentes a la ciencia de los datos, tecnologías de la información, ciencias de la computación y computación, se recomienda realizar un estudio en las personas que participaron para conocer las causas de su deserción y, en el caso de los que culminaron exitosamente, cuáles fueron los factores de éxito.

## REFERENCIAS

- García-Peñalvo, F. J. (2020). El sistema universitario ante la COVID-19: corto, medio y largo plazo. *Universidad*. <https://www.universidadsi.es/sistema-universitario-covid-19/>
- Martin, N., Kelly, N., & Terry, P. (2018). A framework for self-determination in massive open online courses: Design for autonomy, competence, and relatedness. *Australasian Journal of Educational Technology*, 34(2). <https://doi.org/10.14742/ajet.3722>
- Suárez, D. (2012). iTunesU y Coursera: pedagogía universitaria 2.0. *Cuaderno de Pedagogía Universitaria*, 9(18), 6-8. <http://hdl.handle.net/20.500.12060/639>
- Polanco Martínez, G. G., Holguín, I. G., & Vázquez, J. A. (2018). Estudio comparativo entre estudiantes y docentes del Instituto Tecnológico de Chihuahua en la utilización de los recursos *online* masivos y abiertos (MOOC). *Revista de la Facultad de Contaduría y Administración de la Universidad Autónoma de Chihuahua*, 15(46), 87-105. <http://fca.uach.mx/institucionales/2020/02/17/excelencia-administrativa-vol-46.pdf#page=89>
- Universidad Autónoma de Campeche. (2020, 28 de agosto). *Políticas de operación durante la enseñanza remota para continuidad del servicio académico para el ciclo escolar 2020-2021*.

# MOOC para la educación continua en sistemas de información durante la pandemia

Mtra. Diana C. Mex Álvarez, Dra. Margarita Castillo Téllez, Br. Francisco Emmanuel Alfaro Juárez  
diancmex@uacam.mx, mcastill@uacam.mx, al059109@uacam.mx

La educación continua de la comunidad de la Universidad Autónoma de Campeche durante la pandemia del COVID-19 se realizó a través del programa Coursera for Campus. En este trabajo se presenta la metodología y el análisis de los resultados obtenidos de la comunidad universitaria que participó en el programa de la plataforma en el área de sistemas de información.

## Introducción

- La transformación digital era una necesidad y un requerimiento con respecto al avance del conocimiento y de la sociedad [1].
- Todas las instituciones educativas a raíz de la pandemia por el COVID-19 han implementado estrategias donde han hecho uso de las herramientas digitales, entre ellas el empleo de los cursos abiertos masivos en línea o mejor conocidos como MOOC (*Massive Open Online Courses*, por sus siglas en inglés). Este tipo de cursos ofrecen un aprendizaje que se caracteriza por su carácter no lineal y asíncrono, es decir, es posible para el estudiante aprender a su propio ritmo, debido a la facilidad por la no existencia de un espacio físico [2].
- Para la UAC el desarrollo de contenidos propios para crear MOOC de educación continua era imposible, ya que no cuenta con un departamento para este fin. La institución optó por Coursera, que es una plataforma educativa establecida en el año 2011, asociada con prestigiosas universidades y organizaciones de todo el mundo [3].
- Entre las principales ventajas de Coursera están los certificados avalados por las universidades asociadas y profesionales con credenciales prestigiosas, así como el nivel avanzado que ofrece en las diversas temáticas. Sin embargo, la principal desventaja de Coursera ante otras plataformas es el precio elevado, por ello, la UAC buscó el beneficio de cursos y certificados gratuitos para su comunidad, al incorporarse en el programa Coursera for Campus derivado de la contingencia por el COVID-19.

## Método

A continuación presentamos los resultados obtenidos de la incorporación de la Universidad Autónoma de Campeche en el programa Coursera for Campus.

Para ello se presenta la metodología empleada:

Posterior al ingreso de la plataforma, se elige la sección de Análisis y se genera el filtro para obtener la actividad de los estudiantes del 1 mayo al 31 de diciembre del 2020, tal como se muestra en la Figura 1.



Figura 1. Sección de Análisis

Una vez filtrado por fecha, se genera una tabla de récords como se muestra en la Figura 2 con el listado de cursos, su área temática, la fecha de la primera inscripción, la fecha de la última inscripción, el número de inscripciones, número de inscripciones activas, número de participantes que completaron cursos y el porcentaje de participantes que completaron cursos.

Figura 2. Tabla de récords

En la tabla de récords se da clic sobre el título de la columna de área temática para que sea ordenada alfabéticamente y posteriormente descargar los resultados oprimiendo el icono de tres puntos y seleccionando descargar en formato Excel.

Una vez descargado el formato Excel, abrimos el archivo y se nos mostrará una tabla con toda la información de los cursos, podemos utilizar el filtro para seleccionar cursos de alguna de las áreas de conocimiento como se muestra en la Figura 3.

Figura 3. Filtrado en Excel

El procesamiento estadístico de los datos fue realizado a través de Microsoft Excel, con la ayuda de las herramientas que proporciona para los estudios de tipo descriptivo.

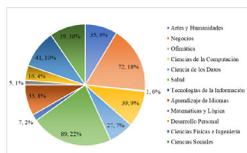


Figura 4. Porcentaje de cursos impartidos en total

Sin embargo, es de gran relevancia que solo 20 de los 74 fueron culminados en su totalidad.

En la Figura 5, se observa 1 curso solicitado sobre computación, 39 sobre ciencias de la computación, 27 sobre ciencia de los datos y 7 de tecnologías de la información. Posteriormente se observa cómo la gráfica decrece respecto a la actividad de los cursos, hasta llegar a 0

completados en el área de computación, 13 en ciencias de la computación, 5 en ciencia de los datos y 2 en tecnologías de la información. De ahí que se manifieste que los MOOC en las ciencias de la computación y tecnologías de la información requieran de estrategias complementarias que aseguren su culminación con éxito.

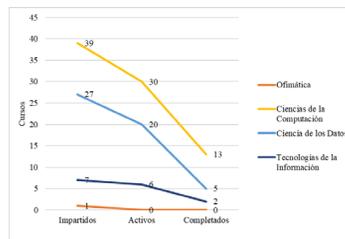


Figura 5. Estado de los cursos

En la Figura 6, se presentan las 146 personas que estuvieron interesadas en los 74 cursos, cuál fue su comportamiento durante el periodo evaluado y que solo 32 terminaron exitosamente.

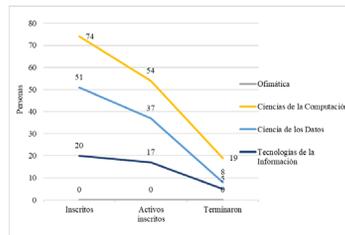


Figura 6. Comportamiento de personas participantes

En conclusión, la eficacia de los cursos y la eficiencia terminal por área se muestra en la Figura 7.

Área	Eficacia de cursos	Eficiencia terminal
Ofimática	0%	0%
Ciencias de la computación	33%	25%
Ciencia de los datos	18%	16%
Tecnologías de la información	28%	25%

Figura 7. Eficacia de cursos y eficiencia terminal por área

## Futuras investigaciones

Para mejorar la eficiencia terminal de los MOOC referentes a la ciencia de los datos, tecnologías de la información, ciencias de la computación y computación, realizar un estudio en las personas que participaron para conocer las causas de su deserción y de quienes culminaron exitosamente, conocer cuáles fueron los factores de éxito.

## Literatura citada

- [1] García-Peñalvo, F. J. (2020). El sistema universitario ante la COVID-19: corto, medio y largo plazo. *Universidad*. <https://www.universidadesi.es/sistema-universitario-covid-19/>
- [2] Universidad Autónoma de Campeche. (2020, 28 de agosto). *Políticas de operación durante la enseñanza*

remota para continuidad del servicio académico durante el ciclo escolar 2020-2021.

- [3] Suárez, D. (2012). *iTunesU y Coursera: pedagogía universitaria 2.0*. Cuaderno de Pedagogía Universitaria, 9(18), 6-8. <http://hdl.handle.net/20.500.12060/639>

## Agradecimientos

Agradecemos a la Universidad Autónoma de Campeche por las facilidades en la realización de esta investigación.



# Revisión de algoritmos de detección de *malware* ofuscado basados en *machine learning*

Erly Galia Villarroel Enríquez  
20182063@aloe.ulima.edu.pe  
<https://orcid.org/0000-0001-8566-0494>  
Universidad de Lima

Recibido: 6 de agosto del 2022 / Aceptado: 9 de octubre del 2022

doi: <https://doi.org/10.26439/ciis2022.6076>

RESUMEN. Los desarrolladores de *malware* cada vez evolucionan más sus técnicas para lograr atacar efectivamente el sistema. Una de estas técnicas es la ofuscación de código, la cual dificulta la detección de *malware* en los mecanismos tradicionales que utilizan los antivirus actuales. Ante ello, se propone la revisión de artículos relacionados con la detección de *malware* ofuscado con *machine learning* para elegir las mejores técnicas de análisis de este tipo de *malware* y emplear los mejores algoritmos para una futura experimentación con ellos.

PALABRAS CLAVE: *malware*, ofuscación, detección, *machine learning*

## REVIEW OF OBFUSCATED MALWARE DETECTION ALGORITHMS BASED ON MACHINE LEARNING

ABSTRACT. Malware developers develop their techniques to attack computer systems effectively. One of these techniques is code obfuscation which makes it difficult to detect malware in the traditional mechanisms used by current antiviruses. This article reviews research related to the detection of obfuscated malware with machine learning to choose the best analysis techniques for this type of malware and use the best algorithms for future experimentation with them.

KEYWORDS: malware, obfuscation, detection, machine learning

## 1. INTRODUCCIÓN

El reporte de FortiGuard Labs (Fortinet, 2022) para América Latina y el Caribe señala que esta región recibió el 10 % del total de intentos de ciberataques que se dieron en el mundo en el 2021, con más de 289 000 millones de intentos de ciberataques, un aumento del 600 % con respecto al 2020. Según datos recabados por FortiGuard Labs, el laboratorio de inteligencia de amenazas de Fortinet, Perú es el tercer país que más intentos de ataques recibió (11 500 millones), después de México y Brasil.

Duo et al. (2022) sostienen que los ciberataques son problemas de seguridad que se presentan en sistemas cibernéticos y que pueden perjudicar el rendimiento del sistema o causar mayores daños en él. Estos ataques pueden ser de denegación de servicio, *phishing* y *malware*, entre otros.

Según diversos autores (Saravia et al., 2019, como se citó en Ashik et al., 2021), el *malware* o *software* malicioso es un programa informático dañino inyectado en programas legítimos para perpetrar acciones ilícitas. Debido al rápido crecimiento de internet y a la aparición de diversos dispositivos conectados a través de redes, el panorama de ataques de *malware* ha aumentado afectando a la privacidad de los usuarios.

Con respecto a las causas que originan las infecciones de *software* malicioso, Ashik et al. (2021) señalan principalmente a las descargas de *software* gratuito. Estas incluyen *freeware* (*software* gratuito) de juegos, navegadores web, antivirus gratuitos, etcétera.

Existen tres formas comunes en la que un *malware* puede entrar a un dispositivo. La primera es el ataque de descarga: cuando el atacante aloja *malware* en un servidor web para infectar a los dispositivos que visiten esa página web. La segunda forma es un ataque de actualización: cuando una aplicación benigna se actualiza tomando características de *malware*. Finalmente, está el ataque de reempaquetado, donde el desarrollador de *malware* empaqueta una aplicación benigna inyectando código malicioso en dicha aplicación (Felt et al., 2014, como se citó en Surendran & Thomas, 2022).

Los ataques de *malware* han ocasionado grandes pérdidas financieras tanto a organizaciones como a individuos (Ashik et al., 2021), así como amenazas de filtración de información personal y robo de información sensible (Lee et al., 2022).

Según Liu et al. (2020), la detección de *malware* es afectada por técnicas como carga dinámica de código, ofuscación de código, entre otras. La ofuscación de código es una técnica que dificulta la detección de comportamiento malicioso en análisis estáticos. Además, facilita al desarrollador de *malware* la creación de múltiples variables de *malware* que pueden causar detecciones erróneas en la mayoría de sistemas existentes (Arp et al., 2014, como se citó en Ouk & Pak, 2022).

Los atacantes usualmente emplean la ofuscación de *malware* para evadir la detección de los programas antivirus. Estos están basados en la coincidencia de patrones de *malware*, pero apenas detectan los nuevos *malware* (Mimura et al., 2016, como se citó en Mimura & Ito, 2022).

Las investigaciones actuales se centran en mejorar la detección de *malware* empleando algoritmos de *machine learning*. Un ejemplo de ello es la investigación de Mahindru & Sangal (2020), quienes extrajeron características de las aplicaciones que recolectaron (benignas o con *malware*) para luego entrenarlas usando distintos algoritmos de *machine learning* y compararlas para determinar el mejor algoritmo para detectar *malware* en dispositivos móviles Android.

Esta búsqueda de mejora en detección incluye al *malware* ofuscado. Según Wu et al. (2021), un desafío abierto es explicar el comportamiento malicioso de un *malware* ofuscado a través del análisis de sus características. Los modelos actuales de *machine learning* solo usan una pequeña porción de estas características para explicar la clasificación de un programa ofuscado como *malware*.

Para Sun et al. (2021), un nuevo desafío en el área de seguridad es el desarrollo de una solución que cuenta con precisión y escalabilidad para detección de *malware* y su clasificación en familias.

Finalmente, Surendran y Thomas (2022) sugieren que se debe reducir la tasa de falsos positivos (cuando una aplicación es clasificada erróneamente como *malware*).

## 2. DESCRIPCIÓN GENERAL DEL TRABAJO

Los desafíos en este campo, como se explicó en la sección de introducción, resaltan la necesidad de una mayor investigación. El presente documento tiene como objetivo analizar las técnicas de preprocesamiento utilizadas para desocultar *malware* ofuscado, además de algoritmos de *machine learning*, para encontrar el mejor algoritmo que pueda detectar *malware* ofuscado en menor tiempo y con la mayor precisión posible.

## 3. CONTRIBUCIONES

Las contribuciones del presente trabajo son las siguientes:

- Realizar una investigación actualizada sobre la detección de *malware* ofuscado.
- Dar una visión sobre las técnicas de preprocesamiento para desocultar *malware* ofuscado.

- Comparar los modelos de *machine learning* usando métricas objetivas como precisión y exactitud.

## REFERENCIAS

- Ashik, M., Jyothish, A., Anandaram, S., Vinod, P., Mercaldo, F., Martinelli, F., & Santone, A. (2021). Detection of malicious software by analyzing distinct artifacts using machine learning and deep learning algorithms. *Electronics*, *10*(14), 1694. <https://doi.org/10.3390/electronics10141694>
- Duo, W., Zhou, M., & Abusorrah, A. (2022). A survey of cyber attacks on cyber physical systems: Recent advances and challenges. *IEEE/CAA Journal of Automatica Sinica*, *9*(5), 784-800. <https://doi.org/10.1109/JAS.2022.105548>
- Fortinet. (2022, 8 de febrero). *América Latina sufrió más de 289 mil millones de intentos de ciberataques en 2021* [Comunicado de prensa]. <https://www.fortinet.com/lat/corporate/about-us/newsroom/press-releases/2022/fortiguard-labs-reporte-ciberataques-america-latina-2021>
- International Business Machines. (2022). *X-Force Threat Intelligence Index 2022*. IBM Security X-Force. <https://www.ibm.com/security/data-breach/threat-intelligence/>
- Liu, K., Xu, S., Xu, G., Zhang, M., Sun, D., & Liu, H. (2020). A review of android malware detection approaches based on machine learning. *IEEE Access*, *8*, 124579-124607. <https://doi.org/10.1109/ACCESS.2020.3006143>
- Lee, D., Jeon, G., Lee, S., & Cho, H. (2022). Deobfuscating mobile malware for identifying concealed behaviors. *Computers, Materials and Continua*, *72*(3), 5909-5923. <http://dx.doi.org/10.32604/cmc.2022.026395>
- Mimura, M., & Ito, R. (2022). Applying NLP techniques to malware detection in a practical environment. *International Journal of Information Security*, *21*(2), 279-291. <https://doi.org/10.1007/s10207-021-00553-8>
- Ouk, P. C., & Pak, W. (2022). High performance classification of android malware using ensemble machine learning. *Computers, Materials and Continua*, *72*(1), 381-398. <http://dx.doi.org/10.32604/cmc.2022.024540>
- Sun, B., Takahashi, T., Ban, T., & Inoue, D. (2021). Detecting Android malware and classifying its families in large-scale datasets. *ACM Transactions on Management Information Systems (TMIS)*, *13*(2), 1-21. <https://doi.org/10.1145/3464323>

- Surendran, R., & Thomas, T. (2022). Detection of malware applications from centrality measures of syscall graph. *Concurrency and Computation: Practice and Experience*, 34(10), e6835. <https://doi.org/10.1002/cpe.6835>
- Wu, B., Chen, S., Gao, C., Fan, L., Liu, Y., Wen, W., & Lyu, M. R. (2021). Why an Android app is classified as malware: Toward malware classification interpretation. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 30(2), 1-29. <https://doi.org/10.1145/3423096>

# Revisión de algoritmos de detección de *malware* ofuscado basados en *machine learning*

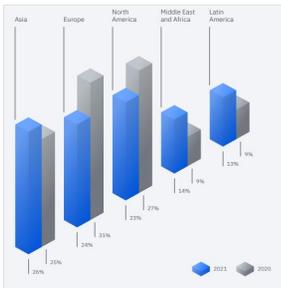
Erlly Galia Villarroel Enríquez  
20182063@aloe.ulima.edu.pe

**Resumen.** Los desarrolladores de *malware* cada vez evolucionan más sus técnicas para lograr atacar efectivamente el sistema, una de estas técnicas es la ofuscación de código que dificulta la detección de *malware* en los mecanismos tradicionales que utilizan los antivirus actuales. Ante ello se propone la revisión de artículos relacionados con la detección de *malware* ofuscado con *machine learning* para elegir las mejores técnicas de análisis de este tipo de *malware* y emplear los mejores algoritmos para una futura experimentación con los mismos.

## Introducción

El reporte de FortiGuard Labs (Fortinet, 2022) para América Latina y el Caribe señala que esta región recibió el 10 % del total de intentos de ciberataques que se dieron en el mundo en el 2021, con más de 289 000 millones de intentos de ciberataques. Según datos recabados por FortiGuard Labs, Perú es el tercer país que más intentos de ataques recibió (11 500 millones), después de México y Brasil. Los ataques de *malware* han ocasionado grandes pérdidas financieras tanto a organizaciones como a individuos (Ashik et al., 2021), amenazas de filtración de información personal y robo de información sensible (Lee et al., 2022). El problema descrito anteriormente ha generado la necesidad de mecanismos para detección de *malware*; sin embargo, los atacantes siempre innovan en técnicas para evitar la detección. Tal como lo señalan Liu et al. (2020), quienes mencionan que la detección de *malware* es afectada por técnicas como carga dinámica de código, ofuscación de código, entre otras. La ofuscación de código es una técnica que dificulta la detección de comportamiento malicioso en análisis estáticos. Además, facilita al desarrollador de *malware* la creación de múltiples variables de *malware* que pueden causar detecciones erróneas en la mayoría de sistemas existentes. (Arp et al., 2014, como se citó en Ouk & Pak, 2022). Los atacantes usualmente emplean la ofuscación de *malware* para evadir la detección de los programas antivirus, estos están basados en la coincidencia de patrones de *malware* que apenas detectan los nuevos *malware* (Mimura et al., 2016, como se citó en Mimura & Ito, 2022).

Figura 1: Ataques de *malware* a nivel mundial, 2021 vs. 2020. Fuente: IBM Security X-Force



## Materiales y métodos

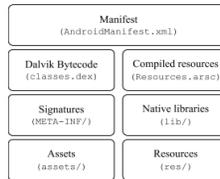
Se revisaron artículos en las bases de datos IEEE Xplore y Scopus que trataban sobre técnicas para detectar *malware* ofuscado en aplicaciones (análisis de *bytecode* y llamadas al sistema) y los principales algoritmos de *machine learning*.

## Resultados

### Detección basada en análisis de *bytecode*

Para el caso de aplicaciones Android, Latchar et al. (2020) sostienen que estas están empaquetadas en forma de archivos APK (Android Package), estos a su vez constan de varios archivos, dentro de los cuales está incluido un archivo ejecutable Dalvik (dex), siendo este el que contiene el *bytecode* o código de *bytes*. Este código es ejecutado por el sistema de Android Runtime (ART), este entorno de ejecución se encarga de transformar el *bytecode* a código máquina para que la aplicación pueda ser ejecutada por el sistema operativo correspondiente.

Figura 2: Estructura de los archivos APK según Chen et al. (2019)



### Detección basada en análisis de llamadas al sistema

Según Surendran y Thomas (2022), las llamadas sensibles a una API generalmente contienen más información que las llamadas al sistema; sin embargo, al utilizar mecanismos que utilizan llamadas al sistema pueden darse erróneas clasificaciones cuando una aplicación benigna invoca llamadas API que son típicas de aplicaciones con *malware*. Por lo tanto, el seguimiento de las llamadas al sistema de una aplicación contiene más información relevante sobre el comportamiento malicioso que las llamadas API sensibles invocadas por la aplicación.

Figura 3: Captura de pantalla de la ejecución en Linux del comando *strace*, el archivo C a ejecutar fue renombrado como "prueba" y se obtuvieron las llamadas al sistema y la frecuencia de estas.

```

$ lsobfd -o /tmp/so -l /usr/lib64/libc.so.6 gcc -o prueba sec.c
$ lsobfd -o /tmp/so -l /usr/lib64/libc.so.6 strace -c ./prueba
# time seconds usecs/call calls errors syscall
-----
0,00 0,000000 0 1 read
0,00 0,000000 0 1 write
0,00 0,000000 0 3 close
0,00 0,000000 0 6 mmap
0,00 0,000000 0 3 mprotect
0,00 0,000000 0 1 munmap
0,00 0,000000 0 1 brk
0,00 0,000000 0 4 pread64
0,00 0,000000 0 1 access
0,00 0,000000 0 1 execve
0,00 0,000000 0 2 1 arch_prctl
0,00 0,000000 0 1 set_tid_address
0,00 0,000000 0 3 openat
0,00 0,000000 0 3 newstatat
0,00 0,000000 0 1 set_robust_list
0,00 0,000000 0 1 prctl
0,00 0,000000 0 1 getrandom
0,00 0,000000 0 1 rseq
-----
100,00 0,000000 0 39 2 total
  
```

Tabla 1: Descripción de algunas llamadas al sistema

Nombre	Descripción
access	Verifica los accesos a un fichero
pread64	Para leer o escribir en un fichero
brk	Cambia el tamaño del segmento de datos
fork	Permite a un proceso crear un proceso hijo
accept	Acepta una conexión sobre un conector (socket)

## Técnicas de *machine learning* utilizadas en detección de *malware* ofuscado

Tabla 2: Comparación de artículos que emplean *malware* ofuscado

Artículo	Dataset	Técnicas de ofuscación	Técnica de ML empleada	Resultados (% accuracy)
AndroDet: An adaptive Android obfuscation detector	Se utilizó el conjunto de datos de AMD y también se recopilaron aplicaciones del repositorio de software "Droid".	Identificar renaming String encryption Control flow obfuscation	SVM	92,02
			SVM	81,41
			SVM	68,32
Impact of Code Deobfuscation and Feature Interaction in Android Malware Detection	Se utilizó el conjunto de datos Drebin y Androzoo.	Code Reduction	LightGBM	99,54
Detection of Malware Applications from Centrality Measures of Sycall Graph	Se utilizó el conjunto de datos Drebin y AMD	Dynamic code loading, renaming and String Encryption	Random Forest	99
Obfuscated VBA Macro Detection Using Machine Learning	Recolectado por los mismos autores a través de búsqueda en Google	Random obfuscation, split obfuscation, encoding obfuscation and logic obfuscation	Random Forest	90,3
Detecting obfuscated JavaScripts using Machine Learning	Utilizaron el dataset "jsDeliver", el cual ofuscaron en la página "javascriptobfuscator.com"	Minificación y compresión	Random Forest	99
			SVM	97

## Conclusiones

A través de la revisión de varios artículos de investigación y documentos de conferencias, podemos denotar la importancia de la aplicación de técnicas de aprendizaje automático en la detección de *malware* ofuscado. Además, destacamos dos importantes técnicas de detección de *malware* ofuscado en aplicaciones Android: el análisis *bytecode* y las llamadas al sistema, para futuros trabajos, estas serán utilizadas en una experimentación utilizando los algoritmos Random Forest y MLP que fueron los que obtuvieron los mejores resultados en todos los documentos revisados.

## Referencias

Amin, M., Shah, B., Sharif, A., Ali, T., Kim, K. I., & Anwar, S. (2022). *Android malware detection through generative adversarial networks. Transactions on Emerging Telecommunications Technologies*, 33(2), e3675.

Chen, Y. C., Chen, H. Y., Takahashi, T., Sun, B., & Lin, T. N. (2021). Impact of Code Deobfuscation and Feature Interaction in Android Malware Detection. *IEEE Access*, 9, 123208-123219.

International Business Machines. (2022). *X-Force Threat Intelligence Index 2022*. IBM Security X-Force. <https://www.ibm.com/security/data-breach/threat-intelligence/>

Mahindru, A., & Sangal, A. L. (2021). MLDroid—Framework for Android malware detection using machine learning techniques. *Neural Computing and Applications*, 33(10), 5183-5240.

Mirzaei, O., de Fuentes, J. M., Tapajador, J., & González-Manzano, L. (2019). AndroDet: An adaptive Android obfuscation detector. *Future Generation Computer Systems*, 90, 240-261.

Sun, B., Takahashi, T., Ban, T., & Inoue, D. (2021). Detecting Android Malware and Classifying Its Families in Large-scale Datasets. *ACM Transactions on Management Information Systems (TMIS)*, 13(2), 1-21.

Surendran, R., & Thomas, T. (2022). Detection of malware applications from centrality measures of sycall graph. *Concurrency and Computation: Practice and Experience*, 34(10), e6855.

Wu, B., Chen, S., Gao, C., Fan, L., Liu, Y., Wen, W., & Lyu, M. R. (2021). Why an android app is classified as malware: Toward malware classification interpretation. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 30(2), 1-29.

Xu, P., & Khairi, A. E. (2021). Android-COCO: Android Malware Detection with Graph Neural Network for Byte-and Native-Code. arXiv preprint arXiv:2112.10938.

## Agradecimientos

Agradezco a mi asesor Juan Gutiérrez por brindarme su apoyo a lo largo de la elaboración de este documento, al profesor Pablo Rojas por brindarme pautas para realizar la exposición, y a mis amigos por animarme a exponer este póster.



**DATOS  
DE LOS AUTORES**

---



## RENZO ANGLES

Doctor en Ciencias con mención en Computación por la Universidad de Chile. Bachiller en Ingeniería de Sistemas por la Universidad Católica de Santa María, Perú. Profesor asociado del Departamento de Ciencias de la Computación de la Universidad de Talca, Chile. Durante el 2013, realizó un posdoctorado en la VU Amsterdam, y participó en el proyecto europeo Linked Data Benchmark Council (LDBC). Sus trabajos de investigación se encuentran en la intersección entre las bases de datos orientadas a grafos y la web semántica. Sus especialidades actuales son la teoría y el diseño de lenguajes de consulta para grafos, así como el análisis de proteínas.

## MOHAMMAD REZA MOUSAVI

Doctor en Ciencias de la Computación por la Universidad Tecnológica de Eindhoven, Países Bajos. Antes de incorporarse al King's College de Londres en el 2021, ocupó cargos en la Universidad de Reikiavik, la Universidad Tecnológica de Eindhoven, la Universidad Tecnológica de Delft, la Universidad de Halmstad, la Universidad Tecnológica Chalmers y la Universidad de Leicester. Entre sus temas de interés están las validaciones basadas en modelos, pruebas y verificación de sistemas ciberfísicos, y líneas de productos de *software* de prueba. Ha liderado iniciativas de investigación y proyectos de colaboración industrial en sistemas de salud y automoción, así como su validación, verificación y certificación.

### TATIANA ANDREEVA

Doctora en Administración por la Universidad Estatal de San Petersburgo, Rusia. Profesora asociada en Gestión y Comportamiento Organizacional, y directora de Investigación en la Escuela de Negocios de la Maynooth University, Irlanda. Su trabajo ha sido publicado en prestigiosas revistas como *Human Resource Management Journal*, *Human Resource Management* y *Journal of World Business*, entre otras. Su campo de investigación aborda los desafíos de la gestión del conocimiento en las organizaciones; actualmente, está desarrollando un proyecto para explorar cómo el cambio al trabajo remoto e híbrido y la creciente digitalización influyen en el intercambio de conocimientos.

### AVID ROMÁN-GONZÁLEZ

Doctor en Procesamiento de Señales e Imágenes por Télécom ParisTech, Francia. Magíster en Automatización Industrial y Humana por la Universidad Paul Verlaine de Metz, Francia. Realizó una estancia posdoctoral en la Universidad Peruana Cayetano Heredia, Perú. Docente principal en la Universidad Nacional Tecnológica de Lima Sur (UNTELS). Miembro sénior IEEE e investigador Renacyt en el grupo Carlos Monge nivel I. Ha realizado labores de investigación para el Centro Nacional de Estudios Espaciales de Francia (CNES) y el Centro Aeroespacial Alemán (DLR). Entre sus temas de interés se encuentra el procesamiento de señales e imágenes, bioingeniería y biométrica, automatización humana, control automático, inteligencia artificial y tecnología aeroespacial.

### ROSARIO GUZMÁN JIMÉNEZ

Magíster en Ingeniería de Sistemas por la Universidad de Lima, Perú. Ingeniera de Sistemas por la Universidad Católica de Santa María. Estudiante del Doctorado en Ingeniería por la Pontificia Universidad Católica del Perú. Profesora e investigadora en el abordaje a problemas complejos en educación, salud y cambio climático, auspiciado por el Instituto de Investigación Científica de la Universidad Lima.

### EDUARDO ALEJANDRO ESCOTTO CÓRDOVA

Doctor y magíster en Humanidades, línea lingüística, por la Universidad Autónoma Metropolitana, México. Licenciado en Psicología con estudios de maestría en Psicobiología por la Universidad Nacional Autónoma de México. Profesor de tiempo completo titular de la Carrera de Psicología y en la Facultad Estudios Superiores Zaragoza de la Universidad Nacional Autónoma de México. Algunos de sus libros y artículos más importantes son “La evaluación cualitativa en la neuropsicología” (2021), “Emotion Classification from EEG Signals Using Wearable Sensors: Pilot Test” (2020), *Recursos semióticos en la enseñanza de las matemáticas* (2018), “El juego en la estimulación temprana del desarrollo en niños preescolares” (2018),

“Los universales semióticos culturales” (2017), “Los recursos semióticos del profesor de estadística asociados al rendimiento académico de los estudiantes” (2017) y el capítulo “Conceptions of Play Activity and its Application” del libro *The Routledge International Handbook of Early Childhood Play* (2017).

#### ALVARO SALDÍVAR (DIVAPATI PREM)

Egresado de la Facultad de Administración de Empresas de la Universidad de Lima, Perú. Coautor del método Tawa Pukllay. Coordinador *yachachiq* de la Asociación Yupanki. Asesor externo de investigación científica para el Instituto de Investigación Científica de la Universidad Lima. Investigador de ciencias andinas.

#### DHAVIT PREM

Maestría en Filosofía con mención en Epistemología por la Universidad Nacional Mayor de San Marcos, Perú. Egresado de la Facultad de Ingeniería de Sistemas de la Universidad de Lima. Autor del método Tawa Pukllay. Director *yachachiq* de la Asociación Yupanki. Asesor externo de investigación científica para el Instituto de Investigación Científica de la Universidad Lima. Inventor, escritor, conferencista y científico independiente.

#### GUILLERMO DÁVILA

Doctor en Ingeniería y Gestión del Conocimiento por la Universidade Federal de Santa Catarina, Brasil. Máster en Ciencias de la Información e ingeniero de sistemas por la Universidad Nacional de Ingeniería, Perú. Profesor de la Carrera de Ingeniería de Sistemas de la Universidad de Lima e investigador del Instituto de Investigación Científica de la misma casa de estudios. También es líder del grupo de investigación Desarrollo Empresarial, Gestión del Conocimiento e Innovación de la Universidad de Lima. Su área de investigación es la gestión del conocimiento organizacional y sus investigaciones han sido publicadas en reconocidos periódicos internacionales. Dos de sus investigaciones fueron premiadas en el prestigioso congreso europeo IFKAD, ediciones 2017 y 2019. En el ámbito profesional, tiene experiencia gestionando proyectos y unidades de negocio en empresas de los sectores de TI, banca y seguros, atracción de inversiones y desarrollo de negocios.

#### INDIRA GUZMÁN

Doctora en Ciencias de la Información y Tecnología por la Universidad de Siracusa, Nueva York, Estados Unidos. Magíster en Gestión de la Información con la beca Fulbright. Ingeniera de sistemas por la Universidad Nacional Técnica de Donetsk, Ucrania. Profesora investigadora en el área de Sistemas de Información Computacionales en la Escuela de Negocios de la Universidad Estatal Politécnica de California, Pomona, Estados Unidos. Cuenta con dieciséis

años de experiencia en el área académica, donde ejerció trabajos de docencia, investigación y dirección de programas académicos en Estados Unidos. Además, tiene una amplia experiencia profesional en la administración de sistemas en la banca. Sus áreas de investigación incluyen el impacto de las tecnologías de la información en las empresas y la sociedad, la cultura ocupacional de los profesionales de TI (ITOC), la brecha de género en TI, auditoría de sistemas y evaluación de riesgos. Es miembro activo de Academy of Management (AOM), Association of Information Systems (AIS), Association of Computing Machinery (ACM), ISACA y el Centro Nacional de Mujeres en Tecnologías de la Información en Estados Unidos (NCWIT). Actualmente, es presidente del Capítulo Latinoamericano de la Asociación de Sistemas de Información. Se desempeña como consultora de investigación del proyecto financiado por el Centro Internacional de Investigaciones para el Desarrollo (IDRC) a través de la Fundación UNISELVA, cuyo objetivo es promover el liderazgo de las mujeres en TI en Bolivia, Perú y Brasil.

### CLAUDIA QUINTANILLA

Ingeniera de sistemas, con especialización en Negocios Electrónicos por ESAN; en Gestión de Tecnología de Información por la Universidad de San Martín de Porres, Perú; en Innovación por la Stanford University; y en Gestión de Proyectos. Es cofundadora de Rextie, la *fintech* líder en el sector de cambio de divisas para pymes y personas del Perú, con más de 3500 millones de dólares transaccionados. Tiene más de quince años de experiencia en diversas áreas del sector financiero; ahora busca aplicar su pasión por el cambio para revolucionar el mundo financiero.

### CRISTIANO MACIEL

Doctor en Computación por la Universidade Federal Fluminense, Brasil. Profesor titular en el Instituto de Computación y en el Programa de Posgrado en Educación, ambos en la Universidad Federal de Mato Grosso, Brasil. Es becario de investigación del Consejo Nacional de Desarrollo Científico y Tecnológico (CNPq). Sus principales intereses de investigación son la ingeniería de *software*, el género, la raza y las tecnologías, y la educación en línea. Desde el 2011, ha estado trabajando en las áreas de investigación de género y tecnologías. Actualmente, supervisa estudiantes de doctorado en el campo de STEM/STEAM. Es uno de los consultores del programa Meninas Digitais (en español, Niñas Digitales) de la Sociedad Brasileña de Computación (SBC). Es coordinador general del proyecto internacional ELLAS, financiado por el International Development Research Centre, Canadá.

### VÍCTOR HUGO AYMA QUIRITA

Doctor y máster en Ingeniería Eléctrica por la Pontificia Universidade Católica do Rio de Janeiro, Brasil. Egresado de la Universidad Nacional de San Antonio Abad del Cusco, Perú. Tiene experiencia en investigación usando aprendizaje automático y aprendizaje profundo.

Actualmente, se desempeña como investigador en el Instituto de Investigación Científica y como docente en la Carrera de Ingeniería de Sistemas de la Universidad de Lima, Perú.

### PEDRO MARCO ACHANCCARAY DÍAZ

Doctor y máster en Ingeniería Eléctrica por la Pontificia Universidade Católica do Rio de Janeiro, Brasil. Egresado de la Universidad Nacional de Ingeniería, Perú. Cuenta con experiencia en proyectos de investigación, desarrollo e innovación (I+D+i) usando teledetección, inteligencia artificial y aprendizaje profundo en las áreas de agricultura, petróleo y gas, y preservación de patrimonio cultural. Actualmente, se desempeña como investigador posdoctoral en el Institute of Geodesy and Photogrammetry (IGP) de la Universidad Técnica de Brunswick, Alemania.

### SMITH WASHINGTON ARAUCO CANCHUMUNI

Doctor en Ingeniería Eléctrica y máster en Ingeniería Mecánica por la Pontificia Universidade Católica do Rio de Janeiro, Brasil. Egresado de la Universidad Nacional de Ingeniería, Perú, en Ingeniería Mecatrónica. Cuenta con años de experiencia en proyectos de investigación, desarrollo e innovación (I+D+i) usando inteligencia artificial, aprendizaje automático, aprendizaje profundo y modelos generativos en las áreas de petróleo y gas. Actualmente, se desempeña como investigador y profesor en el laboratorio de inteligencia computacional de la PUC-Rio.

### PEDRO JUAN SOTO VEGA

Doctor en Ingeniería Eléctrica por la Pontificia Universidade Católica do Rio de Janeiro, Brasil. Máster en Ingeniería Eléctrica por la misma casa de estudios. Graduado en Ingeniería en Telecomunicaciones y Electrónica por la Universidad de Oriente, Cuba. Tiene artículos publicados en varias revistas y congresos internacionales. Ha participado de varios proyectos de investigación e innovación, la mayoría en el Laboratorio de Visión por Computador (LVC) de la PUC-Rio. Actualmente, desarrolla investigación en las áreas de visión artificial, aprendizaje automático, aprendizaje profundo, teledetección y biometría. Es miembro de la International Society of Photogrammetry and Remote Sensing (ISPRS) y del Capítulo Brasileño de la IEEE Geoscience and Remote Sensing Society (GRSS). Cuenta con experiencia en el área de ingeniería eléctrica, con énfasis en telecomunicaciones.

### HERNÁN ALEJANDRO QUINTANA CRUZ

Magíster en Administración de Negocios por la Universidad de Lima, Perú. Profesor en la Universidad de Lima y miembro del comité organizador del CIIS 2022. Hace investigación en temas relacionados con ingeniería de *software* y videojuegos.

### ELVIRA RINCÓN-FLORES

Doctora en Ciencias de la Educación por la Universidad de Salamanca, España. Pertenece al Instituto para el Futuro de la Educación del Tecnológico de Monterrey, México. Es profesora de esta institución desde 1993 en la escuela de Ingeniería y Ciencias, así como en la escuela de Humanidades y Educación.

### BRENDA SANTOS

Magíster en Administración Educativa. Actualmente, trabaja en el Centro de Vinculación y Desarrollo Profesional del Tecnológico de Monterrey, México. Es profesora de Experiencia Profesional en esta casa de estudios desde el 2012. Su proyecto de tesis doctoral es sobre la enseñanza de las matemáticas.

### ALBERTO MATSUURA

Magíster en Dirección Estratégica de Contenidos por la Universidad de Lima, Perú. Es coordinador de Innovación Educativa en la Oficina de Innovación y Calidad Educativa de esta universidad. También es director y productor audiovisual, diseñador gráfico y especialista en contenido digital.

### LILIANA PEDRAJA-REJAS

Doctora en Educación por la Pontificia Universidad Católica de Chile, y doctora en Administración y Dirección de Empresas por la Universidad Politécnica de Valencia, España. Profesora titular de la Universidad de Tarapacá, Chile. Su línea de investigación se centra en educación superior, calidad, liderazgo, cultura organizacional, gestión institucional y formación inicial docente.

### EMILIO RODRÍGUEZ-PONCE

Doctor en Educación por la Universidad Autónoma de Barcelona, y doctor en Ciencias Económicas y Empresariales por Universidad Complutense de Madrid, España. Profesor titular de la Universidad de Tarapacá, Chile. Su línea de investigación se centra en educación superior, calidad y acreditación, liderazgo educativo, aprendizaje organizacional y formación inicial docente.

### CAMILA MUÑOZ-FRITIS

Ingeniera civil industrial por la Universidad de Tarapacá, Chile. Actualmente, es estudiante de la Maestría en Ciencias de la Ingeniería, con mención en Matemáticas Aplicadas, en la Universidad de Chile.

### YANA GONCHAROVA

Estudiante de doctorado en la Universidad de Campania “Luigi Vanvitelli”, Italia. Graduada en la Universidad Estatal de Pyatigorsk en Traducción Literaria (título de especialista). El tema de investigación de su tesis es la imagen de Italia en la cultura rusa entre los siglos XIX y XX. Sus intereses científicos son los estudios literarios, la literatura rusa, la lingüística, la comunicación intercultural, la traducción, la educación en línea y los enfoques educativos basados en las competencias.

### OLGA SHEVELEVA

Estudiante de doctorado en la Universidad Estatal de Dubná, Rusia. Máster en Análisis y Gestión de Sistemas en la misma universidad con la tesis *Desarrollo de la conciencia espacial de los estudiantes con el uso de tecnologías 3D*. Licenciada en Análisis y Gestión de Sistemas por la misma casa de estudios. Su campo de investigación es el desarrollo de un nuevo modelo estructurado para las competencias TIC como parte de la universidad digital intelectual del futuro.

### VLADIMIR DOBRYNIN

Doctor por la Universidad Estatal de Dubná, Rusia. Ha estudiado la especialidad de Matemática en la Universidad Estatal de Kharkiv. Investigador sénior. Su actividad profesional gira en torno a la gestión de sistemas complejos (técnicos, sociotécnicos en condiciones de incertidumbre). Su trabajo científico trata de la modelización de sistemas de control de aeronaves y la búsqueda y exploración de yacimientos de minerales sólidos. Es profesor en la Universidad de Dubná en los campos de teoría de sistemas, análisis informático de sistemas dinámicos, tecnologías de gestión del conocimiento, modelos y métodos de análisis de soluciones de diseño.

### ALISSON AVALOS TAPIA

Bachiller en Ingeniería Empresarial por la Universidad del Pacífico, Perú.

### JUAN DAVID CÁRDENAS

Estudiante de sexto ciclo de Ingeniería de la Información en la Universidad del Pacífico, Perú.

### RODRIGO CABALLERO

Estudiante de sexto ciclo de Ingeniería de la Información en la Universidad del Pacífico, Perú.

### VICTOR ANDRES AYMA QUIRITA

Doctor en Ingeniería por la Pontificia Universidad Católica del Perú (PUCP). Magíster en Procesamiento de Señales, Automatización y Control por la Pontificia Universidad Católica

do Rio de Janeiro, Brasil. Graduado y titulado en Ingeniería Electrónica por la Universidad Nacional de San Antonio Abad del Cusco, Perú. Miembro investigador del Grupo de Inteligencia Artificial (IA-PUCP) y presidente del Capítulo de Ingeniería Electrónica del CIP-CD Cusco. Docente en la Universidad del Pacífico (UP) y de la Escuela de Posgrado de la PUCP (Maestría en Informática). Asimismo, es consultor internacional en análisis de datos, procesamiento de imágenes y asesor e investigador en proyectos de I+D+I. Desde el 2016, sirve como revisor en diferentes revistas indexadas de alto impacto; actualmente, tiene la calificación de investigador nivel I del grupo María Rostworowski del Renacyt.

### JAVIER MORE SÁNCHEZ

Ingeniero electrónico y de telecomunicaciones por la Universidad Nacional de Piura, Perú. Magíster en Ingeniería de las Telecomunicaciones por la Pontificia Universidad Católica del Perú. Magíster en Regulación y Gestión de Servicios Públicos por la Universidad del Pacífico. Máster en Gestión de Infraestructuras por la Universidad Politécnica de Cataluña, España. Estudiante del Doctorado en Ingeniería de Sistemas e Informática en la Universidad Nacional Mayor de San Marcos, Perú. Tiene más de doce años de experiencia en el sector telecomunicaciones. Actualmente, labora en el Organismo Supervisor de Inversión Privada en Telecomunicaciones (OSIPTEL) y es docente en la Universidad de Lima.

### DIANA CONCEPCIÓN MEX ÁLVAREZ

Ingeniera en Sistemas Computacionales por la Universidad Autónoma de Campeche, México. Maestra en Pedagogía con especialidad en Entornos Virtuales de Aprendizaje. Actualmente, es estudiante del Doctorado en Proyectos. Colabora como profesora e investigadora de la Facultad de Ingeniería de la Universidad Autónoma de Campeche, donde imparte las asignaturas de programación en la Licenciatura en Ingeniería en Sistemas Computacionales, Ingeniería en Mecatrónica e Ingeniería en Energía. Cuenta con la distinción de perfil PRODEP, otorgada por la Secretaría de Educación Pública del Gobierno de México. Es líder del Cuerpo Académico, reconocido por la Dirección de Superación Académica de la Secretaría de Educación Pública del Gobierno de México, "Ciencias de la Computación", con clave UNACAM-060.

### MARGARITA CASTILLO TÉLLEZ

Doctora en Energías Renovables. Ingeniera industrial por la Universidad Nacional Autónoma de México. Miembro del Sistema Nacional de Investigadores. Cuenta con la distinción de perfil PRODEP, otorgada por la Secretaría de Educación Pública del Gobierno de México. Colabora en la Universidad Autónoma de Campeche como profesora e investigadora en los programas educativos de Licenciatura en Ingeniería en Energía, Licenciatura en Mecatrónica y Maestría en Ingeniería en Energías Renovables y Eficiencia Energética. Es líder del Cuerpo Académico, reconocido por la Dirección de Superación Académica de la Secretaría de Educación Pública

del Gobierno de México, “Sustentabilidad Energética y Energías Renovables”, con clave UNACAM-057.

### FRANCISCO EMMANUEL ALFARO JUÁREZ

Estudiante de la Licenciatura en Ingeniería Mecatrónica en la Universidad Autónoma de Campeche, México. Realizó sus estudios de educación media superior en el Colegio de Bachilleres del Estado de Campeche, plantel 02 Candelaria. Participa como prestador de servicio social en el Cuerpo Académico “Ciencias de la Computación”, con clave UNACAM-060.



UNIVERSIDAD  
DE LIMA

El V Congreso Internacional de Ingeniería de Sistemas (CIIS 2022) tuvo como objetivo la reflexión, discusión y descripción de aplicaciones relacionadas con los entornos híbridos y las nuevas tecnologías, que cobraron especial relevancia en el contexto de las restricciones por la pandemia del COVID-19. No solo han contribuido a alcanzar resultados en términos económicos, sino también en el potencial innovador que puede surgir de estos entornos, el bienestar de sus miembros o la flexibilidad como competencia que permita adaptarse a futuros cambios.

El CIIS 2022, como en anteriores ediciones, congregó a investigadores de primer nivel. En esta memoria, se recogen tres ponencias magistrales, que estuvieron a cargo de la doctora Tatiana Andreeva, el doctor Mohammad Reza Mousavi y el doctor Renzo Angles. Asimismo, se describen las secciones de paneles, donde expertos nacionales y extranjeros expusieron y discutieron sus puntos de vista sobre temas específicos, como un sistema educativo remoto *online/offline*, la inclusión de mujeres en STEM, los desafíos del aprendizaje profundo en la visión por computador y una plataforma para la gamificación dirigida a estudiantes universitarios. Finalmente, se reúnen los trabajos de investigación presentados en el evento, los cuales tratan sobre sistemas de información, ingeniería de *software*, ciencias de la computación y tecnología.