



Facultad  
de Informática  
y Electrónica



# REVISTA PERSPECTIVAS

REVISTA TÉCNICA CIENTÍFICA DE LA FIE



▣ JULIO-DICIEMBRE 2024

e-ISSN: 2661-6688

VOL 6, N°2



*epoch*  
**HACEMOS  
HISTORIA**





# REVISTA PERSPECTIVAS

REVISTA TÉCNICA CIENTÍFICA DE LA FIE

**JULIO - DICIEMBRE 2024**

e - ISSN: 2661 - 6688

VOL 6, N°2

RIOBAMBA - ECUADOR



*epoch*  
**HACEMOS  
HISTORIA**



**epoch**

Facultad de  
Informática y  
Electrónica

Instituto de  
Investigaciones

Dirección de  
Publicaciones



REVISTA PERSPECTIVAS

PRÓXIMA  
CONVOCATORIA

ENERO - JUNIO 2023

<http://perspectivas.esPOCH.edu.ec/>

## CONVOCATORIA DE RECEPCIÓN DE MANUSCRITOS

---

### PERIODICIDAD SEMESTRAL

#### 1ERA CONVOCATORIA:

Recepción abierta hasta octubre

**Publicación:** Enero

#### 2DA CONVOCATORIA:

Recepción abierta hasta Abril

**Publicación:** Julio



*esPOCH*  
**HACEMOS  
HISTORIA**



**esPOCH**

Facultad de  
Informática y  
Electrónica



## COMITÉ EDITORIAL

---

- **DIRECTOR**

**Omar S. Gómez, Ph.D.**

*Escuela Superior Politécnica de Chimborazo (Ecuador)*

---

- **EDITOR EJECUTIVO**

**Raúl H. Rosero, Ph.D.**

*Escuela Superior Politécnica de Chimborazo (Ecuador)*

---

- **EDITORES ASOCIADOS**

**Raúl Antonio Aguilar Vera, Ph.D.**

*Universidad Autónoma de Yucatán (México)*

---

**Jesús Pardo Calvache, Ph.D.**

*Universidad del Cauca (Colombia)*

---

**Rosa Sumactika Delgadillo Ávila de Mauricio, Ph.D.**

*Universidad Nacional Mayor de San Marcos (Perú)*

---

**George Enrique Figueras Benítez, Ph.D.**

*Universidad Simón Bolívar (Venezuela)*

---

**Luis Antonio Rivera Escriba, Ph.D.**

*Universidade Estadual do Norte Fluminense (Brasil)*

---

**Talia Beatriz Tene Fernandez, Ph.D.**

*Universidad Técnica Particular de Loja (Ecuador)*

---

- **SECRETARIO CIENTÍFICO**

**Mayra A. Pacheco Cunduri, M.Sc.**

*Escuela Superior Politécnica de Chimborazo (Ecuador)*

---

- **COORDINACIÓN PROCESO PUBLICACIÓN**

**Diego Ávila Pesántez, M.Sc.**

*Escuela Superior Politécnica de Chimborazo (Ecuador)*

---

- **REVISORES DE IDIOMA INGLÉS**

**Nelly Padilla Padilla, M.Sc.**

*Escuela Superior Politécnica de Chimborazo (Ecuador)*

---

- **EDICIÓN DE DISEÑO Y DIFUSIÓN**

**Rosa Ramos Jiménez, M.A.**

*Escuela Superior Politécnica de Chimborazo (Ecuador)*

---

- **DISEÑO Y DIAGRAMACIÓN**

**José Luis Heredia Hermida Mgtr**

*Escuela Superior Politécnica de Chimborazo (Ecuador)*



## COMITÉ CIENTÍFICO

---

**Lorena Molina Valdiviezo, Ph.D.**

*Universidad Nacional de Chimborazo (Ecuador)*

---

**Eliana Acurio Méndez, Ph.D.**

*Escuela Politécnica Nacional (Ecuador)*

---

**Miguel Delgado Prieto, Ph.D.**

*Universidad Politécnica de Cataluña (Ecuador)*

---

**Luis Miguel Procel Moya, Ph.D.**

*Universidad San Francisco de Quito (Ecuador)*

---

**Luis Tello Oquendo, Ph.D.**

*Universidad Nacional de Chimborazo (Ecuador)*

---

**Cristian Vacacela Gómez, Ph.D.**

*Universidad Yachay Tech (Ecuador)*

---

**Ciro Radicelli García, Ph.D.**

*Universidad Nacional de Chimborazo (Ecuador)*

---

**Lorena Guachi Guachi Núñez, Ph.D.**

*Universidad Yachay Tech (Ecuador)*

---

**Patricio Humanante Rámos, Ph.D.**

*Universidad Nacional de Chimborazo (Ecuador)*

---

**Edison Taco Lasso, Ph.D.**

*Universidad San Francisco de Quito (Ecuador)*

---

**Edison Espinosa, Ph.D.**

*Filiación: Universidad de las Fuerzas Armadas (Ecuador)*

---

**Maricela Jiménez Rodríguez, Ph.D.**

*Filiación: Universidad de Guadalajara (México)*

---

**Juan Carlos Estrada, Ph.D.**

*Filiación: Universidad de Guadalajara (México)*

---



## CONTENIDO

### 7 PRESENTACIÓN

- 
- 1 - 16** Estado de la investigación en sistemas de administración de colecciones de biodiversidad: un mapeo sistemático de la literatura
- 
- 17 - 24** Matriz de trazabilidad en el levantamiento de requisitos: una revisión sistemática de la literatura
- 
- 25 - 36** Implementación de un controlador para robot de tipo péndulo invertido
- 
- 37 - 46** Análisis de datos para la optimización eficiente de horarios y aprendizaje automático
- 
- 47 - 54** Análisis del comportamiento de datos médicos de covid -19 mediante inteligencia artificial
- 
- 55 - 70** Mejora del aprendizaje basado en la lógica: aplicación de guess-18 para la evaluación de la experiencia del usuario
- 
- 71 - 83** Hacia un Modelo de Gobierno de APIs, Mapeo Sistemático de la Literatura
-



## PRESENTACIÓN

Apreciables lectores,

En este año que inicia, hacemos de su conocimiento la publicación del número 2 del volumen 6 de la revista Perspectivas. En este número se encuentran disponibles publicaciones afines a las áreas de informática, electrónica, telecomunicaciones y diseño. Sabemos que los contenidos ofrecidos en este número serán de interés para nuestra comunidad de lectores.

Reiteramos nuestro agradecimiento todos aquellos quienes han hecho posible mantener la continuidad de esta revista, a nuestros autores por sus contribuciones y por confiar en

este medio de divulgación técnico-científica, a nuestro equipo de revisores que dedican parte de su tiempo en el proceso de revisión con el fin de contar con publicaciones de mayor calidad. Agradecemos también al equipo editorial así como a nuestras autoridades institucionales.

Recordamos a nuestra comunidad que continua abierta la recepción de manuscritos, los cuales tras su correspondiente proceso de revisión y aceptación se publicarán en los números correspondientes de este año.

**Cordialmente,  
EQUIPO EDITORIAL**

**Saber para Ser !**



# ESTADO DE LA INVESTIGACIÓN EN SISTEMAS DE ADMINISTRACIÓN DE COLECCIONES DE BIODIVERSIDAD: UN MAPEO SISTEMÁTICO DE LA LITERATURA

State of research in biodiversity collection management systems:  
A systematic mapping of the literature

Karina Valdés-Iglesias <sup>1</sup>	gcvaldesiglesiaskarina@gmail.com
Jorge Octavio Ocharán-Hernández <sup>2</sup>	jocharan@uv.mx
Juan Carlos Pérez-Arriaga <sup>3</sup>	juaperez@uv.mx
Christian Alejandro Delfín-Alfonso <sup>4</sup>	cdelfin@uv.mx

<sup>1,2,3,4</sup> Universidad Veracruzana.

## RESUMEN

Las colecciones de biodiversidad permiten tomar decisiones para la preservación de un ecosistema. Ya que poseen una relación entre el aumento o disminución de la población de especies ligado a la erosión de su hábitat. Por lo que el crecimiento de estas ha impulsado la necesidad de migrar hacia sistemas de administración de biodiversidad que sigan el ciclo de vida establecido por la Ingeniería de Software. Es crucial tener un conocimiento actualizado sobre el estado del arte en los sistemas de administración de colecciones de biodiversidad, conocer las estrategias que se han implementado en conjunto de los beneficios para investigaciones académicas. Motivando a la futura implementación de los sistemas y proveyendo una base de los elementos a considerar para su construcción. Por esta razón, se llevó a cabo un mapeo sistemático de la literatura en base a la guía de Kitchenham, abordando las características y esfuerzos en este campo. Los hallazgos revelan: (i) El desarrollo de sistemas de administración de colecciones de biodiversidad está experimentando un crecimiento significativo a nivel nacional. Al igual que universidades han comenzado a implementarlos en sus herbarios. (ii) La tecnología predominante empleada es PHP en el *framework* de Laravel, principalmente debido a sus funciones incorporadas que facilitan el desarrollo de aplicaciones con recursos limitados. (iii) Los estándares más

comúnmente utilizados son el International Code of Botanical Nomenclature y el Darwin Core. (iv) Los sistemas de administración de colecciones de biodiversidad ofrecen beneficios sustanciales al mantener la consistencia y accesibilidad de los datos, facilitando así la toma de decisiones para la preservación de la biodiversidad. (v) Los atributos de calidad de mayor interés incluyen: accesibilidad, facilidad de aprendizaje, reconocimiento adecuado, desempeño en tiempo real e interoperabilidad que puede apoyar en la definición de requisitos no funcionales durante el desarrollo de sistemas. (vi) Los sistemas de administración de colecciones de biodiversidad a nivel local muestran una mayor frecuencia de desarrollo en el ámbito botánico en comparación con el zoológico, esto debido a la facilidad de recolección. Estos hallazgos proporcionan una justificación sólida para la implementación de sistemas de administración de colecciones de biodiversidad, junto con pautas para sus funcionalidades principales y atributos de calidad relevantes. Asimismo, ofrecen un punto de partida valioso para el desarrollo de futuros sistemas en este ámbito.

**Palabras Clave:** sistemas de administración de biodiversidad, colección de biodiversidad, bases de datos de biodiversidad, Mapeo sistemático de la literatura, síntesis narrativa.

## ABSTRACT

Biodiversity collections allow making decisions for the preservation of an ecosystem since they have a relationship between the increase or decrease in species population linked to the erosion of their habitat. Therefore, the growth of these collections has driven the need to migrate towards biodiversity management systems that follow the life cycle established by Software Engineering. It is crucial to have up-to-date knowledge about the state of the art in biodiversity collection management systems, understanding the strategies implemented along with the benefits for academic research. This motivates the future implementation of systems, providing a foundation for the elements to consider in their construction. For this reason, a systematic literature mapping was carried out based on the Kitchenham guide, addressing the characteristics and efforts in this field. The findings reveal: (i) The development of biodiversity collection management systems is experiencing significant growth nationally, with universities starting to implement them in their herbaria. (ii) The predominant technology used is PHP in the Laravel framework, mainly due to its built-in functions that facilitate the development of applications with limited resources. (iii) The most used standards are the International Code of Botanical Nomenclature and the Darwin Core. (iv) Biodiversity collection management systems offer substantial benefits by maintaining consistency and accessibility of data, thus facilitating decision-making for biodiversity preservation. (v) Quality attributes of interest include accessibility, ease of learning, adequate recognition, real-time performance, and interoperability that can support the definition of non-functional requirements during system development. (vi) Biodiversity collection management systems at the local level show a higher frequency of development in the botanical domain compared to the zoological, primarily due to the ease of collection. These findings provide a solid justification for the implementation of biodiversity collection management systems, along with guidelines for their main functionalities and relevant quality attributes. Additionally, they offer a valuable starting point for the development of future systems in this domain

**Keywords:** biodiversity management systems, biodiversity collection, biodiversity databases, systematic literature mapping, narrative synthesis

## » I. Introducción

En la actualidad, la conservación de la biodiversidad incluye la protección de especies en riesgo (ej. en peligro de extinción) y el mantenimiento de áreas silvestres. Motivo por el cual es uno de los principales problemas del siglo 21. La biodiversidad es afectada por la destrucción y fragmentación del hábitat, caza ilegal y la explotación de especies, entre otras [1]. Para mitigar la erosión de la biodiversidad se monitorea los proyectos de conservación y sostenibilidad ya sea por organizaciones gubernamentales o educativas para fines de divulgación científica [2]. Ya que la biodiversidad, se ha enfrentado continuamente desafíos que han resultado en procesos de defaunación a nivel global, como señala Dirzo et al. (2014). De tal forma que muchas especies animales, es posible que solo las identifiquemos en colecciones biológicas en universidades públicas y privadas. Es por ello por lo que resulta necesario un inventario de los especímenes de colecciones de biodiversidad que proveen una evidencia física de la presencia y existencia de ellos. Las colecciones de biodiversidad utilizan datos fiables y se dificulta su consulta al no encontrarse registrados a bases de datos normalizadas. Puesto que, al almacenarse en un laboratorio aislado, la divulgación de la colección se limita a nivel local.

La globalización de la biodiversidad es afectada por no aplicar técnicas de escalabilidad en la base de datos que permitan la interoperabilidad y accesibilidad de los datos taxonómicos [3]. La accesibilidad se convierte en un limitante que afecta la reutilización y análisis de la información para investigaciones académicas que ayudan para toma de decisiones para conservación el ambiente [4].

Las colecciones requieren un sistema que administre su base de datos estandarizada y estructurada; los cuales son sistemas de administración de colecciones de biodiversidad. Su objetivo principal es desarrollar un sistema que permita la interoperabilidad de conocimientos en una amplia gama de sistemas locales, e integrarlos en arquitecturas de conocimiento global [5]. Sin embargo, es necesario seguir un método antes de comenzar el desarrollo de este. Momento donde

la Ingeniería de Software ayuda herramientas y técnicas para verificar los requisitos actuales y planificar para los nuevos, incluyendo los atributos de calidad. Elaborar una clasificación antes de pasar al diseño de la base de datos posee como principio la capacidad de recuperación [3] y, posteriormente comenzar con la implementación siguiendo una trazabilidad con los requisitos y diseño.

En relación con nuestro estudio, el objetivo principal es conocer el estado del arte del desarrollo de los sistemas de administración de colecciones de biodiversidad. Identificar los avances al igual que las características propias que benefician a los atributos de calidad mencionados anteriormente. Involucrando la documentación utilizada como los estándares taxonómicos. Junto con los retos actuales orientados a la globalización de la biodiversidad descrita. Buscando motivar el desarrollo de los sistemas a nivel local y proveer puntos de mejora para los implementados en la actualidad.

## » II. Método de investigación

El primer paso fue la elaboración de una búsqueda manual de mapeos y revisiones sistemáticas de la literatura por medio del navegador de Google Scholar. Esto con el fin de conocer el estado de la investigación del tema. Siendo el único resultado una revisión sistemática de la literatura titulada "Aplicación de la visión de Computadora y Machine Learning para la digitalización de especímenes del herbario". Donde su principal objetivo fue buscar técnicas y aplicaciones utilizadas para digitalización del herbario, al igual que los desafíos existentes a partir de las limitaciones de los recursos de la universidad. Con los resultados de la revisión se detallaba una solución, junto con los materiales usados para su implementación [6].

Esa revisión fue centrada en la digitalización de un herbario de una universidad. Los investigadores buscaron técnicas de clasificación de Machine Learning. Apegados a un presupuesto limitado. Se nota la diferencia de perspectiva con nuestro mapeo. Ya que nuestro objetivo es localizar el

estado del arte de los sistemas de administración de colecciones. Proveyendo pautas para la especificación de requisitos funcionales y atributos de calidad; a los investigadores.

Por lo cual desarrollamos el mapeo sistemático de la literatura siguiendo el método de [7], que se divide en tres fases principales: Planeación, conducción y documentación. Para el análisis de los datos se utilizó la síntesis narrativa, que se presentará en los siguientes apartados de la ejecución del método.

- PI-1: ¿Cuáles son los sistemas de administración de colecciones de biodiversidad que se encuentran en uso? PI-2: ¿Qué tecnologías fueron utilizadas en la construcción de estos sistemas?
- PI-2.1: ¿Cuáles fueron los criterios de selección para el uso de estas tecnologías?
- PI-3: ¿Qué estándares taxonómicos fueron utilizados en la construcción de estos sistemas?
- PI-4: ¿Cuáles fueron los beneficios de la digitalización de una colección de biodiversidad para la investigación de la biodiversidad de las especies?
- PI-5: ¿Cuáles son los retos actuales en el desarrollo de sistemas para administración de colecciones de biodiversidad?
- PI-6: ¿Qué tipos de especímenes se encuentran registrados en las colecciones que cuentan con un sistema de administración?

### II-1. Proceso de búsqueda

: En el proceso se siguieron dos estrategias: búsqueda manual y búsqueda automatizada. Con el fin de asegurar la calidad en la recaudación de estudios para la futura extracción. Las cuales detallaremos a continuación:

**1. Búsqueda manual:** Especificadas las preguntas de investigación, se seleccionaron las bases de datos donde se realizará la búsqueda. Las cuales son: IEEE Xplore Digital Library, ACM Digital Library son recomendadas por [7]; Science Direct y Springer Link, son bases de datos multidisciplinarias de apoyo a nuestro tema; y por último BioOne por ser una base de

datos referente a temas biológicos.

- *Construcción de la cadena:* Los journals de SpringerLink representativos de nuestro tema fueron: BMC Bioinformatics, BMC Ecology y Biodiversity and Conservation para comenzar con una búsqueda manual. En la que se identificaron seis estudios primarios de nuestro tema. Donde se localizó palabras claves para la elaboración de la cadena.
- *Evaluación de la cadena:* La validación de la cadena se apoyó de las fórmulas propuestas por Zhang et al., calculando la recuperación y precisión de una cadena en una base de datos permite verificar la calidad de las distintas cadenas formadas. Ya que los estudios primarios reconocidos se encontraron en SpringerLink, es la base donde se escogió la evaluación.

**2. Búsqueda automatizada:** Los mapeos de acuerdo con Zhang deben tener una recuperación de mínimo 0.8 para que una cadena se catalogue fiable para la búsqueda automatizada. Por lo que en la Tabla A1 se encuentran las distintas evaluaciones de cadenas en la base de datos de SpringerLink, la cual puede ser consultada en los anexos. La cadena de búsqueda obtuvo un 0.83 de rendimiento y 0.015 de precisión. Cumpliendo el requerido por Zhang para que una cadena sea considerada fiable, la cual fue la siguiente: **(“Biodiversity“ OR “Biological data“ OR “GBIF“) AND (“Information System“ OR “Portals“ OR “Standard“ OR “Dataset“)**

#### II-2. Proceso de selección:

**1. Criterios de selección:** Con la cadena seleccionada, se escogieron los criterios para la evaluación en cada base de datos. Siendo el principal objetivo de los criterios el asegurar la calidad de los estudios y optimizar el proceso de evaluación con ayuda de los filtros previstos por cada base de datos. Que se describen a continuación:

##### *Criterios de inclusión*

CI-1: Fecha de publicación entre agosto 2017 y

septiembre de 2022.

CI-2: El estudio se encuentra en inglés.

CI-3: El título da indicios de responder alguna de las preguntas de investigación.

CI-4: El *abstract* da indicios de responder alguna de las preguntas de investigación.

CI-5: Texto completo disponible.

CI-6: El texto completo del estudio contesta al menos una de las preguntas de investigación.

##### *Criterios de exclusión*

CE-1: No se incluye información sobre biodiversidad, por ejemplo: médicas, toxicológicas, químicas. Entre otras.

CE-2: El estudio es un resumen, presentación u opinión. CE-3: Estudio duplicado.

Son un total de tres fases para la evaluación de cada estudio. La primera fase son los criterios de inclusión CI-1, CI-2 y de exclusión CE-1, CE-2 donde se buscó descartar los artículos que no son de apoyo a nuestro tema de investigación. En la segunda fase son los criterios de inclusión de CI-3, CI-4 y CI-5 en la búsqueda de respuestas a las preguntas de investigación. La última fase involucra el criterio de inclusión CI-6 y el criterio de exclusión CE-3 asegurando que los documentos finales respondan una o más preguntas de investigación.

**2. Extracción de datos:** La extracción de datos se divide en dos apartados, los cuales son:

- Datos de la publicación: Información necesaria para identificar el artículo. Los campos para considerar: Título, autores, año, fuente, convenio, palabras clave, resumen y tipo de publicación.
- Datos para la síntesis: Campos relacionados a alguna pregunta de investigación. Se localizaron palabras claves de cada pregunta que nos conduzcan a la respuesta que posteriormente fue clasificada en un análisis estadístico. Los campos fueron:
  - a) Sistemas de administración de colecciones de biodiversidad vigentes: Relacionada a la pregunta PI-1.
  - b) Tecnologías seleccionadas en los sistemas de administración de colecciones de biodiversidad: Relacionada a la pregunta PI-2.

- c) Estándares taxonómicos utilizados para los sistemas de administración de colecciones de bio- diversidad: Relacionada a la PI-3.
- d) Beneficios de los sistemas de administración de colecciones de biodiversidad: Relacionada con la pregunta PI-4.
- e) Retos actuales de los sistemas de administración de colecciones de biodiversidad: Relacionada con la pregunta PI-5.
- f) Población y Jerarquía de las colecciones de biodiversidad: Relacionada a la pregunta PI-6.

**3. Síntesis de datos:** Se siguió el proceso de la síntesis narrativa que se basa en el uso de palabras claves y texto para resumir y explicar los resultados por medio de datos estadísticos y descripción de las conclusiones.

Proceso que se siguió se dividió en: Valoración crítica, descripciones y clusters, tabulación y gráficas [8].

#### II-A. Conducción

En la FIGURA que se encuentra en anexos. Fue diseñada con el diagrama de flujo de mapeos sistemáticos ROSES,

s.f. En la primera etapa se muestran los resultados del proceso de selección de las bases de datos contempladas, con SpringerLink con diez, IEEE Xplore con ocho, ACM Digital Library con dos y BioOne junto con Science Direct con uno respectivamente. Desviándonos del proceso de Kitchenham, se realizó una búsqueda en Google Scholar, con el fin de encontrar artículos que no se encontrarán indexados en una de las bases de datos contempladas. Donde se encontró un journal de acceso libre titulado Biodiversity Data Journal que pertenece PenSoft que es una editorial de literatura científica con sede en Sofía, Bulgaria. En la cual se siguió el proceso de los criterios de selección, obteniendo seis artículos. Que al sumarse con los localizados en las bases de datos nos da un total de 29 artículos. Con los estudios seleccionados que se encontrarán citados

en los anexos, se usó la herramienta MAXQDA. La cual proporciona las funcionalidades para generar códigos de nuestros campos de extracción y agruparla por segmentos. Ayudando a organizar la información para la tabulación y elaboración de gráficas para responder nuestras preguntas de investigación.

### ► III. Resultados

#### III-A. Distribución de los estudios

En la búsqueda automatizada de las cinco bases de datos seleccionadas (ver Figura 1). La mayor cantidad de estudios fue en SpringerLink con un total de 10 estudios. Ya que en esta fue que la evaluación de nuestra cadena y donde pertenecen los dos journals representativos. La clasificación de forma descendente fue: IEEE Xplore con ocho, Biodiversity Data Journal con 7, ACM Digital Library con dos estudios y ScienceDirect junto con BioOne con uno.

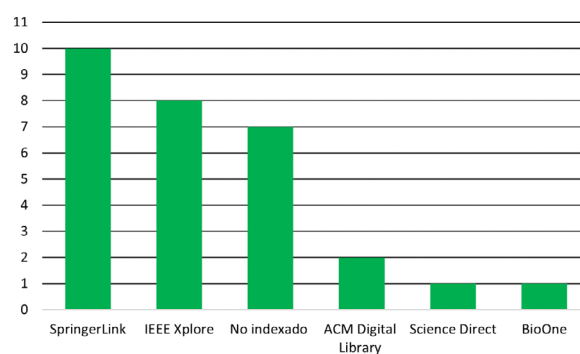


Fig. 1. Distribución de estudios por base de datos.

En la Figura 2, muestra la distribución de los estudios por año. Siendo 2021 el que posee un mayor interés en el tema con un total de nueve estudios el año y 2019 el de menor cantidad con uno. Se observa que no hay una tendencia clara de decremento a pesar de la notable reducción en 2019. Sin embargo, en los siguientes, el número fue en aumento, consolidarse en el 2022 con un total de cinco estudios.

#### III-B. Hallazgos por pregunta de investigación

**III-B1. Sistemas de administración de colecciones de bio- diversidad que se encuentran en uso:** Con

un total de 57 sistemas localizados en los estudios. Se clasificaron en los niveles: Global, País y Local. Que se encuentran en la Tabla A7 que se encuentran en los anexos. Esta división sirve para notar una mayor distribución a nivel país con un total de 33 colecciones, siguiendo con Global con 23 y por último Local con cuatro. Algunos de los países mencionados con colecciones fueron: Estados Unidos, Brasil, Australia. Mientras que locales abarcaban ciudades o universidades como lo eran Florida.

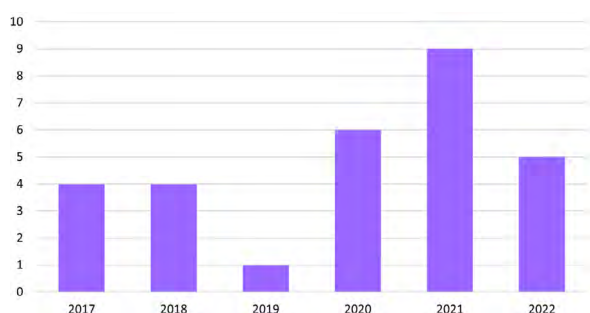


Fig. 2. Distribución de estudios por año de publicación.

**III-B2. Tecnologías utilizadas en la construcción de los sistemas de administración de colecciones de biodiversidad:** En la Figura 3. La tecnología dominante fue PHP en el framework Laravel. Donde algunos criterios de selección que son localizados en el EP13 fueron: Funciones incorporadas que mitigan las amenazas de ciberseguridad, facilidad para la organización de la información, junto con flexibilidad a los requisitos del negocio. Respecto al apartado de ciberseguridad, se realizó una investigación del framework Laravel que provee la gestión de la seguridad por medio de: autenticación, rutas protegidas por privilegios, CSRF y encriptación de datos [9]. Para el mismo apartado de desarrollo web se nombró HTML5 y JavaScript. En aplicaciones móviles solamente se desarrollaron para el sistema operativo de Android. Por lo que se construyeron en el framework de Android Studio que utiliza el lenguaje de Java, siendo el único mencionado en los estudios EP22 y EP25.

Respecto a los gestores de bases de datos los estudios coinciden en MySQL siendo sus criterios mencionados en el estudio EP21: Facilidad de

uso y curva de aprendizaje rápida para aquellos que inician en la introducción de datos o están acostumbrados a utilizar Excel. Por último, en el despliegue que es la última etapa de un ciclo de vida de Software, los artículos mencionaban a Docker, el cual es de gran apoyo por su simplicidad y permite un alto índice de compatibilidad.

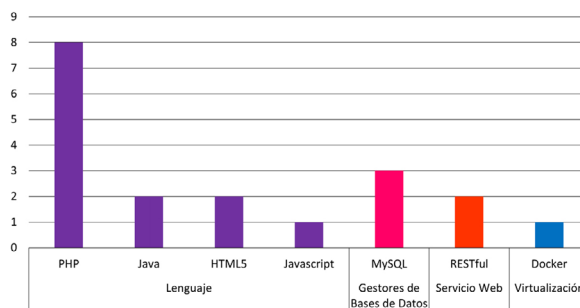


Fig. 3. Distribución de tecnologías de sistemas de administración de colecciones de biodiversidad.

**III-B3. Estándares taxonómicos utilizados en la construcción de los sistemas de administración de colecciones de biodiversidad:**

- **International Code of Botanical Nomenclature (ICBN):** Es el conjunto de normas y recomendaciones que rigen la denominación científica de todos los organismos tradicionalmente tratados como algas, hongos o plantas, fósiles o no fósiles. [10]
- **International Code of Zoological Nomenclature (IZBN):** Proporciona y regula un sistema uniforme de nomenclatura zoológica que garantiza que cada animal tenga un nombre científico único y universalmente aceptado. El mantenimiento de normas internacionales de nomenclatura animal es la función exclusiva de la Comisión. No es función de la Comisión involucrarse en cuestiones taxonómicas, excepto cuando tienen implicaciones nomenclaturales. [11]
- **Darwin Core:** Ofrece un marco de trabajo estable, sencillo y flexible para recopilar datos de biodiversidad provenientes de fuentes diferentes y variables. Desempeña un papel fundamental al compartir, usar y reusar los datos de biodiversidad de libre acceso y en la actualidad representa la gran mayoría de los cientos de millones de registros de presencia

de especies disponibles a través de GBIF.org. [12]

En la Figura 4 muestra la frecuencia de aparición de los estándares en los estudios. Darwin Core e International Code of Botanical Nomenclature (ICBN) posee la misma frecuencia. El último es International Code of Zoological Nomenclature (ICZN) con un total de seis.

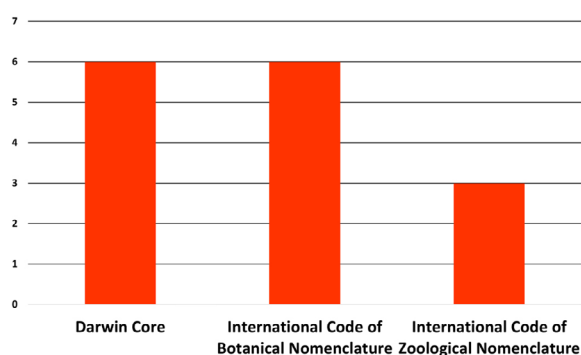


Fig. 4. Distribución de estándares de los sistemas de administración de colecciones de biodiversidad.

### III-B4. Beneficios de la digitalización de una colección de biodiversidad para la investigación

Se extrajeron oraciones cortas que mencionaban distintos beneficios de la digitalización. Con base a ello se elaboró una clasificación por categorías para trazar la relación de los beneficios respecto a impacto externo e interno. Los cuales fueron:

- Consistencia de la información: Coherencia y estandarización de los datos de especímenes presentados en la colección para facilitar la interpretación en futuros estudios biológicos.
- Acceso a la información: Distribuir los datos de la colección de biodiversidad para investigaciones científicas y académicas, buscando mantener activo el estudio del tema.
- Toma de decisiones para la biodiversidad: Por medio de los datos proporcionados por la colección, sea posible la evaluación de los especímenes con el fin de tomar decisiones sobre preservación de la especie, seguimiento y cuidado del ecosistema donde habitan.
- Ahorro de tiempo en tareas manuales: Optimizar el tiempo de subida y distribución de datos, para asignar prioridad a las actividades de investigación y análisis de la

colección.

- Mantenimiento de la colección: Su principal objetivo es apoyar a la conservación y preservación de las especies. Llevando un seguimiento de las especies por especímenes y si son necesarias medidas para evitar el decremento de la especie.

En la Figura 5. El primer lugar es el mantenimiento de la colección con un total de 16 apariciones en estudios. En segundo lugar, se encuentra un empate el acceso a la información y toma de decisiones para biodiversidad, que va de la mano con el mantenimiento. El tercer lugar lo ocupan el ahorro de tiempo en tareas manuales y mientras que la consistencia de la información. El interés de la digitalización de una colección se centra en el seguimiento de las especies para su conservación y distribución, junto con la representación taxonómica para historia del ecosistema mencionada en el EP06. A partir de la información de la colección pueden definirse planes de conservación del ecosistema. Apoyados por estudios académicos afines a la restauración del hábitat. Esto requiere una constante actualización. Lo que conlleva a una necesidad en facilitar la carga de datos para los curadores y proveer los mecanismos de seguridad necesarios que ayuden a la consistencia de la información con base de un estándar taxonómico.

### III-B5. Retos actuales en el desarrollo de sistemas para administración de colecciones de biodiversidad

Los extractos de los artículos relacionados con los retos actuales son clasificados según los siguientes atributos de calidad, de acuerdo con el [13]: Usabilidad, rendimiento, compatibilidad, seguridad, funcionalidad idónea y facilidad de mantenimiento. Los atributos de calidad se desglosaron en sub- atributos de calidad propuestos por el estándar.

Además del ISO 25010, en los estudios EP04 y EP11 se menciona los FAIR principios. Los cuales son directrices para la gestión y administración de datos científicos para

apoyar la reutilización de activos digitales [14]. Relacionado al aumento del apoyo informático y buscar un trato de calidad de la información gracias a la complejidad y velocidad de la creación de los datos. Siendo las directrices mencionadas: Findability (Localizable), Accessibility (Accesibilidad), Inter-operability (Interoperabilidad), Reuse (Reusar). El de interés es Localizable, porque es el único que no se encuentra en los atributos o subatributos del estándar. Por lo que se clasifica la directriz en los subatributos de usabilidad del ISO 25010, por su relación directa al usuario. A continuación, en la Tabla A9 que se encuentran en los anexos. Se presentan las definiciones de los subatributos para una comprensión de la distribución encontrada en los estudios.

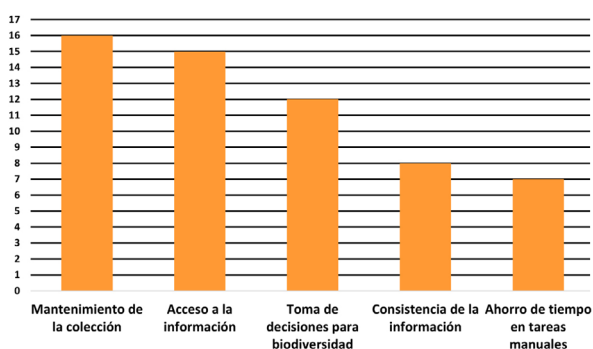


Fig. 5. Distribución de beneficios de los sistemas de administración de colecciones de biodiversidad.

En la Figura 6. Se muestra la distribución de los sub-atributos siendo el de mayor frecuencia accesibilidad con 14, siguiendo con facilidad de aprendizaje, reconocimiento adecuado y obteniendo un empate entre operabilidad, localizable e interfaz de usuario estética. La siguiente fueron los sub-atributos comportamiento de tiempo y capacidad. Siguiendo con compatibilidad que involucra interoperabilidad. El antepenúltimo es seguridad con los apartados de confidencialidad e integración. Finalizando con la exactitud en funcionalidad junto con modificabilidad. Centrándonos en el apartado de usabilidad fue una relación con la necesidad de administrar de la información con el propósito de interpretar para estudios académicos.

La accesibilidad al seguir las pautas conocidas

por los biólogos como lo es el proceso que siguen al curar un espécimen antes de etiquetarlo; ayudando a que los datos sean localizables y reutilizables. Por lo que la finalidad principal de la organización adecuada de la colección es motivar a investigadores y alumnos a la interpretación de la colección en futuras divulgaciones científicas. El rendimiento involucra la gestión de los datos entrantes, sin afectar las funcionalidades del sistema. En compatibilidad se muestra en la correcta y estandarizada representación taxonómica permitiendo sea reutilizable y facilidad en la importación a bases de datos globales.

La seguridad centrada en la confidencialidad e integración de la información. Ya que las colecciones de biodiversidad al tener datos delicados, y sin una correcta autenticación y autorización pueden terminar expuestos a cazadores ilegales y vendedores en el mercado negro. La funcionalidad idónea con exactitud en funcionalidad en las precisiones y retorno en los buscadores avanzados. Al final, se encuentra la facilidad de mantenimiento que involucra el constante ajuste según las ambiciones actuales de la biodiversidad.

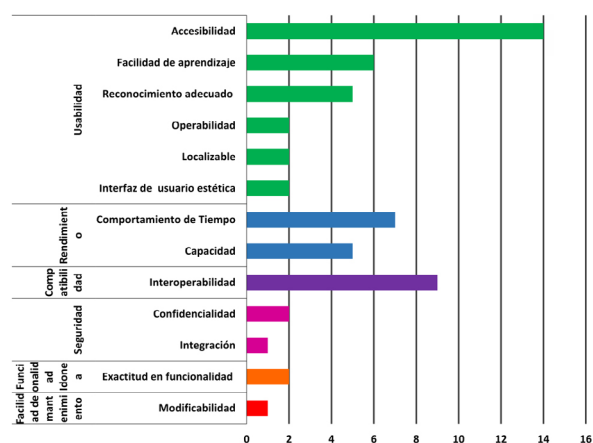


Fig. 6. Distribución de sub-atributos de calidad de los sistemas de administración de colecciones de biodiversidad.

**III-B6. Jerarquía de la colección de biodiversidad en los sistemas de administración:** La necesidad de estandarización de las colecciones de biodiversidad, involucra el conocer la estructura base que es: clase, orden, epíteto, especie y subespecie que es en la que se encuentran



catalogados las especies. En la tabla A11 que se encuentran en los anexos. Se muestran la población de las colecciones detalladas a profundidad en los artículos, al igual que su ubicación.

La botánica es la dominante en ubicaciones, una de las razones es que la recolección de plantas para su conservación es más sencilla. Ya que en las curaciones taxonómicas de mamíferos requiere una constante vigilancia por el movimiento de la especie, junto con métodos de preservación, o de una parte del organismo. Por lo que se puede concluir que es más activa la digitalización de colecciones botánicas a nivel país, por los recursos disponibles e investigaciones en el ámbito de medicina.

#### ► IV. Discusión

El desarrollo de sistemas de administración de colecciones de biodiversidad muestra una mayor frecuencia en países. Ya que algunas de sus ventajas contra los sistemas globales es la delimitación de la colección facilitando la integración de la información en base a estándares taxonómicos. En comparación a los sistemas locales cuentan con una rama amplia de recursos ya sea en el uso de tecnologías para cualquier etapa de implementación; al igual que equipo especializado para los estudios de campo y observación de los especímenes. A pesar de que, el desarrollo de sistemas de colecciones de biodiversidad sea mayor en los países.

En nuestros estudios seleccionados fueron descritas a profundidad las de nivel local. Detallando el procedimiento desde la investigación previa, la selección de tecnologías, junto con las limitantes propias de este nivel. Ya que las colecciones de países se centran a nivel tanto de colección zoológica y botánica, el proceso de implementación es superior al rango de años para el mapeo sistemático. Por lo tanto, que en el apartado de tecnologías no se encontró gran variedad. Porque los únicos artículos que mencionaban pila de tecnologías utilizadas provenían de IEEE Xplore o SpringerLink. El primero al estar centrado a los temas Ciencias de la Computación, Ingeniería Eléctrica y Electrónica, el enfoque de los estudios

era en el desarrollo de los sistemas. Mientras que SpringerLink al ser multidisciplinaria se encontraron los estudios centrados al sistema. Las otras plataformas ofrecían un enfoque biológico, que apoyo a conocer los beneficios del sistema a la propia colección, no su construcción.

La localización de sistemas de colecciones descritos en estudios se agrupaba en América del Norte y Australia. Eso no significa que los sistemas de colecciones solamente se desarrollan en esos sitios. Porque como se observó en la Tabla A7 en distintos países se encontraba un sistema para organizar su biodiversidad. Una de las razones puede ser que no se encuentre un artículo relacionado en las bases de datos consultadas, o su desarrollo se remonte de años atrás y no se encuentra detallados sus retos actuales. Sin embargo, con la información brindada, fue posible el clasificar los retos actuales en atributos de calidad de interés a un nivel local, apoyando a instituciones a tener una referencia para la futura construcción de los suyos.

#### ► V. Amenazas a la validez

##### *V-A. Validez externa*

Un limitante de nuestro estudio recae en acceso a los artículos, ya que algunos fueron descartados al no localizar su texto completo, básicamente por tratarse de revistas cuyo acceso por artículo tiene un costo que oscila entre los 20 y 50 dólares americanos, por tanto, nuestra aproximación muestra resultados parciales. Al igual que en los apartados de retos actuales, la mayoría de los sistemas de administración detallados en los artículos eran de índole local. Siendo nuestro alcance el conocer las tecnologías, retos y estándares utilizados a los desarrollados por Instituciones o pequeñas regiones. Ya que no se especificaban las limitaciones en los sistemas a mayor escala como lo eran GBIF e IDigBio.

##### *V-B. Validez interna*

Una de las principales amenazas a la validez de todo de tipo de mapeo sistemático de la literatura es la cuestión de la fiabilidad. Que pueden

manifestarse en distintas etapas del mapeo, las cuales se describirán a continuación:

*1. Fiabilidad en la búsqueda de resultados:*

- a) Evaluación de los criterios de los estudios seleccionados desde un punto de vista subjetivo: Fue mitigado con la consulta de los coordinadores sobre el contenido del artículo, llegando a un consenso para descartarlo o incluirlo.
- b) Estudios descartados por falta de acceso al texto completo: Fue mitigado al buscar por medio de fuentes externas los estudios ya sea por un navegador o consultando al autor, hasta asegurarse que no fuera posible conseguir el texto completo.

*2. Fiabilidad en la extracción y clasificación de datos:*

- a) Pérdida de datos por traducción de los estudios para su clasificación: Mitigando el sesgo por el cambio de idioma se apoyó de traductores y búsqueda de palabras desconocidas, investigando el contexto con otros textos académicos.
- b) Análisis de los extractos de los estudios: Se consultó con los coordinadores la información clasificada, junto con un respaldo de los segmentos originales para posibles consultas.

*3. Fiabilidad en la síntesis, análisis y visualización de los datos:*

- a) Uso de las técnicas para la síntesis narrativa: El proceso de análisis fue supervisado por los coordinadores. Los cuales fueron revisados y rectificados hasta asegurar la calidad de las gráficas.

## » VI. Conclusiones

En el artículo se presenta por primera vez un mapeo sistemático de la literatura donde se describió el estado actual de los sistemas de administración de colecciones de biodiversidad. Con un total de 29 estudios primarios localizados para responder nuestras preguntas de investigación. Las respuestas a ellas fueron:

- El desarrollo de sistemas de administración de colecciones de biodiversidad es mayor a nivel país. Ya que encuentra limitantes en recursos en sistemas locales.
- La tecnología dominante fue PHP en el Framework de Laravel por sus funciones incorporadas que minimice las amenazas de seguridad y flexibilidad en los requisitos del negocio.
- Los estándares con mayor frecuencia fueron Darwin Core que es usado por GBIF que es la colección más representativa a nivel global. El segundo fue International Code of Botanical Nomenclature.
- Los beneficios de los sistemas de administración de colección de biodiversidad es mantener la consistencia y accesibilidad de los datos para permitir la toma de decisiones para preservación de la biodiversidad.
- Los atributos de calidad de interés son: Accesibilidad, facilidad de aprendizaje, reconocimiento adecuado, comportamiento de tiempo e interoperabilidad.
- Los sistemas de administración de colecciones de bio- diversidad a nivel local tienen mayor frecuencia en desarrollo cuando son de tipo botánica que zoológica.

Los hallazgos principales denotan el enfoque hacía los beneficios del desarrollo de estos sistemas. Aunque se detalla la globalización de las colecciones, no se describe en profundidad en varios de ellos el proceso para la implementación del sistema. En los que se llega a describir, se sigue el ciclo de vida de un sistema, como lo establece la Ingeniería de Software. Apoyando a disminuir la complejidad al comprender el dominio y siguiendo una trazabilidad para asegurar el resguardo correcto de la colección.

Concluimos que nuestro mapeo es una contribución para fomentar una base sobre los sistemas de administración de colecciones de biodiversidad. Alentando a la construcción de sistemas con una justificación sólida de sus beneficios y demostrando que el limitante de recursos no es impedimento para el desarrollo. Detallando las características principales para el futuro esfuerzo en este rubro. Apoyando a la construcción

de sistemas de la misma índole y alcance de universidades, reconociendo la prioridad en los atributos de calidad.

## » VII. Referencias

- [1] S. Vijeta, S. Shikha y S. Anamika, “The principal factors responsible for biodiversity loss”, *Open Journal of Plant Science*, págs. 011-014, ene. de 2021, ISSN: 26407906. DOI: 10.17352/ojps.000026.
- [2] P. Stephenson y A. R. de Paz, “New database enhances the accessibility of global biodiversity information for conservation monitoring”, *Oryx*, vol. 56, págs. 329-330, 3 mayo de 2022, ISSN: 0030-6053. DOI: 10.1017/s0030605322000205.
- [3] J. Soberón y A. T. Peterson, “Biodiversity informatics: Managing and applying primary biodiversity data”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 359, págs. 689-698, 1444 abr. de 2004, ISSN: 09628436. DOI: 10.1098/rstb.2003.1439.
- [4] S. M. Asmara, M. Man y M. T. Abdullah, “A Model of Biodiversity Taxonomy Database (iTaxo) for Managing and Analysing Kenyir Landscape Biodiversity Data”, vol. 769, *Institute of Physics Publishing*, jun. de 2020. DOI: 10.1088/1757-899X/769/1/012017.
- [5] P. Shanmughavel, “An overview on biodiversity information in databases”, en, *Bioinformation*, vol. 1, n.o 9, págs. 367-369, mar. de 2007.
- [6] B. R. Hussein, A. Malik, W.-H. Ong, J. Willem y F. Slik, “Application of Computer Vision and Machine Learning for Digitized Herbarium Specimens: A Systematic Literature Review”, 2019.
- [7] B. A. Kitchenham, D. Budgen y P. Brereton, “Evidence-Based Software Engineering and Systematic Reviews”, 2019.
- [8] J. Popay, H. Roberts, A. Sowden et al., “Guidance on the conduct of narrative synthesis in systematic Reviews. A Product from the ESRC Methods Programme. Version 1”, undefined, 2006. DOI: 10.13140/2.1.1018.4643.
- [9] Laravel, *Security Laravel*. dirección: <https://laravel.com/docs/4.2/security>.
- [10] N. J. ( J. Turland, J. H. Wiersema, F. R. Barrie et al., *Código internacional de nomenclatura para algas, hongos y plantas (Código de Shenzhen)* : adoptado por el decimonoveno Congreso Internacional de Botánica, Shenzhen, China, julio de 2017, pág. 322, ISBN: 9783982013701.
- [11] I. C. on Zoological Nomenclature., W. D. L. Ride e I. U. of Biological Sciences., *International code of zoological nomenclature*. International Trust for Zoological Nomenclature, c/o Natural History Museum, 1999, pág. 306, ISBN: 0853010064.
- [12] GBIF. “¿Qué es Darwin Core y por qué es importante?” (2012), dirección: <https://www.gbif.org/es/darwin-core>.
- [13] *Standards Publication Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE)-System and software quality models*. 2013, ISBN: 9780580702235.
- [14] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg et al., “Comment: The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, vol. 3, mar. de 2016, ISSN: 20524463. DOI: 10.1038/sdata.2016.18.
- [15] H. Zhang, M. A. Babar y P. Tell, “Identifying relevant studies in software engineering”, *Information and Software Technology*, vol. 53, págs. 625-637, 6 jun. de 2011, ISSN: 09505849. DOI: 10.1016/j.infsof.2010.12.010.
- [16] G. M. D. L. Vega, “Colecciones biológicas”. dirección: <https://www.researchgate.net/publication/338954715>.
- [17] E. Turnhout y S. Boonman-Berson, “Databases, Scaling Practices, and the Globalization of Biodiversity”, 2011. dirección: <http://www.gbif.org/index.php?id=269>.
- [18] N. V. Ivanova y M. P. Shashkov, “Biodiversity databases in Russia: towards a national portal”, *Arctic Science*, vol. 3, págs. 560-576, 3 sep. de 2017, ISSN: 2368-7460. DOI: 10.1139/as-2016-0050.
- [19] P. J. Stephenson y C. Stengel, “An

inventory of bio- diversity data sources for conservation monitoring”, PLoS ONE, vol. 15, 12 December dic. de 2020, ISSN: 19326203. DOI: 10.1371/journal.pone.0242923.

[20] B. Jackowiak, M. Lawenda, M. M. Nowak et al., “Open Access to the Digital Biodiversity Database: A Comprehensive Functional Model of the Natural History Collections”, Diversity, vol. 14, 8 ago. de 2022, ISSN: 14242818. DOI: 10.3390/d14080596.

[21] ROSES, ROSES Flow Diagram for Systematic Maps. dirección: ROSES % 20Flow % 20Diagram % 20for % 20Systematic%20Maps.

**Tabla A1**  
**EVALUACIÓN DE CADENAS**

Cadena	Resultados	Estudios relevantes	Precisión	Retorno
("Biodiversity Database. OR "Biodiversity Data. OR "Biodiversity repository. OR "GBIF") AND ("Management System. OR "Software. OR "Standard")	138	2	2/6 * 100 = 33.33 %	2/138 * 100 = 1.44 %
("Biodiversity. OR "Biological data. OR "GBIF") AND ("Information System. OR "Portals. OR "Standard. OR "Dataset")	3784	5	5/6 * 100 = 83.33 %	5/3784 * 100 = 0.33 %
("Biodiversity Database" OR "Biodiversity Data" OR "Biodiversity repository" OR "GBIF") AND ("Management System" OR "Software" OR "Standard")	2842	3	3/6 * 100 = 50 %	3/2842 * 100 = 0.10 %

**Tabla A3**  
**ESTUDIOS SELECCIONADOS**

ID	Título	DOI	Autores	Año
EP01	Database: Taxonomy of plants Nomenclature for borneo biodiversity Information System	10.1109/IAC.2017.8280642	Budiman Edy, Alam Sitti Nur	2017
EP02	Evaluation of the Bioinformatics Resource Portal	10.1109/SEE.2018.8720973	ICITI- Haerullah Jamil, Muh Wati, Masna Saudek	2018
EP03	Performance Analysis of the Resource Loading Time for Borneo Biodiversity Information System	10.1109/IAC.2018.8780515	Budiman Edy, Puspitasari Novianti, Haerullah Jamil, Muh Wati, Masna Saudek, Indra Dolly	2018
EP04	OpenBioMaps – self-hosted data management platform and distributed service for biodiversity related data	10.1007/s12145-022-00818-3	Bán Miklós, Boné Gábor, Máté Bérces, Sándor Barta, Zoltán Kovács, István Ecsedi, Kornél Sipos, Katalin	2022
EP05	An index of ecological value for European arable plant communities	10.1007/s10531-021-02191-x	Fanfarillo E. , Kasperski A.	2021
EP06	Taxonomic shortfalls in digitised collections of Australia’s flora	10.1007/s10531-019-01885-7	Haque Md. , Mohasinul Beaumont, Linda J, Nipperess David A.	2020
EP07	Towards a global list of accepted species VI: The Catalogue of Life checklist	10.1007/s13127-021-00516-w	Donald Hobern, Saroj K. Barik, Les Christidis, Stephen T.Garnett, Paul Kirk, Thomas M. Orrell, Thomas Pape, Richard L. Pyle, Kevin R. Thiele, Frank E. Zachos, Olaf Bánki.	2021

EP08	The persistent multi-dimensional biases of biodiversity digital accessible knowledge of birds in China	10.1007/s10531-020-02024-3	Huang Xiongwei, Lin Congtian, Ji Liqiang	2020
EP09	Ethnobotany Database: Exploring diversity medicinal plants of Dayak Tribe Borneo	10.1109/EEC-I.2017.8239094	Johan Herni, Hairah mmul, Budi-man Edy	2017
EP10	Ethnobotany Database: Exploring diversity medicinal plants of Dayak Tribe Borneo	10.1109/EEC-SI.2017.8239094	Johan Herni, Hairah Ummul, Budiman Edy	2017
EP11	The use of Global Biodiversity Information Facility (GBIF)-mediated data in publications written in Chinese	10.1016/j.gecco.2020.e01406	Luo, Maofang Xu, Zheping Hirsch, Tim Aung, Thant Sin Xu, Wubing Ji, Liqiang Qin, Haining Ma, Keping	2021
EP12	Rapid enhancement of biodiversity occurrence records using unconventional specimen data	10.1007/s10531-018-1584-0	Pearson, Katelin D.	2018
EP13	Biodiversity Repository and Retrieval System for Malay Language	10.1109/AiC54368.2022.9914585	Pethie, Hazimah Nordin, Sharifalillah	2022
EP14	Addressing the implementation challenge of the global biodiversity framework	10.1007/s10531-020-02009-2	Phang, Sui C. Failler, Pierre Bridgewater, Peter	2020
EP15	Evaluation of Borneo's Biodiversity Information System	10.1109/EEC-IS.2018.8692955	Puspitasari, Novianti Budiman, Edy	2018
EP16	A Visualization-based Approach for the Taxonomy Browser Interface	10.1145/3160504.3160523	Rey Marina Fortes, Freitas Carla Maria Dal Sasso	2017
EP17	A comprehensive review on biodiversity information portals	10.1007/s10531-022-02420-x	Saran Chaudharyumar Priyanka Amrpal Vishal Sameer, Sumit, Singh, Tiwari, Kumar	2022
EP18	A Data Chaining on Relational Database: A Case Study with Indonesian Genomics Information System	10.1109/ISI-TIA52817.2021.9502249	Sulistyawan, I Gede Eka, Arifin Achmad, Fatoni Muhammad Hilman	2021
EP19	Towards a global list of accepted species II. Consequences of inadequate taxonomic list governance	10.1007/s13127-021-00518-8	Thomson Scott A., Thiele Kevin, Conix Stijn, Christidis Les Costello Mark John Hobern, Donald Nikolaeva, Svetlana Pyle, Richard L. van Dijk, Peter Paul Weaver, Haylee Zachos, Frank E. Zhang, Zhi-Qiang Garnett, Stephen T.	2021
EP20	GBIF information is not enough: national database improves the inventory completeness of Amazonian epiphytes	10.1007/s10531-022-02458-x	de Araujo Matheus L., Quaresma Adriano C., Ramos Flavio N.	2022
EP21	Digitizing specimens in a small herbarium: Aviable workflow for collections working with limited resources	10.3732/apps.1600125	Harris Kari M., Marsico Travis D.	2017

**Tabla A5**  
**ARTÍCULOS LOCALIZADOS POR SEDE**

Sede	Artículos
International Conference on Informatics and Computing (ICIC)	EP01
Applied Informatics International Conference (AiC)	EP13
Brazilian Symposium on Human Factors in Computing	EP16
Conference on Computing and Sustainable Societies (COMPASS)	EP10

Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)	EP15
International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)	EP09
International Conference on Informatics and Computing (ICIC)	EP03
International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)	EP02
International Conference on Cybernetics and Intelligent System (ICORIS)	EP22
International Seminar on Intelligent Technology and Its Applications (ISITIA)	EP18
Organisms Diversity & Evolution	EP07, EP19
Biodiversity and Conservation	EP05, EP06, EP08, EP12, EP14, EP17, EP20
Applications in Plant Sciences	EP21
Earth Science Informatics	EP04
Global Ecology and Conservation	EP11
Biodiversity Data Journal	EP23, EP24, EP25, EP26, EP27, EP28, EP29

**Tabla A7**  
**SISTEMAS DE ADMINISTRACIÓN DE COLECCIONES DE BIODIVERSIDAD POR NIVEL**

Nivel	Sistemas de Administración de Colecciones de Biodiversidad
Global	Global Biodiversity Information System, OpenBioMaps, National Specimen Information Infrastructure, Global Biodiversity Information Facility (GBIF), Global Biodiversity Framework (GBF), iNaturalist, iDigBio, Biodiversity Heritage Library, Encyclopedia of Life, Avibase database, FishBase, AmphibiaWeb, The Paleobiology database, The Global Invasive Species Database, Reptile Data-Base, Avibase database, SpeciesLink, Biodiversity Information Serving Our Nation, National Specimen Information Infrastructure, World Flora Online
País	Bioinformatics Resource Portal, Borneo's Biodiversity Information System, ArEco, Floristic Quality Index, Natura Capital Index, Arable Land Naturalness Index, Atlas of Living Australia, Australasian Virtual Herbarium (AVH), Australian Plant Census, BirdReport, Catalogue of Life China, Ethnobotany Database, ElephantBook, Biodiversity Repository and Retrieval System, TaxonomyBrowser, SinBiota, Information System about the Brazilian Biodiversity, Indian Bio-resource Information Network, Biodiversity information system for Europe, Indian Bioresource Information Network, India Biodiversity Portal, The Brazilian Biodiversity Information System, Indonesian Genomics Information System, Atlas of Living Australia, Southeast Regional Network of Expertise and Collections (SERNEC) Portal, Symbiota, iDigBio's Specimen Portal
Local	FSU herbarium, Atlas of Florida Plants, STAR Herbarium, New York Botanical Garden

**Tabla A9**  
**DEFINICIÓN Y LOCALIZACIÓN DE SUB-ATRIBUTOS DE CALIDAD**

Sub-atributo de calidad	Descripción	Artículos
Accesibilidad	Grado en el que el producto o sistema puede ser usado por personas de distintas características y capacidades para completar metas específicas al contexto.	EP01, EP14, EP16, EP17, EP19, EP20, EP21, EP23, EP5, EP26, EP27, EP28
Reconocimiento adecuado	Grado en el que los usuarios pueden reconocer un producto o sistema es adecuado para sus necesidades	EP04, EP07, EP15, EP16
Facilidad de aprendizaje	Grado en el que el producto o sistema puede ser usado por usuarios específicos para completar metas de aprendizaje de interacción del sistema	EP01, EP11, EP25, EP27, EP28, EP29
Operabilidad	Grado en el que el producto o sistema tiene atributos para facilitar el control y operaciones.	EP04, EP23

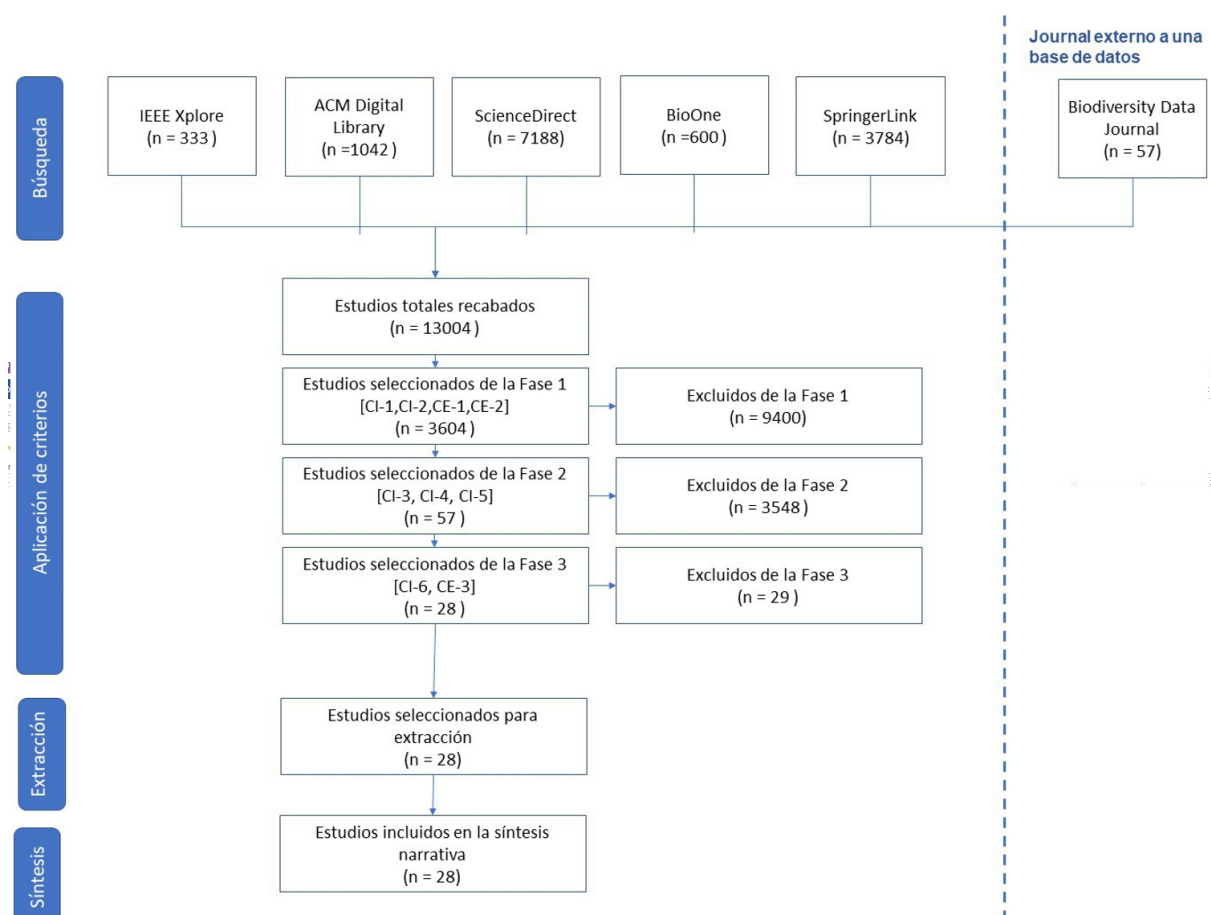
Interfaz de usuario estética	Grado en el que la interfaz del usuario causa satisfacción al interactuar con ella	EP20, EP25
Comportamiento de Tiempo	Grado en que los tiempos y procesamientos de respuesta cumplen con los requisitos	EP03, EP14, EP22
Capacidad	Grado en el que los límites del sistema o parametros del sistema cumplen con los requisitos	EP03, EP04, EP05, EP27
Interoperabilidad	Grado en el que dos o más sistemas, productos o componentes pueden intercambiar información	EP06, EP07 ,EP12, EP16, EP19, EP22, EP24
Confidencialidad	Grado en el que un producto o sistema permite el acceso a los datos solo a personal autorizado	EP04 ,EP07
Integración	Grado en el que un sistema, producto o componente previene el acceso no autorizado o modificación de la información	EP20
Exactitud en funcionalidad	Grado en el que el sistema cumple con específicas tareas y objetivos del usuario	EP13
Modificabilidad	Grado en que un producto o sistema puede modificarse de forma eficaz y eficiente sin introducir defectos sin disminuir la calidad	EP14
Localizable	Los datos deben ser fáciles de encontrar tanto para las personas como para las computadoras.	EP04, EP11, EP29

**Tabla A9**  
**DEFINICIÓN Y LOCALIZACIÓN DE SUB-ATRIBUTOS DE CALIDAD**

Sub-atributo de calidad	Descripción	Artículos
Accesibilidad	Grado en el que el producto o sistema puede ser usado por personas de distintas características y capacidades para completar metas específicas al contexto.	EP01, EP14, EP16, EP17, EP19,EP20,EP21, EP23, EP5,EP26, EP27, EP28
Reconocimiento adecuado	Grado en el que los usuarios pueden reconocer un producto o sistema es adecuado para sus necesidades	EP04, EP07, EP15, EP16
Facilidad de aprendizaje	Grado en el que el producto o sistema puede ser usado por usuarios específicos para completar metas de aprendizaje de interacción del sistema	EP01, EP11, EP25, EP27, EP28, EP29
Operabilidad	Grado en el que el producto o sistema tiene atributos para facilitar l control y operaciones.	EP04, EP23
Interfaz de usuario estética	Grado en el que la interfaz del usuario causa satisfacción al interactuar con ella	EP20, EP25
Comportamiento de Tiempo	Grado en que los tiempos y procesamientos de respuesta cumplen con los requisitos	EP03, EP14, EP22
Capacidad	Grado en el que los límites del sistema o parametros del sistema cumplen con los requisitos	EP03, EP04, EP05, EP27
Interoperabilidad	Grado en el que dos o más sistemas, productos o componentes pueden intercambiar información	EP06, EP07, EP12, EP16, EP19, EP22, EP24
Confidencialidad	Grado en el que un producto o sistema permite el acceso a los datos solo a personal autorizado	EP04, EP07
Integración	Grado en el que un sistema, producto o componente previene el acceso no autorizado o modificación de la información	EP20
Exactitud en funcionalidad	Grado en el que el sistema cumple con específicas tareas y objetivos del usuario	EP13
Modificabilidad	Grado en que un producto o sistema puede modificarse de forma eficaz y eficiente sin introducir defectos sin disminuir la calidad	EP14
Localizable	Los datos deben ser fáciles de encontrar tanto para las personas como para las computadoras.	EP04, EP11, EP29

**Tabla A11**  
**POBLACIÓN DE LOS SISTEMAS DE COLECCIONES LOCALIZADOS EN LOS ARTÍCULOS**

Ubicación	Población	Artículos
Arabía	Plantas medicinales	EP05
Australia	Plantas	EP06, EP25
Estados Unidos	Mamíferos, Plantas del bosque	EP16, EP21
Francia	Plantas medicinales	EP13
Indonesia	Plantas del bosque, Plantas medicinales	EP03, EP09
Norte América	Plantas	EP12
Sudáfrica	Mamíferos	EP10



**Figura A1.** Desarrollo del Mapeo Sistemático de la Literatura.



# MATRIZ DE TRAZABILIDAD EN EL LEVANTAMIENTO DE REQUISITOS: UNA REVISIÓN SISTEMÁTICA DE LA LITERATURA

## Traceability matrix in requirements gathering: a systematic review of the literature

Enrique Alexandre Kattz Villa <sup>1</sup>	ekattz44@gmail.com
Patricia Martínez Moreno <sup>2</sup>	pmartinez@uv.mx
José Antonio Vergara Camacho <sup>3</sup>	jvergara@uv.mx
Luis Antonio Morales Velázquez <sup>4</sup>	luismorales04@uv.mx

<sup>1,2,3,4</sup> Universidad Veracruzana.

### RESUMEN

La falta de una matriz de trazabilidad estandarizada tiene varias desventajas para el proceso de desarrollo de software. En primer lugar, sin una matriz de trazabilidad estandarizada es difícil garantizar la calidad y la integridad de los requisitos. Según [5], la trazabilidad de requisitos es necesaria para garantizar que los requerimientos están completos, coherentes y no redundantes. Sin una matriz de trazabilidad estandarizada, los desarrolladores pueden perder de vista los requerimientos originales y hacer cambios no justificados o que no están en línea con los objetivos por una gestión del cambio en el proyecto. De acuerdo con [6], la trazabilidad de requerimientos es esencial para gestionar el cambio en el proyecto, ya que permite a los desarrolladores identificar qué requerimientos se ven afectados por un cambio y cómo se relacionan con otros requerimientos. Sin una matriz de trazabilidad estandarizada, los cambios pueden ser difíciles de gestionar y pueden tener un impacto negativo en el proyecto en su conjunto. Por lo anterior, se realiza la Revisión Sistemática de la Literatura (RSL), se analizan las diferentes propuestas de autores o empresas sobre la Matriz de Trazabilidad de Requisitos (RTM) y se identifican los puntos débiles y fuertes de cada una, con la finalidad de proponer una RTM unificada a partir de los elementos más comunes usados por las RTM analizadas, la cual esté compuesta de fortalezas que abonen a la calidad e integridad de

los requisitos.

**Palabras Clave:** Requerimientos, Matriz de trazabilidad, Trazabilidad, Estandarizada, Análisis.

### ABSTRACT

The lack of a standardized traceability matrix has several disadvantages for the software development process. First, without a standardized traceability matrix, it is difficult to ensure the quality and integrity of the requirements. According to [5], the traceability of requirements is necessary to ensure that these requirements are complete, consistent and not redundant. Without a standardized traceability matrix, developers may lose sight of the original requirements and make changes that are not justified or that are not in line with the goals of change management in the project. According to [6], the traceability of requirements is essential to manage change in a project, as it allows developers to identify which requirements are affected by a change and how they are related to other requirements. Without a standardized traceability matrix, changes can be difficult to manage and can have a negative impact on the project as a whole. The Systematic Literature Review (SLR) is carried out, the different proposals of authors or companies on the Requirements Traceability Matrix (RTM) are analyzed and the weaknesses and strengths

of each one are identified. All of these in order to propose a unified RTM based on the most common elements used by the RTM analyzed, which is composed of strengths that contribute to the quality and integrity of the requirements

**Keywords:** Literature review, slr, requirements, software engineering.

## ► I. Introducción

La trazabilidad es un concepto abstracto en el desarrollo de software y una herramienta de seguimiento en el ciclo de vida de los requerimientos, sin embargo, con una matriz es mucho más fácil rastrear y asegurar el éxito del proyecto, no obstante, existen matrices propuestas de diversos autores por lo que no existe una matriz estandarizada que ayude en general a los desarrolladores en diversos tipos de proyectos, esto lleva a que los desarrolladores tengan dificultades significativas en el manejo de requerimientos, la comunicación entre los equipos de trabajo no sea clara y que se pierda totalmente la evaluación del impacto de los cambios a lo largo del ciclo de vida del software. La presente investigación tiene el propósito de proponer una matriz de trazabilidad estandarizada basada en la recolección y análisis de otras matrices de trazabilidad propuestas por diferentes autores, a través de aplicar la Revisión Sistemática de Literatura. La principal razón de llevar a cabo esta investigación es que, en la ingeniería de software, los requerimientos son parte vital para lograr desarrollar un software de calidad, y en ocasiones los desarrolladores carecen de una herramienta que ayude a cumplir con el seguimiento de estos requerimientos de una manera eficiente. Lo que se espera lograr con esta investigación es que la propuesta de matriz de trazabilidad este compuesta de conceptos que son claves y precisos para el seguimiento de los requisitos, agilizando y haciendo que se produzcan productos de software de calidad.

Las fases de la Revisión Sistemática de Literatura:

1. Definición de la búsqueda. Se generaron preguntas de investigación, los criterios de inclusión y exclusión, y las cadenas de

búsqueda.

2. Ejecución de la búsqueda. En esta sección se obtuvieron los estudios primarios y se presenta una tabla con los resultados de la búsqueda por repositorio.

3. Discusión de los resultados. Apartado en el que se muestran los resultados obtenidos de la RSL, y se describe el análisis y selección hasta llegar a plantear la propuesta de RTM.

## ► II. Conceptualización

Según [2], en su libro *Business Process Management: Profiting from Process* define a la matriz de trazabilidad como:

“Una tabla que se utiliza para rastrear el progreso de los requisitos a medida que se implementan a través del ciclo de vida del proyecto. Permite el seguimiento de cada requisito desde su origen, a través de la implementación, hasta la aceptación y la verificación.”

*A. Importancia de implementar la matriz de trazabilidad en el levantamiento de requisitos.*

La matriz de trazabilidad es clave en la gestión del cambio, establece las relaciones entre los distintos requisitos desde una perspectiva jerárquica siendo la base de la documentación que se produce a lo largo del proceso. El uso de una matriz de trazabilidad es una gran técnica para hacer esto de manera eficiente [3].

Beneficios de implementar la RTM [8]:

- Ayuda a identificar todas las necesidades del cliente al equipo de desarrollo en cada etapa del proceso de desarrollo de software.
- Es muy fácil encontrar las características faltantes o que se pasan por alto.
- Se hace sencillo detectar los casos de prueba que necesitan ser actualizados, ya que son solicitudes de cambio.
- Asegura que todas las solicitudes se reflejan en los casos de prueba.
- Ofrece la certeza al cliente que el producto ha sido desarrollado, según sus necesidades.

De acuerdo con [1] en donde afirmaron que:

“A diferencia de productos en otras ramas de la ingeniería, un producto de software es maleable y pueden cambiar su estructura y composición en espacio y tiempo durante el proceso de desarrollo de software y es aquí donde la trazabilidad adquiere gran importancia”.

El proceso de la matriz de trazabilidad toma su importancia porque proporciona un vínculo entre los requisitos, las capas comerciales y los objetivos clave del proyecto.

De manera similar, se visualiza o rastrea durante el proceso de desarrollo del proyecto para garantizar que el producto entregado funcione según las especificaciones acordadas desde el inicio.

### » III. Método

Se llevó a cabo esta investigación de enfoque cualitativo de alcance exploratorio, de tipo transversal no experimental. Este tipo de investigación cualitativa “se refiere a un conjunto de métodos y técnicas utilizados para recopilar y analizar datos no numéricos o datos que no se pueden cuantificar fácilmente” [7]. Los datos que se recopilaron fueron con base en el análisis, para después formular la propuesta propia.

#### A. Revisión Sistemática de la literatura (RSL).

Para la revisión de la literatura se realizó la búsqueda, selección, extracción y análisis de las diversas matrices de trazabilidad aplicadas en el levantamiento de requisitos dentro del proceso del desarrollo de un producto de software. El estudio, siguió el proceso de la Revisión Sistemática de Literatura desarrollada por [9] por su enfoque hacia la Ingeniería de Software.

Inicialmente, se definieron las preguntas de investigación (PI), cadenas de búsqueda, y las bases de datos donde se implementaron y los criterios de inclusión (CI) y criterios de exclusión (CE) correspondientes a la etapa de planificación.

Se establecieron dos PI basadas en los indicadores

que se pueden obtener con el análisis de Matrices de Trazabilidad.

**PI1:** ¿Qué elementos integran una matriz de trazabilidad?

**PI2:** ¿Por qué contar con un artefacto de matriz de trazabilidad en el levantamiento de requisitos del software?

**PI1:** Identificar la información base, los elementos repetitivos y/o comunes de las matrices de trazabilidad ocupadas utilizadas en el levantamiento de requisitos. Con respecto a la PI2, se establece con el propósito de conocer los temas relevantes de implementar la Matriz de Trazabilidad en el proceso de desarrollo de software, específicamente en el levantamiento de requisitos.

Posteriormente, se definieron las cadenas de búsqueda por aplicar en las Bases de Datos: Google Académico, IEEE XPLORE y ACM DIGITAL LIBRARY, llevándose a cabo durante el periodo de búsqueda de 4 meses a finales del año 2022:

Cadena 1: (“Matrix traceability”)

Cadena 2: (“Traceability matrix in requirements gathering” OR “Traceability Matrix in software project development”).

Ahora bien, para cada cadena de búsqueda aplicada, el resultado obtenido fue filtrado y revisado conforme a los CI y los CE:

- CI1: El documento está relacionado con la trazabilidad en el software.
- CI2: El título y resumen hacen referencia al seguimiento de los requisitos a través de cada una de las etapas de vida del desarrollo de software.
- CI3: El resumen trata sobre la trazabilidad de requerimientos en el desarrollo de aplicaciones web.
- CE1: El resumen del escrito no trata sobre la ingeniería de requisitos.
- CE2: El título del artículo no responde ninguna

pregunta de investigación.

- CE3: El acceso al documento requiere de pagar una membresía.

El CI1, está estrechamente ligado a la ingeniería de requisitos ya que se especifica que la trazabilidad se aplica directamente en el desarrollo del software, toda vez la trazabilidad está ligada a los requisitos de un proyecto.

El CI2, se relaciona con el título y/o resumen del artículo, siempre y cuando en este mismo se mencionara el aplicar seguimiento a los requisitos.

El CI3, es conforme a los documentos o estudios orientados al desarrollo web, esto con la finalidad de ampliar el enfoque de estudio en cuanto a las matrices presentadas.

El CE1, aquellos resúmenes no enfocados a la ingeniería de requisitos del desarrollo de software se excluyeron.

El CE2, este criterio de exclusión aplicado con base en el título como una forma de filtrar de manera rápida los resultados obtenidos de las cadenas de búsqueda, ya que ocasiones el título se enfoca a otro tipo de trazabilidad.

El CE3, son aquellos documentos de las bases de datos que requieren una cuota de pago para el acceso a los mismos, por lo que fueron descartados.

Con base en lo anterior, en la tabla I se presenta el número total de 1344 documentos recuperados de las bases de datos al utilizar la cadena 1. Se destaca la ACM DIGITAL

LIBRARY con el mayor número de hallazgos con 1102 documentos, representando el 82% del total de los documentos. En menor cantidad se encuentran IEEE XPLORE y GOOGLE ACADÉMICO con 160 y 82 documentos respectivamente. Con respecto a la cadena 2, GOOGLE ACADÉMICO arrojó 72 resultados, por arriba de IEEE XPLORE y ACM DIGITAL LIBRARY con 35 y 10 resultados de búsqueda, respectivamente. Se observa que existe una cantidad considerable de

bibliografía que hace alusión o mencionan en sus documentos la palabra “Matriz de Trazabilidad” sin embargo para efectos de esta investigación se requirió que mostrarán exactamente una matriz de trazabilidad para analizar sus elementos, por lo que al final de la RSL se trabajaron sólo con 16 documentos.

Los resultados de las cadenas de búsqueda dentro de las tres bases de datos fueron las siguientes:

**Tabla I**  
**Resultados de las cadenas de búsqueda en los repositorios**  
**Google Académico, IEEE XPLORE y ACM DIGITAL**  
**LIBRARY.**

Base de datos	Resultados de Cadena 1	Seleccionados	Resultados de Cadena 2	Seleccionados
GOOGLE ACADÉMICO	82	1	72	9
IEEE XPLORE	160	6	35	4
ACM DIGITAL	1102	7	10	3
LIBRARY				
TOTAL	1344	14	117	16

#### ► IV. Resultados

Los resultados obtenidos del análisis de las matrices analizadas se derivan en una propuesta de autoría propia, que para llegar al resultado se siguieron tres pasos:

1. Priorizar los elementos repetidos: requerimiento, casos de prueba y situación del requerimiento, con base en ello se decidió incluirlos en la propuesta ya que varios autores consideran importante su inclusión.
2. En segundo lugar, se hizo hincapié en el seguimiento tanto de los artefactos generados como de casos de prueba, debido a que son fundamentales en los requerimientos al aplicar un cambio e identificar de manera correcta donde se está situado.
3. Por último, se agregaron elementos basados en el estatus general del requerimiento como: prioridad, estado, situación del requisito, responsable y fecha

de entrega, esto con la finalidad que el equipo de desarrollado no pierda el panorama con respeto al tiempo del desarrollo del software.

En respuesta a los resultados obtenidos de la PI1 ¿Qué elementos integran una matriz de trazabilidad? Para [4] “cada proyecto debe diseñar su propia herramienta en función de los requisitos establecidos en las fases de diseño y de la complejidad de la ejecución de las tareas”, sin embargo, se encontró que varias matrices coincidían en una serie de elementos:

- **Identificación o ID:** Normalmente corresponde a un código o número asignado a cada requisito del proyecto.
- **Interesado o Destinatario:** La persona que realiza la solicitud o pedido. También llamada parte relacionada o stakeholder.
- **Prioridad:** No todas las solicitudes tienen la misma importancia. Los hay primarios, secundarios y terciarios, dependiendo del valor que represente el proceso global.
- **Descripción del requisito:** Describe brevemente el contenido y las características clave.
- **Versión:** La matriz debe detallar la evolución del requisito. Debido a esto, se configuran diferentes versiones hasta que se termina.
- **Complejidad:** los requisitos suelen definirse como de complejidad alta, baja o media, según los recursos, el tiempo o el trabajo necesarios para completar el requisito.
- **Estrategia de prueba:** la matriz de trazabilidad también debe establecer el método utilizado para probar diferentes versiones del requisito hasta que el requisito se ejecute y verifique por completo.

En respuesta a la segunda pregunta de investigación PI2 ¿Por qué contar con un artefacto de matriz de trazabilidad en el levantamiento de requisitos del software?, se tienen las siguientes ventajas:

- Ayuda a identificar todas las necesidades del

cliente al equipo de desarrollo en cada etapa del desarrollo de software.

- Es muy fácil encontrar las características que faltan.
- También es muy fácil detectar los casos de prueba que necesitan ser actualizados, ya que son solicitudes de cambio.
- Asegura que todas las solicitudes se reflejan en los casos de prueba.
- Aseverar al cliente que el producto ha sido desarrollado, según sus necesidades.

Para la propuesta final se tomó en cuenta los principales elementos repetidos en las matrices como, por ejemplo: Identificación o ID en este caso se decide cambiar de nombre a Requerimiento ya que es un nombre más sensato, posteriormente Casos de Prueba asociados que si bien no aparece repetidamente en los estudios analizados es un aspecto importante ya que sirve para saber desde que punto un requerimiento debe ser cambiado, en general ayuda a la gestión del cambio. Prioridad, este campo si aparece en los estudios, sin embargo, se decidió cambiar las denominaciones para que fueran más generales y entendibles (Alta, Media y Baja). Artefactos Asociados, campo que muestra los artefactos generados que son ligados al requerimiento, si bien tampoco aparece repetidamente se decidió colocar para que sea de guía en cuanto a la documentación y si esta requerir ajustarse a futuro sería más fácil rastrear lo que se debe modificar. Estado, este campo es especialmente útil en caso de contar con una cantidad grande de requisitos manipulados por una sola persona, lo que resulta benéfico al llevar un mejor control. Situación del Requerimiento asiste a los líderes de proyectos a evaluar el requerimiento y el avance de este. Responsable, este campo se decidió agregar ya que regularmente varias personas están designadas a un mismo proyecto y en ocasiones este tiene muchos requerimientos, de tal manera ayuda al desarrollador encargado en la localización rápida de los mismos. Fecha de entrega, último campo de la propuesta que controla los tiempos de entrega,

enfocado en la puntualidad y cumplimiento de los requisitos, lo que da como beneficio la satisfacción del cliente y que no se incrementen los costos por lo que se decidió agregar para que se tenga visible por todo el equipo desarrollar la fecha límite de entrega, esta fecha es la definitiva en la que el requerimiento ya está listo para ser implementado sin fallas ni detalles.

Como producto final se obtuvo la siguiente propuesta de la RTM (Figura 1), en la que se pretende que los stakeholders, el equipo de desarrollo junto con el jefe de proyecto tenga acceso, estos dos últimos con opción para editar los campos de la matriz, sin embargo, los stakeholders aseguran que su producto está en desarrollo

Nombre del proyecto							
Requerimiento	Casos de prueba asociados	Prioridad	Artefactos asociados	Estado	Situación del requerimiento	Responsable	Fecha de entrega
RF01	ID del caso de prueba y el nombre	Alta Media Baja	CU001 Iniciar sesión	Activo Inactivo Cancelado	En espera Entregado Aceptado	Persona encargada de ejecutar el requisito	Límite de entrega

Fig. 1. Matriz de trazabilidad estandarizada.

Conceptos claves que responden la PI1:

- **Requerimiento:** Es el identificador (ID), nombre con el que se identifica al requerimiento. Ejemplo: RF01. El sistema permite iniciar sesión al empleado.
  - **Casos de prueba asociados:** son los que se asocian al funcionamiento del requerimiento, se recomienda expresar el ID del caso de prueba y el nombre de este.
  - **Prioridad:** Campo ligado a tres palabras claves para valorar la atención que se le quiere dar a un requisito.
- Prioridad: Alta, media y baja, considera a cuál actividad se le dará atención inmediata en comparación a otras los que conlleva a la toma de mejores decisiones dentro del proyecto.
- **Artefactos asociados:** son aquellos artefactos que están ligados al requerimiento, pueden ser: casos

de uso, diagramas de secuencia, de comunicación, etc. Expresados con ID o clave del artefacto y un nombre corto pero descriptivo. Ejemplo: CU001 Iniciar sesión.

- **Estado:** Hace referencia al estado en el que se encuentra el requerimiento, este puede ser “Activo” siendo lo más común, pero también puede ser “inactivo” o “cancelado”, debido a que el requerimiento puede ser cancelado o cambiado por los stakeholders.
- **Situación del requerimiento:** Esto hace referencia a tres situaciones que puede estar el requerimiento: “en espera” (y puede referirse a que está en desarrollo), “entregado” (se refiere a que fue entregado al jefe de proyecto o equipo de trabajo para ser revisado) y, por último, “aceptado” (esto quiere decir que fue aprobado y está finalizado).

- **Responsable:** Es el nombre de la persona encargada de llevar a cabo todas las acciones necesarias para ejecutar el requerimiento.
- **Fecha de entrega:** Es la fecha límite que se tiene para entregar el requerimiento.

Con respecto a la PI2, fue respondida, en el apartado 1 y 2 de este trabajo.

## » V. Conclusión

La matriz de trazabilidad proporciona un medio eficiente para garantizar la integridad y la consistencia de los requerimientos, lo que permite identificar cualquier brecha o superposición en los requisitos, evitar inconsistencias y ambigüedades que afecten negativamente la calidad del producto final. Al visualizar las relaciones entre los diferentes componentes se facilita la detección temprana de problemas y se mejora la comunicación entre los diferentes interesados del proyecto.

Además, la matriz de trazabilidad permite rastrear los cambios y las actualizaciones en los requerimientos a lo largo del desarrollo de un producto de software. A medida que se realizan modificaciones en los requisitos se sugiere utilizar la matriz para evaluar el impacto de estos

cambios en otros elementos del sistema, esto ayuda a gestionar de manera efectiva los cambios y minimiza el riesgo de introducir errores o inconsistencias en el proceso.

En este análisis de matrices de trazabilidad se identifica la necesidad de esta, y a pesar de ello son mínimos los equipos de desarrollo de software que la implementan como artefacto en la etapa del levantamiento de requisitos. Finalmente, al aplicar la RSL se consiguió identificar cada uno de los elementos comunes y repetitivos en las diversas matrices de trazabilidad, con los cuales, estos se unificaron de tal manera que se logró proponer una RTM estandarizada con la finalidad de aportar un artefacto confiable para la industria del software en el proceso de levantamiento de requisitos.

En definitiva, la implementación de una matriz de trazabilidad estandarizada es una inversión clave para cualquier proyecto de levantamiento de requisitos ya que permite una mejor planificación, seguimiento y control, lo que se traduce en un proyecto más eficiente, eficaz y exitoso.

## ► VI. Agradecimientos

A cada uno de los investigadores anónimos que fungieron como revisores de este trabajo, por sus comentarios y aportaciones para mejorarlo.

## ► VII. Referencias

- [1] Anaya, R., Tabares, M., & Arango, F. Una revisión de modelos y semánticas para la trazabilidad de requerimientos. *Revista EIA*, 2006.
- [2] Burlton, R. *Business Process Management: Profiting from Process*, 2001.
- [3] Carvajal, M. M. *La Trazabilidad en el Proceso de Requerimientos de Software*. Heredia, San José, Costa Rica, 2019.
- [4] García, P. Matriz de trazabilidad de requisitos (RTM), 2021. Obtenido de <https://visuresolutions.com/es/requirements-traceability-matrix>
- [5] Gotel, O., & Finkelstein, A. An Analysis of the Requirements Traceability Problem. *International Conference on Requirements Engineering*, 94-10, 1994.
- [6] Hebig, & Brinkkemper. *Improving Software Product Management Processes: a detailed view of the Product Software Knowledge Infrastructure*, 2010.
- [7] Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. *Metodología de la investigación* (4ta. ed.). México D.F.: McGraw-Hill, 2003.
- [8] Hassnanin, M. A comparative study on traceability Approaches in software development. *ITEE Journal*, IV, 1-4, 2015.
- [9] Kitchenham, B., Budgen, D., Brereton, P., Turner, M., Charters, S., & Linkman, S. Large-scale software engineering questions—expert opinion or empirical evidence? *IET software*, 1(5), 161-171, 2007.





# IMPLEMENTACIÓN DE UN CONTROLADOR PARA ROBOT DE TIPO PÉNDULO INVERTIDO

## Implementation of a controller for an inverted pendulum type robot

Josepablo Cruz Baas <sup>1</sup>	jcruzbaas@gmail.com
Jorge Ríos Martínez <sup>1</sup>	jorge.rios@correo.uady.mx
Ernesto Olguín Díaz <sup>2</sup>	ernesto.olguin@cinvestav.edu.mx
Francisco Moo Mena <sup>1</sup>	mmena@correo.uady.mx

<sup>1</sup> Universidad Autónoma de Yucatán (Facultad de Matemáticas).

<sup>2</sup> CINVESTAV (Unidad Saltillo).

### RESUMEN

En la robótica móvil de interiores sobresalen los robots con ruedas porque su mecánica es más simple y requiere menor trabajo de mantenimiento comparados con los robots articulados de piernas. En este trabajo se describen las fases para la implementación de un software de control para un robot péndulo invertido usando el algoritmo regulador cuadrático lineal (*LQR*, por sus siglas en inglés). La arquitectura del software se puede descomponer por grandes bloques de abstracción, combinando bibliotecas de software embebido y piezas de software desarrolladas especialmente para el proyecto. La gran mayoría de proyectos con el péndulo invertido utilizan un control PID, principalmente debido a la simplicidad de su implementación. Este método es funcional, pero poco escalable, consume tiempo y, aunque algunas implementaciones presentan un excelente rechazo a perturbaciones externas, no existe una garantía de estabilidad si algún parámetro del robot cambia. El prototipo desarrollado podría servir como base de experimentación para futuras investigaciones sobre robótica móvil o control automático.

**Palabras Clave:** Péndulo invertido; software de control; *LQR*

### ABSTRACT

In indoor mobile robotics, wheeled robots stand out due to their simpler mechanics and lower

maintenance requirements compared to legged articulated robots. This work describes the phases for implementing control software for an inverted pendulum robot using the Linear Quadratic Regulator (*LQR*) algorithm. The software architecture can be decomposed into large abstraction blocks, combining embedded software libraries and software components developed specifically for the project. Most inverted pendulum projects use a PID control, primarily due to the simplicity of its implementation. This method is functional but lacks scalability, consumes time, and, although some implementations exhibit excellent disturbance rejection, there is no guarantee of stability if any robot parameter changes. The developed prototype could serve as an experimental foundation for future research in mobile robotics or automatic control.

**Keywords:** Inverted pendulum; control software; *LQR*

### ► I. Introducción

La robótica se ha convertido en una herramienta ubicua en varios campos de la industria y la sociedad, como una tecnología que brinda enormes beneficios económicos y *sociales*. En la robótica móvil de interiores sobresalen los robots con ruedas porque su mecánica es más simple y requiere menor trabajo de mantenimiento

comparados con los robots articulados de piernas.

El robot péndulo invertido, robot balancín o *two wheeled inverted pendulum robot*, en inglés, es un robot con dos ruedas que se puede mover de forma independiente alrededor del mismo eje de rotación, obteniendo así una alineación de sentido paralelo. El modelo del péndulo invertido (fig. 1) permite analizar el problema de equilibrio que presenta ese robot porque el cuerpo del robot puede abstraerse como el péndulo, y las ruedas del robot como el eje de rotación del péndulo, de ahí el nombre del robot [1].

Métodos clásicos de control bastan para estabilizar el ángulo de inclinación (o balanceo) del robot, pero no para estabilizarlo mientras se controla. Es un sistema sub-actuado porque hay más variables a controlar que acciones de control. Para conseguir un control preciso tanto de traslación como de inclinación, es necesario un controlador de retroalimentación de múltiples salidas. En este trabajo se describen las fases para la implementación de un software de control para un robot péndulo invertido usando el algoritmo regulador cuadrático lineal (*LQR*, por sus siglas en inglés). El método *LQR* proporciona ganancias de retroalimentación controladas de manera óptima para permitir el diseño de sistemas de alto rendimiento y estabilidad en circuito cerrado.

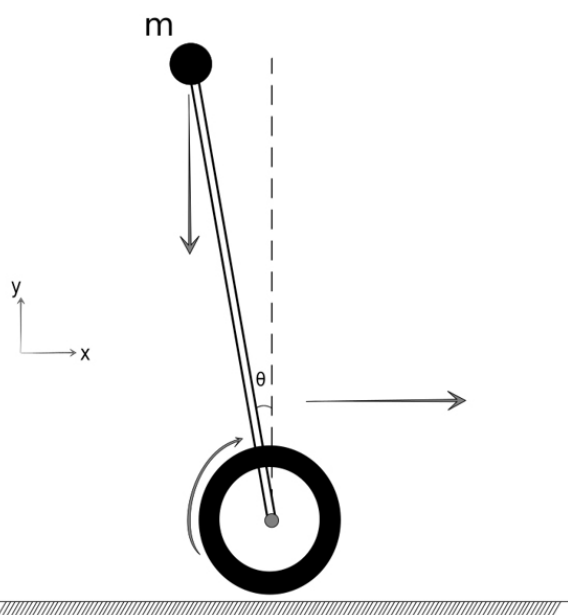


Fig. 1. Modelo del péndulo invertido sobre una rueda o monociclo.

## ► II. Antecedentes

La mayoría de los proyectos que controlan de manera automática un mecanismo de tipo péndulo invertido utilizan un algoritmo de control clásico, el proporcional, integral y derivativo (PID, por sus siglas en inglés). El PID es simple de implementar y no requiere un modelo dinámico para sintonizar sus ganancias permitiendo estabilizar al robot a través de prueba y error. Este método es funcional, pero poco escalable, consume tiempo y, aunque algunas implementaciones presentan un excelente rechazo a perturbaciones externas, no existe una garantía de estabilidad si algún parámetro del robot cambia (por ejemplo: reemplazar las ruedas, o añadir/quitar algún peso). Para dar garantías de estabilidad se requiere analizar un modelo matemático, por ejemplo, el *modelo matemático simplificado de 2 grados de libertad* para el péndulo invertido (fig. 1) reduce el problema a un sistema holonómico cuyo espacio de configuraciones puede ser descrito usando una variable de posición unidimensional y una variable de inclinación del cuerpo del robot.

Según los trabajos revisados, las ecuaciones que describen la dinámica de esos sistemas se obtienen de los fundamentos de mecánica de Newton y de la mecánica de Lagrange. El libro [2] postula que el modelado a través de coordenadas generalizadas de la mecánica Lagrangiana es mucho más simple cuando la cantidad de cuerpos rígidos en el modelo es considerable. En la tabla I, se encuentran los trabajos revisados clasificados por grados de libertad y fundamentos de mecánica clásica.

Existen propuestas nuevas de control utilizando redes neuronales que muestran buenos resultados [12],[16]. Sin embargo, se requiere de un proceso de entrenamiento para representar la dinámica del robot sin certeza de la precisión del modelo, y la capacidad de cómputo necesaria para ejecutar el control en tiempo real supera las capacidades de un microcontrolador aumentando el costo del robot.

De igual forma, existen métodos de control no lineal que aumentan considerablemente la precisión del controlador incluso para el sistema

de 4 grados de libertad [3],[5],[6], pero requieren inevitablemente un filtro tipo Kalman para garantizar un control robusto.

Los controladores que se pueden implementar en un microcontrolador son: el controlador PID, el controlador LQR y el controlador de lógica difusa [17].

Fundamentos de mecánica	Modelo de 4 grados de libertad	Modelo de 2 grados de libertad
Lagrange	[3],[4],[5],[6],[7]	[8],[1],[9]
Newton/Euler	[10]	[11],[12],[13],[14],[15]

Con respecto a los sistemas embebidos de control en la literatura se encontraron trabajos basados en un procesador tipo ARMTM como la tarjeta Beagle BoneTM o la tarjeta Raspberry PiTM [5],[14],[16]. Existen ejemplos de sistemas corriendo en un ArduinoTM en [12],[13],[15],[18]-[20], en Lego MindstormsTM [21] o incluso en tarjetas FPGA [9],[10].

### ► III. Diseño del robot

El robot RAP (fig. 2), construido para este proyecto, posee hardware electrónico que emplea componentes profesionales de bajo costo. Es un prototipo de robot móvil compuesto por un módulo de control (MC) y un módulo de inteligencia básica (MIB). El módulo de control, como su nombre lo indica, es la placa de control, un sistema embebido basado en un microcontrolador de arquitectura ARM<sup>TM</sup> que interactúa directamente con los sensores y actuadores esenciales del robot (acelerómetro, giroscopio, magnetómetro y encoders). Dicho sistema tiene la obligación de estabilizar el balanceo del robot y convertir comandos en movimiento. El módulo de inteligencia básica es el módulo que gobierna al MC considerando que el sistema es estable y controlable. Está basado en un microprocesador de arquitectura ARM<sup>TM</sup> con sistema operativo Linux. Envía comandos de acción para que el robot RAP se traslade. Además, el módulo de inteligencia básica se puede comunicar con otros sistemas computacionales para escalar las habilidades del robot. Este diseño divide los requerimientos en

módulos independientes y autónomos, siendo el pilar del sistema el módulo de control; sin él, no es posible convertir el sistema físico en un robot móvil.

El presente trabajo se enfoca en la primera parte del proyecto, el desarrollo del controlador automático para el MC del robot RAP. Se sintetiza un modelo simplificado para describir la dinámica de la posición, la velocidad y el ángulo de inclinación del RAP, linealizándolo para un rango seguro de operación y estabilizándolo con un controlador LQR de tiempo discreto. Al final, se desarrolla un prototipo robótico péndulo invertido de dos ruedas diferenciales con un sistema de control embebido, que podrá servir como base de experimentación para futuras investigaciones sobre robótica móvil o control automático.

#### A. Diseño mecánico del robot

El robot RAP se compone de un carro, un cuerpo y las ruedas. El carro es la estructura que sujeta las ruedas, la batería, la electrónica de potencia y la electrónica de control. Está conformado por elementos de acrílico cortados a láser. Las ruedas del robot son ruedas de patín o scooter, con un diámetro de 144 milímetros, un grosor de 29 milímetros y un peso de 235 gramos. Cada una se encuentra acoplada al eje del reductor de un motor de corriente directa de 12 volts. Ambos motores comparten especificaciones: el reductor tiene una razón de 30:1, poseen un encoder digital de efecto hall de 64 unidades por revolución, acoplado al eje del motor antes del reductor. Ambos motores sin algún tipo de carga alcanzan una velocidad de 330 rpm con un consumo de 200 miliamperios, [mA]. Por otro lado, los motores entregan un par de fuerza de hasta 14 kilogramos por centímetro con un consumo de 5.5 amperios (esta carga es el límite máximo de los motores, el cual podría dañarlos), los motores son de la marca Pololu [22].

El robot con el cuerpo acoplado tiene una altura de 67 centímetros, sin el cuerpo tiene una altura de 20 centímetros; ambas medidas a partir del piso. El centro de masa del sistema con cuerpo tiene una altura de 17.27 centímetros a partir del eje de las ruedas, con un peso de 3.071 kilogramos.

Tabla II. Variables del modelo matemático con valores medidos y estimados		
Variable	Nombre	Valor
$m_b$	Masa del cuerpo del robot	3.071kg
$m_w$	Masa de ambas ruedas	0.470 kg
$I_b$	Inercia del cuerpo	0.091593 kg m <sup>2</sup>
$I_w$	Inercia total de ambas ruedas	0.056870 kg m <sup>2</sup>
$R$	Radio exterior de ambas ruedas	0.072 m
$l_m$	Altura del centro de masa del cuerpo	0.1727 m
$\mu_d$	Coefficiente de fricción cinética del piso	0.8
$g$	Aceleración de la gravedad	9.8 m/s <sup>2</sup>

**B. Diseño eléctrico y electrónico del robot.**

El sistema *RAP* necesita 3 tensiones diferentes de corriente directa para funcionar: 3.3, 5 y 12 volts. La electrónica de potencia consume 12 volts. Para alimentar el sistema se utiliza una batería LiPo de la marca Wild Scorpion. La batería es de 4 celdas, con una capacidad de 4200 miliamperios hora, [mAh]. El voltaje nominal de la batería es de 14.8 volts, sin embargo, el voltaje con carga completa es de 16 volts, y 14 volts cuando se ha descargado. La capacidad de carga es de entre 5 y 10 C, la capacidad de descarga es de hasta 40C [1C = 4.2 A]. Para alimentar la electrónica digital se tiene un regulador de tipo conmutado que baja la tensión de 12 volts a 5 volts, y en la placa de control se cuenta con un regulador extra de tipo lineal, para bajar la tensión de 5 volts a 3.3 volts.

Para gobernar ambos motores se requieren dos puentes H. de la marca *Freescale Semiconductor*<sup>TM</sup> modelo *MC33926*. Estos circuitos integrados soportan una salida de hasta 5 amperios cada uno. Cuentan con un monitor de corriente para que el microcontrolador, a través del ADC (convertidor de analógico a digital), pueda sentir la corriente que es entregada a los motores.

El control de potencia se hace con dos canales PWM, uno para cada dirección de giro. La frecuencia máxima de estas señales, que el circuito integrado soporta, es de 20 kHz.

La electrónica digital de control del robot *RAP*, consiste en un microcontrolador de la marca *Texas Instruments*<sup>TM</sup>, modelo *Tiva C TM4C123GH6PM*, un sensor de inercia con acelerómetro, giroscopio y magnetómetro de la marca *InvenSense*<sup>TM</sup> modelo

*MPU-9250*, y un módulo puente de comunicación UART a bluetooth, de la marca *Guangzhou HC Information Technology*<sup>TM</sup>, modelo *HC-06*. El microcontrolador *TM4C123G* es un microcontrolador de arquitectura *ARM*<sup>TM</sup> *Cortex-4MF*, de 32 bits con unidad de coma flotante, acompañado del sistema operativo en tiempo real *TI-RTOS*, propiedad de *Texas Instruments*<sup>TM</sup> y de libre uso para los microcontroladores de su marca.

» **IV. Diseño del controlador**

Esta sección pretende ilustrar el proceso de diseño del controlador, sin embargo, por falta de espacio, importantes detalles del análisis matemático no se incluyeron, pero están disponibles enviando un correo al autor correspondiente. El modelo matemático para trabajar es un sistema lineal de ecuaciones diferenciales conocido como representación en el espacio de estado.

**A. Modelo dinámico**

A partir del modelo lineal del sistema se definen las siguientes variables para simplificar el análisis.

$$\begin{aligned}
 H_1 &\triangleq ((m_w + m_b) R^2 + I_w) \\
 H'_2 &\triangleq m_b l_m R \\
 H_3 &\triangleq l_b \\
 D_1 &\triangleq 2R\mu_d \\
 g_2 &\triangleq -m_b g l_m
 \end{aligned}$$

Sean  $\vec{x}$  el vector de estado y  $\vec{y}$  el vector de salida. La representación del modelo dinámico en el espacio de estado es la ecuación (1).

$$\vec{\dot{x}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{-g_2 H'_2}{h_1} & \frac{H_3 D_1}{h_1} & 0 \\ 0 & \frac{g_2 H_1}{h_1} & \frac{-H'_2 D_1}{h_1} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \theta \\ \dot{\alpha} \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{H_3}{h_1} \\ \frac{H'_2}{h_1} \end{bmatrix} u$$

$$\vec{y} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \theta \\ \dot{\alpha} \\ \dot{\theta} \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \theta \\ \dot{\alpha} \\ \dot{\theta} \end{bmatrix}$$

En la tabla II se muestran las variables obtenidas mediante mediciones físicas del robot y sus piezas. Excepto el coeficiente de fricción, el cual es un valor estimado por tablas y condiciones relativas.

El análisis se restringió a un espacio de configuraciones de dos grados de libertad (fig. 1), cuyas variables son: la posición unidimensional del robot *RAP*, y el ángulo de inclinación del cuerpo del robot *RAP*. Para que dichas restricciones tengan sentido, se asume que las ruedas diferenciales del robot *RAP* rotan exactamente a la misma velocidad.

Con el modelo lineal del sistema en representación de estado obtenido para una región de operación conocida, se calcula el rango de la matriz de controlabilidad, para saber si el sistema es controlable. La región de operación del robot *RAP*, consiste en definir el rango admisible del ángulo de inclinación del robot, el cual se define en la vecindad alrededor del cero con variaciones muy pequeñas.

Se pretende ahorrar la mayor cantidad de recursos de cómputo en el sistema embebido para poder construir sobre él otra clase de servicios además del control de la planta, por ejemplo, un muestreador en tiempo real y un sistema de configuración. El proyecto mínimo viable consistió en estabilizar la posición, velocidad y ángulo de inclinación del cuerpo, así que, ponderando los requerimientos, el sistema de control cuya implementación en software embebido es más simple y ligera, es el sistema de control *LQR*, pues consiste en una única multiplicación del vector de estado  $\vec{x}$  y la matriz de control  $K$  en cada periodo de control. Utilizando *LQR* y definiendo dos matrices  $Q$  y  $R$  para la función de costo, de forma que se maximice el tiempo de respuesta y se minimice la energía de la señal de control, se encuentra el valor de la matriz de control  $K$  que estabiliza el sistema dinámico linealizado.

Posterior a esto, utilizando el software *Octave*, se codifica una simulación para el sistema lineal del robot *RAP*. Esto es, ingresar las ecuaciones del sistema en representación de estado a *Octave* (tanto en lazo abierto (LA) como en lazo cerrado

(LC)), asignar un vector de estado inicial donde exista un ángulo de inclinación diferente de cero, graficar las respuestas del sistema sin controlador y con controlador, y comparar las respuestas.

Antes de implementar el sistema en el *RAP*, es necesario discretizarlo para observar los efectos de una frecuencia de control baja y garantizar que el sistema será estable a frecuencias determinadas por el programador. Para discretizar el sistema de control, se hace uso de la relación entre la variable  $x$  y la variable  $z$ , propuesta por el método bilineal (también conocido como Tustin). Se asigna un periodo de muestreo adecuado, y se programa al robot *RAP* para emitir por el puerto serial los valores de las variables de estado en tiempo real, capturadas en dicho periodo de control. Esta información produce las gráficas del comportamiento dinámico del sistema real previo al estabilizador.

Para obtener el vector de estado del sistema dinámico, se hace uso de un sensor de inercia MPU9250 digital (que cuenta con un acelerómetro, un giroscopio y un magnetómetro) y un encoder en cada motor del robot *RAP*. La estimación del estado se consigue a través de un filtro complementario, para la estimación del ángulo de inclinación del cuerpo del robot *RAP* y el promedio de la velocidad de las ruedas del robot *RAP*, para la estimación de la posición y velocidad del robot.

La intención del controlador es estabilizar las variables  $\alpha$ ,  $\dot{\alpha}$ ,  $\theta$  y  $\dot{\theta}$ , correspondientes a la posición y velocidad del carro, y la inclinación del péndulo. Por lo tanto, el sistema de control se simplifica dado que la entrada del sistema es el origen del espacio de estado.

La arquitectura del sistema de control se presenta en la figura 3.

En el diagrama se considera que el sistema ya está discretizado, por lo tanto, las entradas y las salidas de cada bloque son digitales, con un periodo de muestreo de  $T_s = 0.010$  segundos. El bloque denominado *muestreador*, es el bloque que almacena las variables de estado y la salida

del sistema en tiempo real para compartirlas con el sistema de comunicación. El control de torque de los motores convierte la señal de torque  $T_u$  resultado del controlador  $LQR$  y produce un torque efectivo  $U_k$  que cambia el estado dinámico de la planta. El control de motores se asume lineal para este proyecto, de forma que  $T_u = U_k$ .

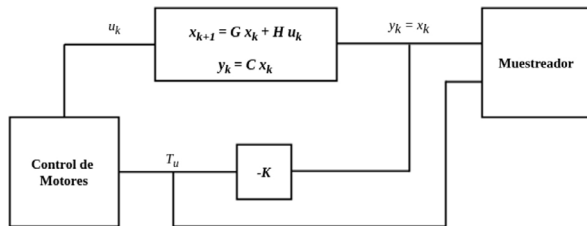


Fig. 3. Diagrama a bloques de la propuesta de control discreto.

El microcontrolador *Tiva C TM123* cuenta con dos unidades de medición para encoders, lo que permite independizar al CPU principal del proceso de muestreo de las ruedas. A su vez, de entre todos los periféricos interesantes que posee, sobresalen para el proyecto, un conjunto de buses IIC y buses UART. La arquitectura del  $\mu C$  permite implementar un sistema operativo en tiempo real *Texas Instruments* ha implementado sobre esta arquitectura su propio RTOS de nombre *Ti-RTOS*, que además de proveer interfaces para la administración lógica de tiempos y recursos, también posee controladores para todos los periféricos internos, éstos controladores son parte de una biblioteca llamada *TivaWare*, la cual viene instalada en una ROM dedicada en el propio  $\mu C$ , lo que acelera el proceso de desarrollo de prototipos.

En este proyecto se ha utilizado tanto la biblioteca como el *RTOS* para agilizar el prototipado, compilando el firmware desde el entorno de desarrollo *Code Composer Studio*, el cual también es propiedad de *Texas Instruments*.

### ➤ V. Arquitectura de software

Como se puede apreciar en la fig. 4, el software se puede descomponer por grandes bloques de abstracción. Los dos bloques más bajos son todas las bibliotecas de software embebido provistas por el fabricante, mientras que las capas superiores, son piezas de software desarrolladas especialmente

para el proyecto.

Las piezas más importantes, son los controladores de periféricos externos del robot. Éstos son los drivers que mueven a ambos motores y permiten codificar y decodificar los datos del sensor IMU. Estos controladores se implementan a través de los drivers internos provistos por el fabricante.

La siguiente capa consiste en el sistema de control de lazo cerrado o simplemente sistema de control, y el sistema de medición de variables de ambiente o entorno, también llamado *environment*. Estas piezas de software comparten el mismo nivel porque interactúan directamente con los drivers del robot.

La capa más alta del sistema está dedicada al sistema de comunicación, aquí los procesos sólo consisten en intercambio de información entre el microcontrolador y la computadora externa.

A continuación, se describen brevemente las 3 piezas de software que dan vida al proyecto.

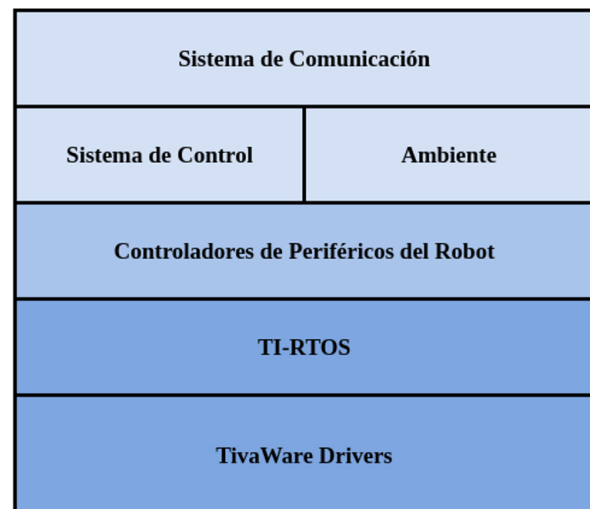


Fig. 4. Bloques lógicos de la arquitectura del software, la posición de abajo hacia arriba, indica el nivel de abstracción, donde el nivel más alto corresponde a flujo de datos, y el nivel más bajo, a comunicación y control del hardware.

#### A. Sistema de medición del entorno

Este sistema se encarga de leer cada sensor con una frecuencia de 2 milisegundos, que posteriormente procesa para construir una estructura privada denominada variables de entorno. Para que

cualquier proceso, externo al sistema de entorno, pueda acceder a esta estructura, se realiza una petición a través de la interfaz *Environment\_read(variableID, ptrBuffer)*, misma que desactiva momentáneamente las interrupciones del CPU para copiar dichos datos desde la estructura hasta el buffer. La función de lectura es de tipo no reentrante<sup>1</sup> y, de no ser por el bloqueo de interrupciones, podría incurrir en una condición de carrera<sup>2</sup> que dañaría la estructura de datos, afectando al sistema de control.

### B. Sistema de control de lazo cerrado

El sistema de control realiza una petición al sistema de entorno para leer las variables de estado estimadas, realiza la corrección de signo en el ángulo de inclinación y hace un llamado al *controlLoop* y al sistema de muestreo.

El *controlLoop*, es una función sin argumento y que no retorna nada, su propósito es actualizar la variable local del sistema de control  $u$ , la cual almacena la señal que deberá ser enviada a cada motor. El cálculo de la señal,  $u$ , es el resultado de multiplicar el vector de estado por la matriz de ganancias (resultado de la propuesta de control) y acondicionarlo usando el factor de conversión  $Km = 0.00291566$ .

El sistema de muestreo consiste en rellenar una trama y agregarla a una cola a través del sistema de comunicación. Este proceso se activa o desactiva con una petición desde el sistema de comunicación.

Tanto el *controlLoop* como el sistema de muestreo se ejecutan en el periodo de control; esto es cada 10 milisegundos. La prioridad de este proceso es la más alta de todo el firmware.

## » VI. Sistema de comunicación

Es el proceso de menor prioridad del firmware. En él, se configura un puerto UART a 115200 baudios con una codificación de 8 bits y 1 bit de

paro. El sistema de comunicación no tiene periodo de ejecución, sino que se ejecuta bajo demanda y únicamente cuando el sistema tiene tiempo libre.

A través del driver del periférico interno UART, implementado por el fabricante, se construye un hilo para recibir datos, y un hilo para transmitirlos. El hilo receptor, verifica que los bytes recibidos tienen el formato adecuado, en cuyo caso almacena una copia de estos datos y hace un llamado al despachador.

El despachador es una función que identifica el endpoint<sup>3</sup> donde va dirigido el paquete, dentro de una tabla de funciones. Luego hace el llamado al procesador del endpoint y espera su respuesta.

El procesador del endpoint es una función que se almacena en una lista denominada, lista de endpoints o lista de puertos. Recibe como argumento el puntero a un buffer de datos y la longitud de dichos datos. El desarrollador puede definir el formato de los datos que puede recibir (dentro de las capacidades de la especificación del sistema de comunicación) y las respuestas que se envían a través de una trama.

Por su parte, el hilo emisor se activa cuando se ha recibido una trama para enviar y sólo si el CPU tiene tiempo muerto. Este hilo empaqueta la trama en un buffer dedicado y envía el puntero al driver UART del RTOS para su transmisión. El driver UART del RTOS implementa transacciones DMA (del inglés Direct Memory Access) las cuales agilizan las transacciones entre la RAM y un periférico interno.

### A. Requerimientos

El sistema de comunicación debe programarse con un protocolo de transmisión que provoque una baja carga de procesamiento por parte del RTOS. También se debe transmitir y recibir bytes sin formato definido, con especial atención a variables de tipo flotante de doble precisión. Al mismo tiempo, el sistema de comunicación debe ser capaz de transmitir muestras entre el RAP y la

<sup>1</sup> Una función reentrante, es una función que puede ser interrumpida y vuelta a ejecutar sin repercusiones en el resultado.

<sup>2</sup> Una condición de carrera sucede cuando dos procesos cuyo orden de ejecución es desconocido comparten un mismo recurso que debe consumirse en un orden determinado.

<sup>3</sup> En software, un endpoint es una abstracción para identificar canales de comunicación entre procesos, los canales pueden ser buffers, puertos, interfaces, etc.

computadora mientras recibe comandos de acción.

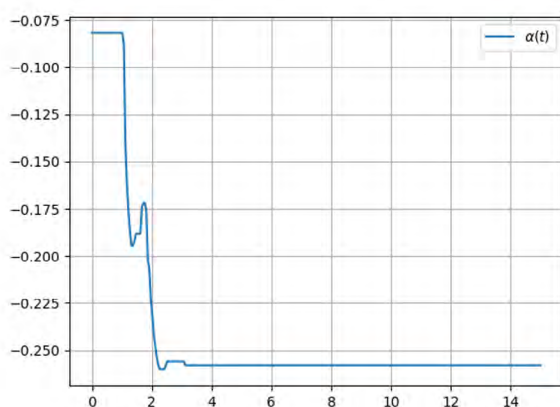
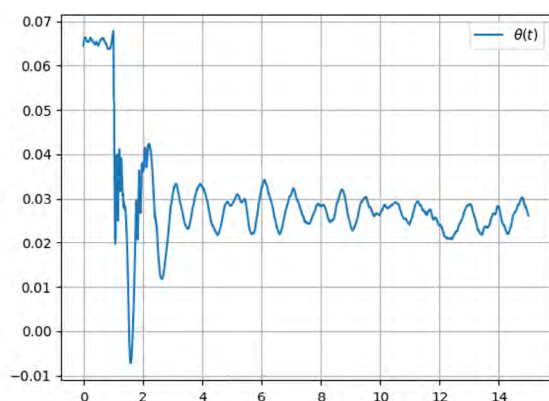
## B. Tramas

Para que lo anterior sea posible, el sistema de comunicación se basa en la transmisión de paquetes de datos denominados tramas.

Una trama se compone por 2 bytes dedicados a la cabecera, de 1 a 255 bytes destinados a los datos crudos y de 1 a 5 bytes opcionales, para la suma de verificación.

La cabecera de la trama dedica 1 byte al símbolo de inicio de trama, y 1 byte para indicar la longitud de la carga de datos (payload en inglés, o datos crudos).

La suma de verificación, por su parte, dedica 5 bytes opcionales para una suma de verificación tipo CRC (no implementada por el momento) y 1 byte no opcional, para el símbolo de final de trama.



**Fig. 5.** Coordenadas generalizadas muestreadas en tiempo real desde el sistema de comunicación del RAP (discretización bilineal). En la figura de arriba se encuentra la inclinación del robot en radianes. Abajo, la posición angular de la rueda virtual, también medida en radianes

## C. Canales

El sistema de comunicación no admite tramas cuyo *payload* tenga longitud 0, debido a que el primer byte del *payload* es destinado a identificar el canal o *endpoint* de comunicación. Es decir, el primer byte del *payload* indica la función que decodificará los datos de la trama. El desarrollador puede implementar sus propias funciones y definir los argumentos codificados en la trama.

El firmware del *RAP* tiene instalados 6 *endpoints*, desarrollados especialmente para el proyecto.

- **Ping:** con identificador 0 y sin recibir datos. Este *endpoint* transmite una trama de respuesta con su propio identificador.
- **startSampler:** con identificador 1, recibe 4 bytes en formato entero sin signo para solicitar una cantidad fija de muestras, o una cantidad indefinida de muestras si la cantidad es 0; las muestras son enviadas en forma de tramas con identificador 1. La codificación consiste en 5 valores en formato flotante de doble precisión cuyo orden es [variable  $\alpha$ , variable  $\theta$ , variable  $\dot{\alpha}$ , variable  $\dot{\theta}$ , variable  $u$ ], la variable  $u$ , es el valor de torque aplicado a la rueda virtual del robot  $\equiv [\vec{x}, u]$ .
- **stopSampler:** con identificador 2 y sin recibir datos. Este *endpoint* detiene el proceso de muestreo independientemente de la cantidad de tramas restante.
- **startControl:** con identificador 3 y sin recibir datos. Activa la ejecución del *controlLoop*.
- **stopControl:** con identificador 4 y sin recibir datos. Detiene la ejecución del *controlLoop*.
- **setK:** con identificador 5, recibe 4 valores en formato flotante de doble precisión cuyo orden es:  $[K_{\alpha}, K_{\theta}, K_{\dot{\alpha}}, K_{\dot{\theta}}]$ . Estos valores se cargan como la matriz de ganancias del *controlLoop*.



## » VII. Resultados

El modelo dinámico del péndulo invertido sobre ruedas diferenciales, reducido a un modelo discretizado, holonómico y lineal e invariante en el tiempo, resultó en una excelente aproximación a la dinámica del RAP, y permitió la implementación de un controlador LQR de tiempo discreto que resultó satisfactorio, pues es capaz de estabilizar las variables críticas del RAP, las cuales son: el ángulo de las ruedas  $\alpha$  y su velocidad  $\dot{\alpha}$ , relacionadas con la posición del RAP, así como el ángulo  $\theta$  de inclinación del péndulo. Para comprobar la efectividad del estabilizador se perturba al robot RAP con ligeras desviaciones del ángulo de inclinación de su cuerpo y, empujando a éste para obligarlo a moverse de lugar. Por un lado, las gráficas resultantes de esta actividad deben mostrar que las variables de estado convergen, por otro, el robot RAP debe presentar resistencia perceptible al movimiento y evitar desplomarse, con lo cual, se puede afirmar que el sistema físico es estable en lazo cerrado y el controlador funciona. Se probaron tres controladores distintos: Controlador LQR de tiempo continuo, Controlador LQR del sistema discretizado por el método ZOH y Controlador LQR del sistema discretizado por el método bilineal.

Todos los controladores fueron instalados a través del sistema de comunicación utilizando un script de python versión 3, solicitando al muestreador 1500 capturas, equivalente a 15 segundos de operación. El entorno de ejecución es una máquina corriendo el sistema operativo Arch Linux y el puerto serial identificado como `/dev/rfcomm0`. El script de muestreo se corre con una redirección del `stdout` hacia un archivo de texto con formato CSV. El comando correspondiente es: `python3 muestreador.py $>$ samples.csv`. En esta sección solamente se muestran las gráficas del controlador que se obtuvo por el método ZOH, para el cual se aplicó una ganancia de acondicionamiento de 0.0245 unidades y la matriz instalada fue:

$$K = [-0.022051 \ -1.904296 \ -0.034785 \ -0.353189]$$

En la fig. 5 se observa que el ángulo de inclinación del RAP es muy cercano a cero y que el ángulo de las ruedas se estabiliza, lo que físicamente se

traduce en un robot RAP erguido sobre sus ruedas y estacionado en una zona cercana al origen.

Aunque cada método de discretización produjo una matriz de ganancias diferente, la implementación de ésta, para todos los casos, requirió de una ganancia de acondicionamiento para evitar saturar la señal de los actuadores. Al final, este factor, diferente para cada controlador, conllevó a una matriz de ganancias  $K$  muy similar. Las ganancias del controlador discretizado por el método de bilineal requirieron el factor de reducción menos significativo. Es por esta razón que se prefirió este resultado.

Es importante mencionar que la planta física presenta un efecto de giro sobre el acimut cuando el RAP retorna al origen. Esto se debe al controlador de lazo abierto implementado en cada motor, pues a pesar de que comparten especificaciones, estos presentan un desgaste y una respuesta diferente a la señal de control. Por lo que usando este método no es posible garantizar que las ruedas giren exactamente igual. Además, el control de posición también es impreciso, pues el RAP no retorna al origen real, sino a una región próxima al origen.

## » VIII. Conclusión

En este trabajo se presentó un método para implementar un controlador en el caso de un modelo dinámico para un robot péndulo invertido con dos variables de configuración, la posición unidimensional del robot y el ángulo de inclinación del cuerpo. Para linealizar el modelo dinámico del RAP se utilizó el punto de equilibrio inestable del sistema y la suposición de que el ángulo del cuerpo del robot tendrá variaciones despreciables.

Posteriormente, se reorganizaron las ecuaciones en forma de *representación de estado* para analizar la controlabilidad y la estabilidad del sistema, lo que permitió confirmar que el sistema dinámico linealizado del RAP es controlable.

Se utilizó el software *Octave* y los paquetes *control* y *signal* para comprobar que es posible controlar el RAP con un controlador LQR calculando una matriz de control  $K$  de tiempo continuo y

validando su estabilidad de forma numérica.

En la implementación, se propuso una frecuencia de control de 100 Hertz, y, usando el software *Octave*, se discretizó el sistema dinámico del *RAP* con el método bilineal. Posteriormente, se calculó una nueva matriz de control  $K$  pero esta vez de tiempo discreto, misma que pudo validarse de forma numérica al producir un comportamiento estable. La ejecución del controlador  $LQR$  consiste en multiplicar la matriz de control  $K$  por el vector de estado  $\bar{x}$  en cada periodo de control.

Por otro lado, discretizar el sistema permitió observar el comportamiento del controlador para una frecuencia conveniente, elegida por el programador, asegurando la estabilidad de la planta en frecuencias bajas de control. Además, usar frecuencias bajas permitió implementar el controlador en un microcontrolador y dividir el tiempo de ejecución en múltiples tareas, sin sobrecargar el sistema embebido. Como resultado, fue más sencillo garantizar que el proceso de control se ejecute en su respectivo periodo con el menor desfase posible, añadiendo mayor funcionalidad al firmware del *RAP*.

Con el fin de probar la estabilidad del *RAP*, se programó el firmware en la tarjeta de control del *RAP* usando el puerto de programación y se instaló la matriz de control  $K$  utilizando el sistema de comunicación. Este prototipo de robot se construyó desde cero y se documentó el proceso del controlador para que sirva como una plataforma de experimentación de control automático en un hardware real. El software de control es abierto lo que permite implementar estrategias distintas o modificar parámetros para comparar resultados.

#### A. Problemas abiertos

Una importante actualización al proyecto debe ser añadir el control de velocidad y considerar el ángulo de giro (el acimut del *RAP*), para lo cual, el modelo propuesto no es suficiente, porque el giro del robot y el efecto de las llantas individuales introducen nuevos fenómenos sobre la dinámica del sistema inercial, además de convertirse en un sistema, conocido en robótica como no

holonómico.

Con un control de velocidad y giro, será posible construir un sistema externo de navegación a través del sistema de comunicación, añadiendo al prototipo *RAP* el módulo de inteligencia básica o *MIB*. El *MIB* se programará sobre el estándar *Robot Operating System* (ROS2) con lo cual será posible construir rutinas de evasión de obstáculos y rutinas de servicios y aplicarse como plataforma de experimentación en robótica móvil.

## » VII. Referencias

- [1] P. Frankovský, L. Dominik, A. Gmitterko, and I. Virgala, “Modelling of two-wheeled self-balancing robot driven by gearmotors” *International Journal of Applied Mechanics and Engineering*, vol.22, 2017.
- [2] E. Olguín-Díaz, “3D Motion of Rigid Bodies: A Foundation for Robot Dynamics Analysis”, ser. *Studies in Systems, Decision and Control*. Springer International Publishing, 2018. [Online]. Available: <https://books.google.com.mx/books?id=9vt9DwAAQBAJ>.
- [3] G. Rigatos, K. Busawon, J. Pomares, and M. Abbaszadeh, “Nonlinear optimal control for the wheeled inverted pendulum system” *Robotica*, vol. 38, no. 1, pp. 29–47, 2020.
- [4] S. Kim and S. Kwon, “Dynamic modeling of a two-wheeled inverted pendulum balancing mobile robot” *International Journal of Control, Automation and Systems*, vol. 13, no. 4, pp. 926–933, 2015.
- [5] C.-F. Hsu and W.-F. Kao, “Double-loop fuzzy motion control with cog supervisor for two-wheeled self-balancing assistant robots” *International Journal of Dynamics and Control*, pp. 1–16, 2020.
- [6] H. Ahmadi Jeyed and A. Ghaffari, “A nonlinear optimal control based on the sdre technique for the two-wheeled self-balancing robot” *Australian Journal of Mechanical Engineering*, pp.1–9, 2020.
- [7] R. D. Huerta Gil, “Control de un péndulo invertido sobre dos ruedas de tres grados de libertad” Thesis, Universidad Nacional

- Autónoma de México, 2014.
- [8] M. Fauziyah, Z. Amalia, I. Siradjuddin, D. Dewatama, R. P. Wicaksono, and E. Yudaningtyas, "Linear quadratic regulator and pole placement for stabilizing a cart inverted pendulum system" *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 3, pp. 914–923, 2020.
- [9] M. R. Bageant, "Balancing a two-wheeled segway robot" Thesis, Massachusetts Institute of Technology, 2011.
- [10] F. Grasser, A. D'arrigo, S. Colombi, and A. C. Rufer, "Joe: a mobile, inverted pendulum" *IEEE Transactions on industrial electronics*, vol. 49, no. 1, pp. 107–114, 2002.
- [11] V. Mudeng, B. Hassanah, Y. T. K. Priyanto, and O. Saputra, "Design and simulation of two-wheeled balancing mobile ro-bot with pid controller" *International Journal of Sustainable Transportation*, vol. 3, no. 1, pp. 12–19, 2020.
- [12] H. Nabil, "Supervised neural network control of real-time two wheel inverted pendulum" *Journal of Advanced Engineering Trends*, vol. 38, no. 2, pp. 131–146, 2020.
- [13] O. E. Aburas and A. H. Ahmed, "A guide to implement a two wheeled robot using pole-placement on arduino" *The International Journal of Engineering and Information Technology (IJEIT)*, vol. 6, 2020.
- [14] M. Velazquez, D. Cruz, S. Garcia, and M. Bandala, "Velocity and motion control of a self-balancing vehicle based on a cascade control strategy" *International Journal of Advanced Robotic Systems*, vol. 13, Apr. 2016.
- [15] H. Hellman and H. Sunnerman, "Two-wheeled self-balancing robot," Thesis, KTH Royal Institute of Technology, 2015.
- [16] C. Dengler and B. Lohmann, "Adjustable and adaptive control for an unstable mobile robot using imitation learning with trajectory optimization" *Robotics*, vol. 9, no. 2, p. 29, 2020.
- [17] R. P. M. Chan, K. A. Stol, and C. R. alkyard, "Review of modelling and control of two-wheeled robots" *Annual Reviews in Control*, vol. 37, pp. 89–103, Apr. 2013.
- [18] J. Morantes, D. Espitia, O. Morales, R. Jiménez, and O. Avi-les, "Control system for a segway" *International Journal of Applied Engineering Research (IJAER)*, vol. 13, 2018.
- [19] D. M Abd-elaziz, "Conventional fuzzy logic controller for ba-lancing two-wheel inverted pendulum" *Journal of Advanced Engineering Trends*, vol. 38, no. 2, pp. 107–119, 2020.
- [20] M. Moness, D. Mahmoud, and A. Hussein, "Real-time mamdani-like fuzzy and fusion-based fuzzy controllers for balancing two-wheeled inverted pendulum" *Journal of Am-bient Intelligence and Humanized Computing*, pp. 1–17, 2020.
- [21] T. Johnson, S. Zhou, W. Cheah, W. Mansell, R. Young, and S. Watson, "Implementation of a perceptual controller for an inverted pendulum robot," *Journal of Intelligent & Robotic Systems*, pp. 1–10, 2020.
- [22] Pololu, "Gearmotor 4752" Accedido en 01-10-2019 a <https://www.pololu.com/product/4752/specs>, 2019.



# ANÁLISIS DE DATOS PARA LA OPTIMIZACIÓN EFICIENTE DE HORARIOS Y APRENDIZAJE AUTOMÁTICO

## Data analysis for efficient schedule optimization and Machine Learning

Rogelio Escobedo Mitre <sup>1</sup>	isc.rogelio.em@tectijuana.edu.mx
Ángeles Quezada Cisnero <sup>2</sup>	angeles.quezada@tectijuana.edu.mx
Bogart Yair Marquez Lobato <sup>3</sup>	bogart@tectijuana.edu.mx
Arnulfo Alanis Garza <sup>4</sup>	alanis@tectijuana.edu.mx

<sup>1,2,3,4</sup>Tecnológico Nacional de México (Instituto Tecnológico de Tijuana)

### RESUMEN

En los últimos años, la integración de técnicas de aprendizaje automático (AP), en los sistemas de gestión escolar ofrece varias oportunidades para mejorar la eficiencia y la toma de decisiones en el ámbito educativo. Al aplicar el AP en la educación, se pueden obtener beneficios significativos. Sin embargo, es importante tener en cuenta que la integración exitosa de técnicas de AP en los sistemas de gestión escolar requiere una infraestructura de datos sólida, la recopilación adecuada de datos y la consideración de cuestiones éticas y de privacidad. Además, el AP no debe reemplazar la interacción humana en la educación, sino complementarla y mejorarla, brindando a los educadores y estudiantes herramientas adicionales para el éxito educativo. Debido al incremento en la retícula de la carrera de Ingeniería en Sistemas Computacionales es necesario llevar a cabo una predicción más confiable y de manera automática de los horarios por semestre. Para abordar el problema de generación de horarios de manera manual, se llevó a cabo un análisis exhaustivo del proceso existente. Esto implica recopilar información relevante sobre cómo se realiza actualmente la generación de horarios por semestre en la institución educativa. Se estudio el enfoque actual utilizado para generar los horarios, además analizar los problemas y las limitaciones asociadas al proceso manual. Se investigarán diferentes técnicas de Machine Learning que podrían aplicarse al problema de generación de horarios. Esto podría incluir algoritmos de

optimización, algoritmos de agrupamiento o clasificación, algoritmos genéticos u otros enfoques de aprendizaje automático que puedan adaptarse al problema específico.

**Palabras Clave:** aprendizaje automático, automatización, horarios.

### ABSTRACT

In recent years, the integration of machine learning (ML) techniques into school management systems offers several opportunities to enhance efficiency and decision-making in the educational field. Applying ML in education can yield significant benefits. However, it is important to note that the successful integration of ML techniques into school management systems requires a robust data infrastructure, proper data collection, and consideration of ethical and privacy issues. Furthermore, ML should not replace human interaction in education but rather complement and improve it by providing educators and students with additional tools for educational success. Due to the increased complexity in the curriculum of the Computer Systems Engineering program, it is necessary to carry out a more reliable and automated prediction of semester schedules. To address the manual scheduling generation problem, a comprehensive analysis of the existing process was conducted. This

involved gathering relevant information on how semester schedules are currently generated in the educational institution. The current approach used for scheduling was studied, along with an analysis of the problems and limitations associated with the manual process. Various ML techniques that could be applied to the scheduling generation problem were investigated. This could include optimization algorithms, clustering or classification algorithms, genetic algorithms, or other machine learning approaches that can be adapted to the specific problem.

**Keywords:** Machine learning, automation, schedules.

## ► I. Introducción

En los últimos años, el avance de la tecnología ha transformado muchos aspectos de nuestra sociedad, y el ámbito educativo no ha sido una excepción. La integración de técnicas de Aprendizaje Automático (Machine Learning) en los sistemas de gestión escolar ha surgido como una solución prometedora para mejorar la eficiencia y la toma de decisiones en las instituciones educativas [1] [2].

En la actualidad, las instituciones educativas se enfrentan a desafíos constantes en la gestión y organización de sus actividades académicas. El seguimiento y análisis manual de datos estudiantiles, la planificación de horarios y la adaptación de los programas educativos a las necesidades individuales de los estudiantes son tareas que demandan un tiempo considerable y pueden resultar propensas a errores.

A su vez la falta de atención en la programación de horarios en instituciones de educación básica y media, resalta la importancia de considerar cuándo es más propicio el aprendizaje, a diferencia de la programación de horarios universitarios, la programación de horarios escolares se centra en las clases y no en los estudiantes [17].

En la literatura científica y en la industria, se han realizado avances significativos en el desarrollo de sistemas de gestión escolar basados en Aprendizaje Automático. Algunos estudios se han centrado en la predicción del rendimiento académico de los

estudiantes, utilizando algoritmos de regresión y clasificación para identificar factores que influyen en el éxito o fracaso escolar. Otros trabajos han abordado la optimización de la planificación de horarios, aplicando algoritmos de agrupamiento y programación lineal para asignar eficientemente recursos y minimizar conflictos [3] [4].

Además, se han desarrollado sistemas de recomendación de asignaturas y cursos, utilizando técnicas de filtrado colaborativo y análisis de contenido para sugerir opciones académicas adaptadas a las preferencias y habilidades de cada estudiante [5] [6].

En universidades como Waterloo han desarrollado un sistema que se basa en una filosofía "dirigida por la demanda", donde los estudiantes eligen primero sus cursos y el sistema intenta encontrar el mejor horario para maximizar el número de solicitudes satisfechas donde el problema se descompone en subproblemas manejables que se resuelven secuencialmente utilizando una heurística ávida para asignar horarios a secciones y un algoritmo de relajación lagrangiana para asignar aulas, sin embargo la asignación de profesores no se hace automática a las secciones de los cursos; ese proceso generalmente se realiza manualmente con antelación [18].

Se sabe, con antelación, que la situación de horarios automáticos no es una situación nueva, y tampoco algo esporádico, ya que universidades como Don Bosco ya que es sabido que el proceso de elaboración de horarios académicos sigue siendo realizado de manera manual, ya que personal docente y administrativo se reúne previo al inicio del período a iniciar para analizar en detalle las estadísticas relacionadas con la población estudiantil, el cuerpo docente y las instalaciones disponibles. Este análisis tiene como objetivo principal la creación de horarios que permitan una gestión eficiente de los recursos institucionales [19].

La programación de horarios académicos plantea un desafío particular que reside en la problemática más amplia de la asignación de recursos. En la comunidad científica, se conoce este desafío

como el "Problema de Programación de Horarios Universitarios". Este tipo de problemas implica la generación de horarios para tareas específicas, con el objetivo de cumplir de manera óptima con condiciones y requisitos particulares. A lo largo del tiempo, se han abordado estos problemas de programación de horarios utilizando diversos enfoques, tales como algoritmos basados en la Colonia de Hormigas, Búsqueda Tabú, Coloreo de Grafos y Algoritmos Genéticos, entre otros [20].

En este contexto, el presente proyecto de análisis de datos para la optimización eficiente de horarios y aprendizaje automático tiene como objetivo abordar estas problemáticas para ofrecer una solución integral y adaptada a la carrera de Ingeniería en Sistemas Computacionales, que permita optimizar la gestión académica con respecto a la cantidad de materias abiertas por semestre para mejorar la oferta académica para estudiantes de tal forma que se brinde un mejor rendimiento en el avance reticular. A través de un enfoque innovador y personalizado, se busca establecer una base sólida para la transformación de las prácticas educativas y sentar las bases para futuros avances en la gestión escolar impulsada por la inteligencia artificial

## ► II. Conceptos relacionados

### II. I. Aprendizaje Automático

El aprendizaje automático, conocido como Machine Learning en inglés (ML), es una disciplina dentro del campo de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos capaces de aprender y mejorar automáticamente a través del análisis de datos. Estos algoritmos tienen la capacidad de identificar patrones complejos y realizar predicciones o tomar decisiones basadas en ellos, sin requerir una programación específica para cada tarea [7]. El aprendizaje automático se basa en principios estadísticos y computacionales, utilizando técnicas que permiten a las máquinas procesar y analizar grandes volúmenes de datos con el fin de extraer información relevante. Al utilizar estos datos, los algoritmos pueden descubrir relaciones y regularidades ocultas, lo que les permite aprender y adaptarse a medida que

se les suministra más información.

Ciclo de vida del ML. Existe un ciclo básico para poder realizar proyectos de machine learning, como se muestra en la Figura 1.

Figura 1  
Componentes de ML



**Fuente:** Elaboración propia (imagen de referencia tomada de <https://keepcoding.io/blog/ciclo-de-vida-de-un-proyecto-en-machine-learning/>)

### II. II. Principales algoritmos de ML

#### Algoritmo de aprendizaje supervisado

El algoritmo de aprendizaje supervisado es una técnica de ML en la que se utilizan conjuntos de datos etiquetados para entrenar un modelo y realizar predicciones sobre nuevos datos no etiquetados. Este tipo de algoritmo se basa en la idea de que existe una relación entre las características o variables de entrada y una variable de salida deseada, y el objetivo es aprender esta relación a partir de los datos de entrenamiento [9].

Los algoritmos más utilizados en el aprendizaje supervisado incluyen árboles de decisión, clasificación de Naive Bayes Ingenuo, regresión por mínimos cuadrados, regresión logística, métodos de ensamble y Máquinas de Soporte Vectorial por sus siglas en inglés SVM (support vector machines). Cada algoritmo tiene sus propias ventajas y desventajas, y su elección dependerá del problema específico y los datos disponibles.

#### Árboles de decisión:

Los árboles de decisión son ampliamente utilizados en el análisis de datos y la toma de decisiones automatizada, ya que son fáciles de interpretar y pueden manejar conjuntos de datos grandes y complejos. Además, son una base importante para algoritmos más avanzados, como Random Forests y Gradient Boosting, que combinan múltiples árboles de decisión para mejorar su rendimiento

predictivo. [11]

Clasificador Bayesiano Ingenuo (Naive Bayes):

Es un algoritmo de clasificación en el campo del aprendizaje automático y la minería de datos. Este clasificador se basa en el teorema de Bayes y asume que las características (o variables) que se utilizan para la clasificación son independientes entre sí, lo que se conoce como una suposición "ingenua (naive)" [12].

Regresión por mínimos cuadrados:

Es un método estadístico utilizado para modelar la relación entre una variable dependiente (o respuesta) y una o más variables independientes (o predictores) en un conjunto de datos. El objetivo principal de la regresión por mínimos cuadrados es encontrar una ecuación o modelo que minimice la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo. En otras palabras, se busca ajustar una línea o una superficie de manera que se minimice la suma de los errores cuadrados. [13]

Regresión logística:

Es un método estadístico y un modelo de regresión utilizado para analizar la relación entre una variable binaria dependiente (es decir, una variable que toma dos valores, típicamente 0 y 1) y una o más variables independientes. A diferencia de la regresión lineal, que se utiliza para predecir valores continuos, la regresión logística se emplea para problemas de clasificación, como la predicción de la probabilidad de pertenecer a una de las dos categorías [14].

### ► III. Propuesta

Debido al avance de la tecnología el desarrollo de sistemas de gestión escolar ha sido una respuesta a la necesidad de optimizar la administración y el manejo de la información en las instituciones educativas. Tradicionalmente, estas tareas se realizaban de manera manual y consumían una gran cantidad de tiempo y recursos. Sin embargo, con el avance de la tecnología, se han desarrollado

soluciones informáticas que permiten automatizar y simplificar estos procesos.

En cuanto al desarrollo del contexto de intervención, es importante mencionar que existen diferentes sistemas de gestión escolar disponibles en el mercado. Sin embargo, la implementación de técnicas de aprendizaje automático (machine learning) en este tipo de sistemas es una tendencia emergente que ofrece ventajas adicionales.

El Análisis de datos propuesto para la optimización eficiente de horarios utilizará técnicas de aprendizaje automático para analizar y procesar los datos relacionados con la gestión académica y administrativa de una institución educativa.

El proyecto de Análisis de datos para la optimización eficiente de horarios y Aprendizaje Automático busca abordar esta problemática, ofreciendo una solución innovadora y eficiente para mejorar la gestión y el rendimiento de las instituciones educativas. Al aprovechar las capacidades del machine learning, se espera que el sistema pueda proporcionar análisis predictivos y recomendaciones basadas en datos, permitiendo a las instituciones educativas tomar decisiones informadas y mejorar su eficacia.

En resumen, el contexto de intervención del proyecto se enmarca en la necesidad de modernizar y optimizar la gestión escolar a través del uso de técnicas de machine learning. Con el objetivo de superar los desafíos y limitaciones de los sistemas tradicionales, el proyecto busca ofrecer una solución innovadora que mejore la eficiencia, la toma de decisiones y el rendimiento general de las instituciones educativas.

Se llevarán a cabo diversas actividades durante el desarrollo del proyecto. Estas actividades incluirán la recopilación y análisis de datos relevantes, la identificación de los requerimientos específicos de nuestra institución educativa, específicamente en el departamento de Sistemas y Computación y de la carrera de Ingeniería en sistemas computacionales, la creación de algoritmos de machine learning adaptados a las necesidades del sistema de gestión escolar, y la implementación



de un entorno de prueba para validar y ajustar la solución propuesta.

## » IV. Metodología

En la Figura 2 se describe el proceso metodológico, llevado a cabo en esta investigación, el cual consiste en los siguientes pasos: definición de la muestra, obtención de datos, análisis de datos y presentación de los datos.

Figura 2  
Proceso metodológico de la investigación



Fuente: *Elaboración propia.*

### IV.I. Definición de la muestra

En este proyecto, se utilizó un muestreo estratificado para seleccionar los periodos que formaron parte del estudio diagnóstico. La muestra fue seleccionada considerando la

información con la que se cuenta con respecto a las bases de datos almacenadas, considerando también diferentes estratos relevantes para el estudio, como, por ejemplo, semestre cursado del alumnado, periodo de ingreso del mismo (enero-junio o agosto-diciembre), rendimiento académico, entre otros.

El proceso de selección de la muestra fue ejecutado de manera cuidadosa. En particular, se puso un énfasis especial en la información contenida en las bases de datos institucionales, contenidas en el sistema escolar desde 11 años atrás y las bases de datos almacenadas y procesadas en archivos de Excel, las cuales se convirtieron en una fuente invaluable de datos para llevar a cabo este estudio.

Uno de los principales criterios utilizados para la estratificación de la muestra fueron los semestres cursados por el alumnado, tomando como base las cantidades de materias ofertadas por semestre y la cantidad de alumnos que tomaron dichas materias. De esta manera, se aseguró que cada estrato de semestre estuviera representado en la muestra en proporción a su importancia relativa en la población

estudiantil.

Además, se consideró el periodo de ingreso de los estudiantes, dividiéndolo en dos categorías: aquellos que ingresaron en el primer semestre del año (enero-junio) y aquellos que lo hicieron en el segundo semestre (agosto-diciembre). Esta división se basó en la suposición de que el contexto académico podría variar según el momento en que los estudiantes comenzaron su formación, lo que podría influir en su rendimiento y en sus experiencias educativas.

El rendimiento académico de los estudiantes también se convirtió en un criterio de estratificación crucial. Se categorizaron los estudiantes en función de su desempeño académico, considerando aspectos como el promedio de calificaciones, la tasa de aprobación y otras métricas relacionadas. Esta estratificación permitió analizar si existían diferencias significativas en el rendimiento de los estudiantes en función de sus logros previos.

### IV.II. Obtención de datos

El análisis de datos es una parte crucial en la toma de decisiones informadas, y la calidad de la información utilizada es esencial para obtener resultados precisos. En este contexto, la información utilizada se obtuvo de múltiples fuentes, principalmente de fuentes primarias. La fuente principal de datos en este caso fue el Sistema de Información Escolar (SIE). Este sistema, que actúa como un repositorio central de información relacionada con cuestiones académicas y de planificación educativa, proporcionó una base sólida para el análisis.

Un aspecto importante a destacar es el acceso directo a los horarios almacenados en el sistema SIE. Esto permitió una recopilación de datos eficiente y precisa, ya que los horarios son componentes esenciales para el análisis, especialmente en un entorno educativo. El acceso a estos datos en tiempo real y de manera directa garantizó la integridad y actualidad de la información utilizada.

Además, para contar con un panorama más completo y para rastrear tendencias y patrones a lo largo del

tiempo, se aprovecharon los archivos de Excel que contenían el historial de horarios recopilados y actualizados semestre tras semestre. Estos archivos históricos proporcionaron una perspectiva a largo plazo y permitieron evaluar la evolución de los horarios y su impacto en la planificación educativa.

En resumen, el análisis de datos se basó en una sólida combinación de fuentes primarias, incluyendo el sistema SIE, la colaboración del departamento de servicios escolares, el acceso a los horarios en tiempo real y el uso de archivos históricos en formato Excel.

A continuación, se presenta una representación de los horarios correspondientes al periodo académico AGO-DIC 2023, junto con la estructura utilizada por el sistema para generar y formatear los informes que sirven como fundamento para su posterior archivo en formato Excel, como se muestra en la figura 2.

Figura 3  
Concentrado de materias, indicando el catedrático asignado, así como grupo, horario y aula

REAL	GRUPO	MATERIA	CR	PLAN	PAG.	CATEDRÁTICO	NECES	OCUPAC	REP	LUNES	HORA/AULA
	AED-1285SC1A	FUND DE PROGRAMACION	5.00	161	01A			0000000	0	110011009111	
	ACF-0901	CALC. DIFERENCIAL	5.00	161	01A			0000000	0	0800110009111	
	ACC-0906	FUNDAM D INVESTIG	4.00	161	01A			0000000	0	1200110009111	
	ACA-0907	TALL DE ETICA	4.00	161	01A			0000000	0	080009009111	
	AEP-10418C1A	MATEM DISCRETAS	5.00	161	01A			0000000	0	110012009111	

Fuente: *Elaboración propia.*

#### IV.III. Obtención de datos

La información se ha sometido a un proceso de organización, con el objetivo de crear una estructura coherente que facilite el análisis. Esta organización se ha llevado a cabo utilizando diversas herramientas y técnicas, como tablas, hojas de cálculo y otros formatos apropiados para asegurar que los datos sean de fácil acceso y comprensión.

Para profundizar en el análisis, se realizó un estudio de correlación entre las variables presentes en el conjunto de datos. Esta fase fue esencial para identificar las relaciones y dependencias entre las variables. Para lograrlo, se emplearon técnicas avanzadas, como la matriz de correlación y gráficos de dispersión, que proporcionaron una visión más clara de cómo las variables se

relacionan entre sí.

Por último, se llevaron a cabo comparaciones minuciosas entre grupos o categorías de datos para identificar diferencias significativas. Estas comparaciones se llevaron a cabo siguiendo un enfoque estadístico sólido, lo que permitió obtener conclusiones fundamentadas y relevantes.

En la figura 2 se muestra la información de grupos por semestre y periodo, en la que se pueden observar las cantidades de grupos creadas a partir de la información recopilada:

Figura 4  
Tabla comparativa del historial de materias abiertas por semestre a lo largo de un periodo 7 años.

semestre	2016	2017	2018	2019	2020	2021	2022
1	3+r	4+r	4+r	3+2r	3+2r	4+3r	5
2	4+r	4+r	4+r	3+2r	3+2r	4+2r	4+r
3	2+r	2+r	3+r	3	3	3+1r	4
4	4	2	3	3	3	3+2r	4
5	3	2	2	2	2	3	3
6	3	2	3	2	2	3	3
7	2	2	2	2	2	2	3
8	2	2	2	2	2	3	2+r
9	1	1	1	2	2	2	2

#### V. Presentación de los datos

Para poder visualizar mejor la información y poder hacer un análisis estadístico se generaron tablas en Excel, así como gráficas para poder visualizar de forma clara y precisa dicha información generada durante el análisis de datos.

Durante la fase de presentación de resultados, se ha puesto un fuerte énfasis en la creación de informes que destaquen de manera clara y concisa los hallazgos esenciales obtenidos a través del análisis. Estos informes se han estructurado de manera que sean fácilmente accesibles y comprensibles para los interesados y partes involucradas en el proyecto.

En particular, para mejorar la visualización de los datos y permitir un análisis estadístico más efectivo, se ha llevado a cabo la creación de tablas en formato Excel. Estas tablas han sido diseñadas con precisión y han sido organizadas de manera lógica, lo que facilita la revisión de los datos y la realización de cálculos estadísticos cuando es necesario.

Además, se han creado gráficos y visualizaciones adecuadas para representar visualmente la información generada durante el análisis de datos. Estas representaciones visuales permiten una comprensión más rápida e intuitiva de los patrones, tendencias y relaciones presentes en los datos. Se ha prestado especial atención a la claridad y la precisión en la presentación de estos gráficos, asegurando que sean herramientas efectivas para comunicar los resultados a las partes interesadas.

En resumen, se ha realizado un esfuerzo significativo en la generación de informes que simplifican la comunicación de los hallazgos clave, aprovechando gráficos, tablas y visualizaciones apropiadas. El objetivo final es proporcionar una plataforma sólida para la toma de decisiones basadas en datos y garantizar que los resultados del análisis sean fácilmente accesibles y comprensibles para todos los involucrados.

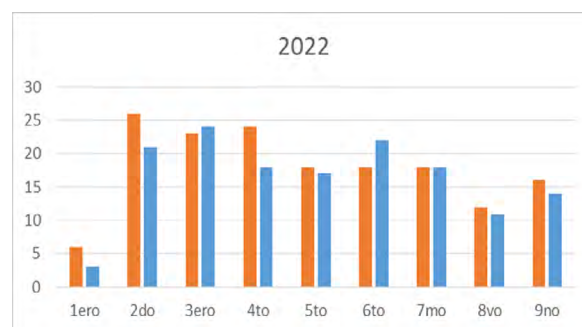
La información con la que se trabajó para poder realizar el proceso de análisis y evaluación de los periodos desde 2012 se obtuvo del programa SIE y se trabajó directamente en una hoja de cálculo donde se ordenó la información de tal forma que se pueda visualizar y trabajar para próximos semestres.

Primero, se debe observar cuántos grupos en total del primer semestre del periodo inmediato anterior. Enseguida se evalúa el segundo semestre del mismo periodo para analizar cuántos posibles repetidores serán candidatos a tomar las mismas materias del semestre en el que se requiere trabajar. Posteriormente se evalúa el segundo semestre del año inmediato anterior de tal forma que se pueda ver el avance y las diferencias/similitudes que se pueden presentar, tomando en cuenta el periodo de ingreso de los jóvenes (ingreso directo en agosto, o ingreso posterior al semestre propedéutico).

Después de analizar los periodos de cada semestre se muestra la cantidad de grupos, por semestre, de 3 periodos consecutivos, que son los que se requiere analizar para poder tomar las decisiones pertinentes para el siguiente semestre, como se muestra en la Figura 4.

La gráfica nos muestra la cantidad de grupos por semestre que se proyectaron en el periodo enero-junio y agosto-diciembre 2022, con esta información se genera la proyección de grupos para el semestre enero-junio 2023, iniciando el primer semestre hasta el noveno. Con esta información se puede proyectar la cantidad de grupos que deberán ofertarse tomando en consideración los grupos de alumnos que estén en situación normal en cuanto a su carga académica, así como alumnos que tengan alguna materia reprobada.

Figura 5  
Comparativo de cantidad de materias abiertas por semestre en el año 2022



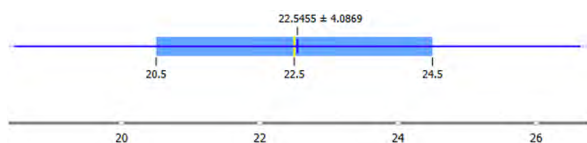
## ► V. Resultados

Posterior al análisis y procesamiento de la información los datos obtenidos de la sección anterior se procedieron a utilizar herramientas como Box Plot para generar diagramas de caja y bigotes, ya que con esta herramienta gráfica nos permite visualizar la distribución y las estadísticas resumidas de un conjunto de datos. Por lo tanto, es muy útil para identificar la simetría, la dispersión y la presencia de valores atípicos en un conjunto de datos de una manera visual y puede proporcionar una comprensión rápida de la distribución de los datos sin necesidad de una representación detallada de cada punto de datos.

Para lograr una proyección más precisa de la cantidad de materias a abrir por semestre, se utilizó la herramienta de Box Plot, donde se analizó la trayectoria desde agosto de 2012 hasta enero de 2023, abarcando 22 semestres consecutivos, es decir, un historial de 11 años. Esto nos permite obtener un promedio de la cantidad de materias a abrir por semestre.

Con los datos recopilados para el segundo semestre, se obtuvo que existe un rango mínimo de 14 grupos y un máximo de 32 grupos. La herramienta utilizada arroja un promedio de 22.5 grupos, con una variación de más o menos 4 grupos por período, como se muestra en la figura 4.

Figura 6  
Diagrama caja y bigotes



Posteriormente se utilizó la herramienta Python para el desarrollo de un módulo que muestre en pantalla el resultado del análisis de la información proporcionada. El despliegue de la plataforma se inició con la incorporación de las librerías esenciales. La incorporación de Pandas, Random Forest y Numpy, conocidas por su profunda capacidad y funcionalidad, sentaron las bases sólidas sobre las cuales se construyó el marco de trabajo. Estas bibliotecas desempeñan un rol de importancia crítica en las fases de procesamiento, análisis y modelado de los datos, asegurando la robustez y efectividad de todo el sistema.

La subsiguiente etapa implicó la transferencia de datos al sistema. Se procedió a la importación de conjuntos de datos preexistentes que albergaban información pertinente a cursos, horarios de clases, capacidad de aulas y la demanda estudiantil.

Figura 7  
Cargado de datos.

```
# Carga de datos
df = pd.read_csv('../data/period_sem_table.csv')
dfBeta = pd.read_csv('../data/period_sem_table_modified.csv')
df = dfBeta.round()
```

Con el fin de obtener un conocimiento más profundo de los datos, se llevó a cabo un análisis exploratorio exhaustivo. Durante este proceso, se identificaron patrones y tendencias significativas que contribuyeron a una comprensión más clara de las necesidades y preferencias de la institución con respecto a la distribución de grupos y horarios.

Figura 8  
Exploración de datos

	período 1	período 2	período 3	período 4	período 5	período 6	período 7	período 8	período 9	período 10
0	3.0	2.0	3.0	1.0	4.0	2.0	2.0	1.0	2.0	2.0
1	15.0	24.0	19.0	19.0	19.0	32.0	14.0	26.0	22.0	25.0
2	18.0	19.0	22.0	18.0	15.0	18.0	24.0	14.0	13.0	15.0
3	12.0	13.0	17.0	18.0	18.0	18.0	17.0	23.0	22.0	13.0
4	18.0	14.0	12.0	12.0	12.0	17.0	18.0	16.0	18.0	12.0
5	14.0	12.0	12.0	12.0	12.0	12.0	14.0	18.0	14.0	14.0
6	14.0	14.0	12.0	13.0	14.0	12.0	12.0	12.0	19.0	12.0
7	13.0	13.0	10.0	15.0	15.0	12.0	11.0	12.0	13.0	12.0
8	9.0	9.0	9.0	2.0	2.0	7.0	6.0	5.0	4.0	6.0

El modelo de aprendizaje automático fue entrenado utilizando la técnica de Random Forest, la cual ha demostrado ser altamente eficaz en la tarea de predecir y recomendar configuraciones de grupos y horarios.

El modelo generó proyecciones basadas en la información de entrada y los resultados se extrajeron en un formato CSV. Estas proyecciones ofrecieron recomendaciones de alto valor para la planificación de grupos estudiantiles.

Figura 9  
Predicciones.

```
# Redondea los valores del array al entero más cercano
rounded_arr = np.around(predictions)

# Imprime el array redondeado
#print(rounded_arr)

for i in range(9):
    print(f"Semestre {i+1}: {rounded_arr[i]}")
```

```
[9]
...
Semestre 1: 6.0
Semestre 2: 22.0
Semestre 3: 19.0
Semestre 4: 18.0
Semestre 5: 15.0
Semestre 6: 17.0
Semestre 7: 16.0
Semestre 8: 15.0
Semestre 9: 11.0
```

## ▶ VI. Conclusiones y trabajo futuro

En esta investigación, se aborda el proceso de generación de horarios por semestre para la carrera de Ingeniería en Sistemas Computacionales. Actualmente, este proceso se realiza de manera manual al final de cada semestre con el objetivo de proyectar la cantidad de grupos que se deben habilitar para el siguiente periodo académico.

Aunque hasta ahora el proceso manual ha brindado un pronóstico confiable, requiere una cantidad significativa de horas hombre para llevarlo a cabo.

Para resolver este desafío y mejorar la eficiencia del proceso de generación de horarios, se puede considerar la aplicación de técnicas de aprendizaje automático.

El desarrollo de sistemas de software para resolver este problema siempre es de alto valor. Una herramienta propuesta en esta investigación es generar los horarios de manera automática usando aprendizaje automático, como lo es un algoritmo de aprendizaje supervisado. Como sabemos la matrícula de la carrera de Sistemas Computacionales ha estado en constante crecimiento por tal motivo es necesario que se automatice este proceso.

Como trabajo futuro se llevará a cabo el uso de técnicas de aprendizaje supervisado como lo es el uso de algoritmos de clasificación de bayes ingenuo (Naïve Bayes), para que con esas técnicas se lleve a cabo el aprendizaje y se haga la predicción de manera automática con la información que actualmente nos proporciona el sistema SIE.

En resumen, el análisis del proceso manual existente y la presentación de información relevante permitirían identificar los desafíos y las limitaciones en la generación de horarios. Basándose en esto, se explorarían soluciones de aprendizaje automático y se desarrollaría un sistema automatizado para mejorar la eficiencia y la precisión en la generación de horarios en la institución educativa.

## » VII. Referencias

- [1] J. O. Yunga Pedraza, «Estudio del estado del arte sobre la predicción de deserción universitaria usando machine learning,» de Universidad Salesiana, Ecuador, 2023.
- [2] A. U. Castaneda, «Un viaje hacia la inteligencia artificial en la educación,» Realidad y Reflexion, pp. 121-136, 2022.
- [3] C. Russo, «Tratamiento Masivo de Datos Utilizando Técnicas de Machine Learning,» REDI, pp. 131-134, 2016.
- [4] A. D. Luca, «Uso de la Técnica de Transfer Learning en Machine Learning para la Clasificación de Productos en el Banco Alimentario de La Plata,» SEDICI, pp. 1-16, 2021.
- [5] B. A. A. BENITEZ, GENERACION DE HORARIOS MEDIANTE SISTEMAS, México, D.F.: Instituto Politécnico Nacional. Centro de Investigación en Computación, 2007.
- [6] E. B. Cañón, «Modelo predictivo del progreso en el aprendizaje de los estudiantes de uniminuto aplicando técnicas de machine learning,» de Scielo, Mexico, 2021.
- [7] D. Hinestroza Ramírez, «El Machine Learning a través de los tiempos, y los aportes a la humanidad,» Universidad Libre, pp. 1-17, 2019.
- [8] keepcoding, «keepcoding.io,» 2 diciembre 2022. [En línea]. Available: <https://keepcoding.io/blog/ciclo-de-vida-de-un-proyecto-en-machine-learning/>.
- [9] O. Simeone, «A Very Brief Introduction to Machine Learning With Applications to Communication Systems,» de IEEE, 2018.
- [10] B. Zarco García, «Algoritmos de clasificación supervisados y semi-supervisados: análisis y comparativa,» UPM, pp. 1-12, 2020.
- [11] "Introduction to Machine Learning with Python" de Andreas C. Müller y Sarah Guido. Libro "Pattern Recognition and Machine Learning" de Christopher M. Bishop.
- [12] "Introduction to Information Retrieval" de Christopher D. Manning, Prabhakar Raghavan, y Hinrich Schütze.
- [13] "An Introduction to the Analysis of Variance" de Ronald A. Fisher.
- [14] "Logistic Regression: A Self-Learning Text" de David G. Kleinbaum y Mitchel Klein. Libro "Introduction to the Practice of Statistics" de David S. Moore, George P. McCabe y Bruce A. Craig.
- [15] "The Elements of Statistical Learning" de Trevor Hastie, Robert Tibshirani y Jerome Friedman.
- [16] "Support Vector Machines" de Nello

- Cristianini y John Shawe-Taylor.
- [17] Víctor Fabio Suarez, Omar Danilo Castrillón, «Diseño de una metodología basada en técnicas inteligentes para la distribución de procesos Académicos en ambientes de trabajo job shop.
  - [18] Michael W. Carter, “A Comprehensive Course Timetabling and Student Scheduling System at the University of Waterloo” 2001
  - [19] Elías Ventura, Eva Marcela, Mendoza Pacas, Carlos Rafael, “Análisis y diseño de un planificador automatizado de horarios universitarios”, 2002
  - [20] Mireya Flores Pichardo, “Revisión de Algoritmos Genéticos Aplicados al Problema de la Programación de Cursos Universitarios” 2011.

# ANÁLISIS DEL COMPORTAMIENTO DE DATOS MÉDICOS DE COVID -19 MEDIANTE INTELIGENCIA ARTIFICIAL

## Behavioral Analysis of COVID-19 Medical Data using Artificial Intelligence

Roberto Carlos Hernández Aparicio <sup>1</sup>	roberto.hernandezap@alumno.buap.mx
María del Carmen Santiago Díaz <sup>2</sup>	marycarmen.santiago@correo.buap.mx
Ana Claudia Zenteno Vázquez <sup>3</sup>	ana.zenteno@correo.buap.mx
Judith Pérez Marcial <sup>4</sup>	judith.perez@correo.buap.mx
Gustavo Trinidad Rubín Linares <sup>5</sup>	gustavo.rubin@correo.buap.mx

<sup>1, 2, 3, 5</sup> Benemérita Universidad Autónoma de Puebla (Facultad de Ciencias de la Computación)

### RESUMEN

La pandemia del COVID-19 ha generado una crisis sanitaria a nivel mundial, y contar con herramientas que permitan gestionar eficientemente la enfermedad es de vital importancia. En este contexto, la inteligencia artificial ofrece una oportunidad única para analizar grandes volúmenes de datos médicos y obtener información valiosa que pueda contribuir a la toma de decisiones médicas y mejorar la gestión de la pandemia. En este trabajo se aplican técnicas de inteligencia artificial para modelar los resultados obtenidos a partir de bases de datos en México de COVID- 19.

**Palabras Clave:** COVID-19, Inteligencia Artificial, Selección Múltiple, Árbol de Decisión.

databases in Mexico.

**Keywords:** COVID-19, Artificial Intelligence, Multiple Choice, Decision Tree

### ► I. Introducción

La pandemia del COVID-19 ha generado una crisis sin precedentes en la salud mundial, afectando a millones de personas y desafiando a los sistemas de salud en todo el mundo [1]. Los datos médicos relacionados con esta enfermedad son inmensos y complejos, lo que dificulta su análisis y comprensión por medios convencionales [2]. Aquí es donde la inteligencia artificial muestra su potencial, ya que tiene la capacidad de procesar grandes volúmenes de datos y encontrar patrones ocultos que pueden ser cruciales para comprender la propagación del virus, la efectividad de las medidas de contención y la evolución de la enfermedad en diferentes poblaciones. Por ejemplo, que ofrece la Inteligencia Artificial a los gobiernos es que pueden predecir con precisión los recursos necesarios en ubicaciones específicas, identificar cómo se puede mejorar el tratamiento en los puntos críticos y, en última instancia, detener la propagación del virus [3].

### ABSTRACT

The COVID-19 pandemic has generated a global health crisis, and having tools that allow the disease to be efficiently managed is of vital importance. In this context, artificial intelligence offers a unique opportunity to analyze large volumes of medical data and obtain valuable information that can contribute to medical decision-making and improve management of the pandemic. In this work, artificial intelligence techniques are applied to model the results obtained from COVID-19

La investigación en el campo de la inteligencia artificial aplicada a la salud ha experimentado un crecimiento significativo en los últimos años. Los modelos de aprendizaje automático y las técnicas de minería de datos se han utilizado exitosamente para abordar problemas de diagnóstico, predicción y optimización de tratamientos. En el contexto específico del COVID-19, estas técnicas pueden ser aplicadas para identificar patrones de comportamiento en los datos, como la tasa de propagación del virus, las características clínicas de los pacientes más afectados, las comorbilidades asociadas y la eficacia de diferentes estrategias de intervención [4].

También hay estudios sobre la frecuencia de características clínicas dentro de México en el estado de Puebla lo que nos indica que fue realizado dependiendo su sexo y edad en general para observar que síntomas están asociados con la mortalidad en el COVID-19 [5]. Además de encontrar otro estudio referente a los adolescentes y su experiencia durante el confinamiento de la pandemia el cual solo mide aquellos de entre la edad de 14 a 25 años [6]. Mencionaremos que se observó la historia de la evolución de la pandemia para ponernos en un contexto general del mundo y de nuestra región de América Latina la cual cuales fueron las características de las personas con las cuales la mortalidad se elevaba [7]. Por último, el gobierno de México tiene una página web en la que nos muestra estadísticas del COVID-19 la cual se pueden aplicar filtros muy generales de cada estado, pero no algo en específico [8].

Uno de los desafíos fundamentales que se enfrentará en este proyecto de pasantía investigativa es la calidad y disponibilidad de los datos. La recolección y curación de datos precisos y actualizados de fuentes confiables serán aspectos críticos para asegurar la validez y confiabilidad de los resultados. Una vez superados los desafíos iniciales, la etapa de diseño y exploración de modelos teóricos será crucial para el éxito del proyecto. La elección de los algoritmos de inteligencia artificial adecuados para el análisis de datos de COVID-19 será fundamental para obtener resultados precisos y útiles. Además, es esencial considerar las limitaciones y especificidades de los

datos médicos, ya que esto influirá en la selección del modelo más adecuado para cada caso.

El análisis y evaluación teórica del modelo será otra fase crucial del proyecto. Además de medir la calidad de las predicciones, también será importante evaluar la robustez de los resultados frente a diferentes escenarios y datos de entrada. Esto permitirá determinar la confiabilidad de los modelos y la posibilidad de aplicarlos en situaciones reales de manera efectiva.

Se exploraron los datos médicos de COVID-19 en estudiantes de edades comprendidas entre los 18 y 28 años, con el propósito de reducir los porcentajes de error en el análisis. Se llevó a cabo una investigación exhaustiva para identificar patrones y tendencias significativas en la población estudiantil joven afectada por la enfermedad.

Es importante destacar que, si bien la inteligencia artificial puede proporcionar herramientas poderosas para el análisis y predicción de datos médicos de COVID-19, su implementación debe ser cuidadosa y ética. La privacidad y la protección de datos deben ser consideraciones primordiales, y se deben seguir todas las regulaciones y normativas pertinentes para garantizar un enfoque responsable y seguro en el uso de la inteligencia artificial en el campo de la salud.

La inteligencia artificial (IA) ha desempeñado un papel fundamental en el análisis del comportamiento de la COVID-19 a través de datos médicos. A través de una variedad de aplicaciones, la IA ha contribuido significativamente en la predicción de la propagación del virus, el diagnóstico temprano de casos, la evaluación del riesgo de gravedad en pacientes, la optimización de tratamientos personalizados, el análisis de imágenes médicas, la secuenciación genómica del virus, la simulación de políticas de control y la predicción de la supervivencia a largo plazo. Estos trabajos han permitido una comprensión más profunda de la pandemia, facilitando la toma de decisiones más informadas por parte de profesionales de la salud y formuladores de políticas, y han contribuido a la gestión y mitigación efectiva de la propagación de la COVID-19 [9, 10].



El análisis de la pandemia de COVID-19 a través de modelos estadísticos como la Regresión Logística (Selección Múltiple) y Árboles de Decisión ha sido objeto de estudio en diversos artículos científicos. Estos modelos se han aplicado para abordar una amplia gama de cuestiones, desde la predicción de la propagación del virus y la identificación de factores de riesgo hasta la toma de decisiones clínicas y la evaluación de políticas de control. La Regresión Logística ha sido utilizada para analizar datos demográficos y médicos a fin de entender la dinámica de la infección y sus consecuencias en diferentes poblaciones. Los Árboles de Decisión, por su parte, han ayudado a los profesionales de la salud a diagnosticar la COVID-19 y a establecer protocolos de tratamiento, así como a evaluar las estrategias de mitigación. Estos enfoques se han convertido en herramientas cruciales en la lucha contra la pandemia, proporcionando información valiosa y permitiendo una toma de decisiones basada en datos más precisa [10, 11].

En conclusión, este proyecto de pasantía investigativa representa una oportunidad emocionante para explorar el potencial de la inteligencia artificial en el análisis y predicción de datos de COVID-19. Los resultados obtenidos pueden tener un impacto significativo en la toma de decisiones médicas y en la gestión de la pandemia a nivel nacional e internacional. Al utilizar técnicas avanzadas de inteligencia artificial, se espera obtener información valiosa que contribuya a la comprensión y contención de la enfermedad, permitiendo una respuesta más efectiva y rápida ante futuras emergencias sanitarias. Además, este proyecto puede sentar las bases para futuras investigaciones en el campo de la salud y la inteligencia artificial, avanzando hacia un enfoque más sólido y holístico para enfrentar los desafíos de la salud pública en el siglo XXI

## » II. Metodología

La metodología consistió en una investigación preliminar y recopilación exhaustiva de datos sobre COVID-19 en estudiantes jóvenes, seguida de la preparación y limpieza rigurosa de los datos. Luego, se diseñaron modelos de inteligencia artificial y se evaluó su viabilidad para predecir

patrones de comportamiento relacionados con la enfermedad, aportando información relevante para la toma de decisiones con inteligencia artificial. A continuación, se muestran los pasos a seguir en la metodología utilizada obsérvese en la Figura 1:

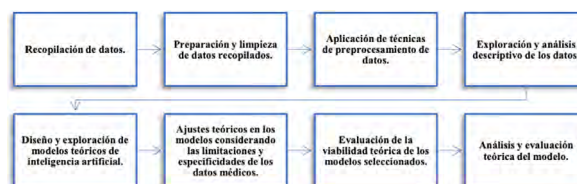


Fig. 1. Metodología propuesta.

### A. Recopilación de datos

La recopilación de datos involucró una investigación preliminar exhaustiva, que abarcó la revisión de diversas fuentes bibliográficas y bases de datos tanto nacionales como internacionales. En este proceso, se determinó la base de datos del gobierno mexicano sobre COVID-19, definiendo su implementación. El propósito central fue delimitar el alcance del proyecto y seleccionar las fuentes de datos pertinentes para abordar la información esencial sobre la propagación de COVID-19 entre estudiantes jóvenes. Adicionalmente, se estableció una comunicación efectiva con las entidades encargadas de la recopilación de datos, garantizando la obtención de los permisos y autorizaciones necesarios para asegurar el acceso a la información de manera legal y ética.

### B. Preparación y limpieza de datos recopilados

Una vez obtenidos los datos, se llevó a cabo un proceso minucioso para eliminar el ruido y los valores atípicos que pudieran introducir sesgos o distorsiones en el análisis posterior. Cada registro fue exhaustivamente revisado para asegurar su integridad y coherencia con el contexto del estudio. Además, se aplicaron técnicas de normalización y estandarización a los datos con el fin de homogeneizarlos, es decir, ajustarlos a una misma escala o rango de valores [12]. Esto permitió realizar una comparación efectiva entre las distintas variables, garantizando que todas contribuyen equitativamente al análisis sin ser afectadas por magnitudes o unidades dispares.

Al normalizar y estandarizar los datos, se facilitó el proceso de modelado y análisis descriptivo, aumentando la precisión y comprensión de los resultados obtenidos.

#### *C. Aplicación de técnicas de preprocesamiento de datos*

En esta etapa, se utilizaron técnicas avanzadas de preprocesamiento de datos para los desafíos específicos presentes en los datos médicos de COVID-19: valores faltantes y variables categóricas. La imputación de valores faltantes consiste en estimar y completar los datos que faltan en el conjunto de información, de manera que no haya vacíos en los registros, lo que es crucial para mantener la integridad de los datos y evitar sesgos en los análisis posteriores.

Por otro lado, la codificación de variables categóricas implica convertir las características cualitativas en una forma numérica para que los modelos de inteligencia artificial pueda procesarlas adecuadamente. Al hacerlo, se establece una relación lógica entre las diferentes categorías, lo que facilita el análisis y la comparación de las variables en el contexto del estudio.

En conjunto, estas técnicas de preprocesamiento aseguran que los modelos de inteligencia artificial reciban datos coherentes y completos, lo que mejora la calidad de los resultados y la precisión de las predicciones. Al abordar de manera efectiva los desafíos inherentes a los datos médicos de COVID-19, se obtiene una base sólida para el análisis y la toma de decisiones informadas en la investigación

#### *D. Exploración y análisis descriptivo de los datos*

Mediante la utilización de técnicas de visualización, como gráficos y diagramas, se exploraron los datos médicos de COVID-19 en estudiantes de edades entre 18 y 28 años. Se identificaron tendencias, distribuciones y correlaciones entre variables clave, proporcionando una comprensión profunda del comportamiento de los datos. Esta etapa permitió descubrir patrones iniciales y plantear hipótesis que serían fundamentales para guiar el

proceso de modelado.

#### *E. Diseño y exploración de modelos teóricos de inteligencia artificial*

En esta fase, se diseñaron los modelos de inteligencia artificial, centrándose en los algoritmos de Selección Múltiple y Árbol de Decisión; debido a su capacidad demostrada para manejar datos complejos y realizar predicciones precisas en casos médicos [14]. Estos modelos han sido ampliamente utilizados en investigaciones relacionadas con el COVID-19 y han demostrado un alto rendimiento en la clasificación y análisis de datos epidemiológicos. Además, la selección de hiperparámetros y la validación cruzada permitieron ajustar los modelos de manera óptima para abordar los desafíos específicos presentes en los datos médicos de la población de estudiantes jóvenes. La consideración de la interpretabilidad del Árbol de Decisión fue un factor crucial, ya que proporciona información clara sobre el proceso de toma de decisiones y las razones detrás de las predicciones, lo que resulta fundamental para una toma de decisiones informada en el contexto de la salud pública y la gestión de la pandemia.

#### *F. Ajustes teóricos en los modelos considerando las limitaciones y especificidades de los datos médicos*

Esta etapa se enfocó en realizar ajustes teóricos con respecto a secciones del proceso programado en los modelos para abordar las características específicas y limitaciones de los datos médicos de COVID-19. Se exploraron técnicas de selección de características y reducción de dimensionalidad para mejorar la capacidad predictiva del modelo, sin comprometer su capacidad de generalización. Se priorizó la precisión y el desempeño de los modelos, sin perder de vista la relevancia y el significado clínico de las predicciones.

#### *G. Evaluación de la viabilidad teórica de los modelos seleccionados*

La viabilidad de los modelos de inteligencia artificial se evaluó considerando su capacidad para capturar patrones de comportamiento

relevantes en los datos médicos de COVID-19 en estudiantes de edades entre 18 y 28 años. Se compararon los resultados con investigaciones previas y se utilizaron métricas de evaluación adicionales, como la curva ROC que representa la formación suficiente para escoger el mejor valor umbral. Además, nos aporta herramientas para, no solo minimizar los falsos positivos y los falsos negativos, sino para escoger de qué preferimos tener menos: falsas alarmas de misiles o, por el contrario, el número de ellos que no son detectados [13]; y el área bajo la curva (AUC), para medir el rendimiento y la eficacia de los modelos. Esta etapa permitió validar la efectividad de los modelos en la predicción de patrones de comportamiento relacionados con COVID-19.

#### H. Análisis y evaluación teórica del modelo

Se realizó un análisis detallado del desempeño de los modelos de Selección Múltiple y Árbol de Decisión, se dividió el conjunto de datos en conjuntos de entrenamiento y prueba utilizando técnicas de validación cruzada. Luego, se aplicaron ambos modelos a los conjuntos de prueba y se evaluaron sus predicciones utilizando diferentes métricas, poniendo énfasis en los resultados obtenidos y su interpretación. Se evaluaron métricas clave, como sensibilidad, especificidad y precisión, para determinar la capacidad de los modelos para predecir patrones de comportamiento en los datos médicos de COVID-19 en estudiantes jóvenes. Los resultados se contextualizan y se discutieron sus implicaciones prácticas para la toma de decisiones en salud pública y la planificación de intervenciones.

El análisis detallado del desempeño de ambos modelos permitió identificar sus fortalezas y debilidades. El modelo de Selección Múltiple destacó por su alta sensibilidad, lo que lo convierte en una herramienta útil para detectar casos positivos de COVID-19 en estudiantes jóvenes. Por otro lado, el modelo de Árbol de Decisión mostró una alta especificidad y mayor interpretabilidad, lo que lo convierte en una opción valiosa para clasificar correctamente a los

casos negativos y comprender el razonamiento detrás de las predicciones. La combinación de ambos modelos ofrece una visión más completa y precisa del comportamiento de COVID-19 en esta población, lo que puede ser de gran utilidad para la toma de decisiones en salud pública y la planificación de intervenciones.

En resumen, esta metodología, centrada en el uso de los modelos predictivos de Selección Múltiple y Árbol de Decisión, permitió realizar una investigación detallada en la Predicción de Patrones de Comportamiento en Datos Médicos de COVID-19 mediante Inteligencia Artificial en estudiantes de edades entre 18 y 28 años. Los resultados obtenidos a partir de esta metodología pueden tener un impacto significativo en la comprensión de la propagación del virus y la efectividad de las intervenciones en esta población, contribuyendo así a una mejor gestión de la pandemia y al diseño de estrategias de salud pública más efectivas.

### ► III. Resultados

En cuanto al proceso de la prueba, inicialmente descargamos la base de datos y aplicamos un filtro para eliminar las columnas innecesarias. Luego, realizamos un filtrado específico para seleccionar únicamente a pacientes con edades entre 18 y 28 años. Posteriormente, obtuvimos un conjunto de datos listos para analizar cómo puede observarse en la Tabla 1.

Tabla I  
Muestra de aplicación de valores y eliminación de valores

Tipo Paciente	Edad	Sexo	Diabetes	EPOC	ASMA
2	18	2	2	2	2
2	24	2	2	2	2
2	28	2	2	2	2

A continuación, aplicamos otro filtro para generar la Figura 2, enfocándonos en aquellos individuos que fallecieron y dieron positivo en la prueba de COVID-19. Esta gráfica nos proporcionó información importante para nuestro estudio.

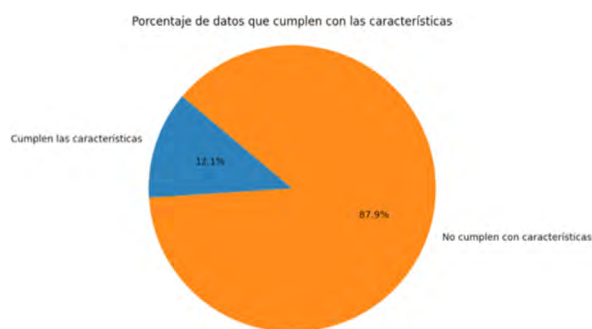


Fig. 2. Porcentaje obtenido a partir de las características.

Al observar esto antes de aplicar algún modelo aplicaremos a los datos un submuestreo aleatorio el cual implica el objetivo de balancear el conjunto de datos, se opta por eliminar instancias de la clase mayoritaria [14] para equilibrar las clases puede haber un desequilibrio en la cantidad de pacientes fallecidos y no fallecidos, se aplica para equilibrar las clases y obtener una cantidad similar de muestras para ambas clases.

Ahora se aplica ya el modelo Árbol de decisión el cual nos muestra los datos observe la Tabla 2 los cuales se explicarán a continuación:

Resultados para la clase "1" (fallecido):

- Precisión: 0.58 (aproximadamente 58% de las predicciones positivas para "1" fueron correctas)-
- Recall: 0.61 (aproximadamente 61% de las muestras "1" fueron correctamente identificadas).
- F1-score: 0.59 (el F1-score es una media ponderada entre precisión y recall, en este caso, aproximadamente 59%).
- Support: 18 (existen 18 muestras de la clase "1" en el conjunto de prueba).

Resultados para la clase "2" (no fallecido):

- Precisión: 0.63 (aproximadamente 63% de las predicciones positivas para "2" fueron correctas).
- Recall: 0.60 (aproximadamente 60% de las muestras "2" fueron correctamente identificadas).
- F1-score: 0.62 (el F1-score es una media ponderada entre precisión y recall, en este caso, aproximadamente 62%).

- Support: 20 (existen 20 muestras de la clase "2" en el conjunto de prueba).

Tabla 2  
Resultados de aplicar Árbol de decisión.

Precisión en el conjunto de la prueba		60%		
	Precisión	Recall	F1-Score	Support
1	0.58	0.61	0.59	18
2	0.68	0.6	0.62	20
Exactitud Global			0.61	38
Promedio Ponderado		0.61	0.61	38

Al obtener estos resultados exitosos se aplicó a un caso de estudio un solo individuo el cual es un masculino de 28 años que fue intubado además de presentar obesidad, diabetes, cardiovascular y UCI el cual dio POSITIVO a COVID-19 su probabilidad de fallecer es del 100% las características el modelo nos predice esto.

Optamos por no aplicar el modelo de Selección Múltiple debido a que el árbol de decisión nos proporcionó resultados satisfactorios para cumplir con nuestros objetivos. Decidimos enfocarnos en profundizar en su interpretación y justificar su uso para nuestro estudio, considerando limitaciones de tiempo y recursos. La complejidad del modelo de Selección Múltiple y la efectividad obtenida con el árbol de decisión fueron factores clave para tomar esta decisión

## » IV. Conclusiones

En conclusión, este proyecto de pasantía investigativa ha sido una invaluable contribución al campo de la salud y la inteligencia artificial, brindando una visión detallada y exhaustiva del proceso de aplicación de técnicas avanzadas para la predicción de patrones de comportamiento en datos médicos relacionados con COVID-19 en estudiantes de edades comprendidas entre los 18 y 28 años. La combinación de una sólida investigación preliminar, rigurosa preparación y limpieza de datos, exploración analítica y diseño de modelos teóricos ha permitido obtener resultados prometedores y relevantes en la lucha contra la pandemia. Los resultados obtenidos han

sido comparados con otros estudios y modelos, y con contundencia se ha demostrado que los modelos de Selección Múltiple y Árbol de Decisión utilizados en este proyecto superan en precisión y efectividad a otros enfoques previamente aplicados en investigaciones similares. Estas evidencias refuerzan la confianza en la aplicabilidad y potencialidad de estos modelos para contribuir significativamente en la comprensión y manejo de la propagación del COVID-19 en la población estudiantil.

Los modelos de Selección Múltiple y Árbol de Decisión han demostrado ser efectivos en la predicción de comportamientos asociados con la pandemia, este estudio puede sentar las bases para futuras investigaciones en el campo de la salud y la inteligencia artificial, avanzando hacia un enfoque más sólido y holístico para enfrentar los desafíos de la salud pública en el siglo XXI. Proporcionando una herramienta valiosa para anticipar tendencias y mejorar la toma de decisiones médicas y la gestión de la crisis sanitaria.

Además, se ha abierto una oportunidad para seguir refinando y mejorando los modelos predictivos, abordando desafíos específicos asociados con los datos médicos de COVID-19 y ampliando el alcance de la investigación a otras poblaciones y regiones. La continua evolución de estos modelos permitirá un análisis más detallado y una toma de decisiones más informada en futuras emergencias sanitarias.

El uso de datos gubernamentales de COVID-19 de México ha sido fundamental para el éxito del proyecto. La colaboración y disponibilidad de estas fuentes confiables de información han permitido un análisis preciso y robusto, brindando una base sólida para el desarrollo de modelos de inteligencia artificial confiables y aplicables. El conocimiento adquirido puede ser esencial para el diseño de políticas públicas y estrategias de salud que prevengan y controlen la pandemia. La capacidad de anticipar patrones de comportamiento en la población estudiantil de 18 a 28 años permitirá tomar decisiones proactivas y diseñar intervenciones más efectivas para proteger la salud y el bienestar de la población.

## » V. Agradecimientos

Un reconocimiento a la Vicerreoría de Investigación y Estudios de Posgrado por la beca otorgada mediante el programa jóvenes investigadores a Roberto Carlos Hernández Aparicio, estudiante de la carrera de Ingeniería en Ciencias de la Computación.

## » VI. Referencias

- [1] Información básica sobre la COVID-19. (n.d.). <https://www.who.int/es/news-room/questions-and-answers/item/coronavirus-disease-covid-19>.
- [2] Enfermedad por el Coronavirus (COVID-19) | OPS/OMS | Organización Panamericana de la Salud. (n.d.). <https://www.paho.org/es/enfermedad-por-coronavirus-covid-19>.
- [3] Jurídico, F. (2021, 5 junio). La inteligencia artificial podría ser la solución contra el coronavirus Foro jurídico. Foro Jurídico. <https://forojuridico.mx/la-inteligencia-artificial-podria-ser-la-solucion-contr-el-coronavirus/>.
- [4] Asale, R.-. (n.d.). epidemia | Diccionario de la lengua española. «Diccionario De La Lengua Española» - Edición Del Tricentenario. <https://dle.rae.es/epidemia>.
- [5] Hernández-Morales, M. D. R., Maldonado-Castañeda, S., Mancilla-Hernández, E., Amaro-Zarate, I., Aguirre-Barbosa, M., & Nazarala-Sanchez, S. (2023). [Frequency of clinical characteristics and factors associated with mortality in patients hospitalized for COVID-19 in Puebla, Mexico]. *Revista Alergia Mexico (Tecamachalco, Puebla, Mexico : 1993)*, 69(2), 67–71. <https://doi.bibliotecabuap.elogim.com/10.29262/ram.v69i2.1146>.
- [6] Sánchez-Xicotencatl, C. O., Campillo-Labrandero, M., Esparza-Meza, E. M., Stincer-Gómez, D., Rojo-Solís, A. L. T., & Aveleyra-Ojeda, E. (2022). Experiencias de los adolescentes frente al confinamiento y la pandemia de la COVID-19. *Revista de Psicopatología y Psicología Clínica*, 27(3), 169–178. <https://doi.org/10.5944/rppc.30938>.
- [7] Jacalyn Duffin. (2022). COVID-19

- : A History. McGill-Queen's University Press. <https://ebSCO.bibliotecabuap.elogim.com/login.aspx?direct=true&db=e000xww&AN=3561910&lang=es&site=eds-live>.
- [8] COVID-19 Tablero México. (n.d.). COVID - 19 Tablero México. <https://datos.covid-19.conacyt.mx/>.
- [9] Elsheikh, A. H., Saba, A. I., Panchal, H., Shanmugan, S., Alsaleh, N. A., & Ahmadein, M. (2021). Artificial intelligence for Forecasting the Prevalence of COVID-19 Pandemic: An Overview. *Healthcare*, 9(12), 1614. <https://doi.org/10.3390/healthcare9121614>.
- [10] Santos, A. J. D., Almeida, D., De Jesus, E. M., Santos, J. B. D., Da Silva, M. M., & Barreto, M. (2020). Inteligência artificial e COVID-19. En EDUFBA eBooks. <https://doi.org/10.9771/9786556300757.009>.
- [11] Vaishya R, Javaid M, Khan IH, et al. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab. Syndr. Clin. Res. Rev.* 2020 [acceso em 2021 ago 23]; 14(4):337-339. Disponible en <https://doi.org/10.1016/j.dsx.2020.04.012>.
- [12] Lalmuanawma, S., Hussain, J., & Chhakchhuak, L. (2020). Applications of Machine learning and Artificial intelligence for COVID-19 (SARS-COV-2) pandemic: a review. *Chaos, Solitons & Fractals*, 139, 110059. <https://doi.org/10.1016/j.chaos.2020.110059>.
- [13] ¿Qué es la minería de datos? La minería de datos, explicada - AWS. (n.d.). Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/data-mining/>.
- [14] Barrientos Martínez, R. E., Cruz Ramírez, N., Acosta Mesa, H. G., Rabatte Suárez, I., Gogeoascoechea Trejo, M. del C., Pavón León, P., & Blázquez Morales, S. L. (2009). Árboles de decisión como herramienta en el diagnóstico médico. *Revista Médica de la Universidad Veracruzana*, 9(2). [https://www.uv.mx/rm/num\\_anteriores/revmedica\\_vol9\\_num2/contenido/index.htm](https://www.uv.mx/rm/num_anteriores/revmedica_vol9_num2/contenido/index.htm).
- [15] J. Curvas ROC (Receiver-Operating-Characteristic) y sus aplicaciones. Ana Rocío del Valle Benavides, Universidad de Sevilla, Departamento de Estadística e Investigación Operativa.
- [16] Batista, G.E.A.P.A, Prati, R.C, Monard, M.C. A study of the behavior of several methods for balancing machine learning training data, *SIGKDD Explor Newslett*, 6(1),20-29, (2004).
- [17] Información referente a casos COVID-19 en México - datos.gob.mx/busca. (n.d.). <https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico>.

# MEJORA DEL APRENDIZAJE BASADO EN LA LÓGICA: APLICACIÓN DE GUESS-18 PARA LA EVALUACIÓN DE LA EXPERIENCIA DEL USUARIO

## Enhancing Logic-Driven Learning: Applying GUESS-18 for User Experience Evaluation

David Bonilla Carranza <sup>1</sup>	jose.bcarranza@academicos.udg.mx
Adriana Peña Pérez Negrón <sup>2</sup>	adriana.ppnegron@academicos.udg.mx
Mirna Muñoz Mata <sup>2</sup>	mirna.munoz@cimat.mx

<sup>1</sup> Universidad de Guadalajara (Centro Universitario de Ciencias Exactas e Ingenierías)

<sup>2</sup> Centro de Investigación en Matemáticas (Unidad Zacatecas)

### RESUMEN

Este artículo detalla un estudio que examina la experiencia del usuario en un juego educativo de ingeniería de software a través de la Escala de satisfacción de la experiencia del usuario del juego (GUESS-18). En él participaron 111 estudiantes de ingeniería informática que jugaron un juego diseñado para mejorar sus habilidades de software a través de tareas interactivas. El GUESS-18 ayudó a medir el impacto del juego en la satisfacción del usuario, mostrando mejoras notables en la capacidad de resolución de problemas y el compromiso de los estudiantes. El estudio también destaca el potencial de GUESS-18 como una métrica valiosa para contextos educativos, lo que indica su utilidad más amplia más allá del simple entretenimiento. Los hallazgos positivos sugieren nuevas direcciones para incorporar herramientas de experiencia de usuario en software educativo, con el objetivo de mejorar la calidad de la tecnología educativa en el campo de la ingeniería informática.

**Palabras Clave:** GUESS-18, experiencia de usuario, juegos educativos, educación en ingeniería de software, aprendizaje interactivo, diseño de juegos.

### ABSTRACT

This article details a study that examines the user experience in a software engineering educational game through the Game User Experience Satisfaction Scale (GUESS-18). It involved 111 computer engineering students who played a game designed to improve their software skills via interactive tasks. The GUESS-18 helped measure the game's impact on user satisfaction, showing notable gains in students' problem-solving abilities and engagement. The study also highlights the potential of GUESS-18 as a valuable metric for educational contexts, indicating its broader utility beyond just entertainment. The positive findings suggest new directions for incorporating user experience tools into educational software, with the goal of enhancing the quality of educational technology in the field of computer engineering.

**Keywords:** GUESS-18, User Experience, Educational Gaming, Software Engineering Education, Interactive Learning, Game Design.

## ► I. Introducción

Technological advances led to changes in educational and assessment processes, especially in the field of programming and Computer Science. Traditional assessments based on sequential questions may not accurately reflect students' abilities, because external factors might influence their answers. With the aim to find a didactic and playful way to assess programming students' skills, use video games has been proposed as an assessment approach [1].

Software engineering is evolving rapidly, with a growing emphasis on interactive and hands-on learning approaches. Video games may be the answer to better assessment methods. "Logic-Driven Learning,"[2] an interactive video game designed for this purpose, embodies this shift by offering a platform where users can engage with software engineering concepts through gameplay. This game aims to enhance students' programming logic and understanding of software engineering principles in an engaging and dynamic environment. The importance of such tools in education is increasingly recognized, particularly in addressing the high attrition rates in software engineering courses, which often stem from a lack of practical, real-world application of theoretical knowledge.

To effectively measure and enhance the educational impact of "Logic-Driven Learning," it is crucial to assess not only the learning outcomes but also the user experience and satisfaction. This is where the Game User Experience Satisfaction Scale (GUESS),[3] a comprehensive tool for evaluating user satisfaction in gaming environments, becomes integral. The original GUESS, with its 55 items, provided an extensive framework for such assessments. However, its length posed practical challenges in situations requiring rapid and repeated evaluations, such as in iterative game design. Consequently, the development of GUESS-18, an abbreviated version of the original scale, offers a more practical solution without compromising on comprehensiveness or reliability.

In this study, we present how "logic-based

learning" is developed and we have used GUESS-18 tests to evaluate how game design and content influence user experience and learning outcomes. This integration aims to provide a dual perspective: firstly, evaluating the game's effectiveness in imparting software engineering skills and, secondly, understanding the elements of user experience that contribute to or detract from the educational value of the game. By applying GUESS-18, we aim to validate this tool in a new context – educational gaming for software engineering – and to gather insights that could guide future enhancements to the game's design. This approach aligns with the emerging trend in educational technology that emphasizes the importance of user experience in learning outcomes, aiming to create more effective, engaging, and user-friendly educational tools.

## ► II. Related work

In this section are analyzed related works from three perspectives included in this research: (A) programming centered systems, (B) programming focused video games, and (C) video games focused on teaching software engineering good practices.

### A. Programming-centered systems

Currently there are several web pages that consist of solving logical algorithms allowing developers to choose the programming language they want to practice. These types of pages are used by large companies such as IBM to evaluate developers. Among these pages we can mention [4]:

1. LeetCode is an online platform that focuses on improving programming skills and preparing individuals for technical interviews.
2. CodeSignal is a platform that conducts assessments to facilitate the technical hiring process, from the initial evaluation to the interview.
3. HackerRank supports a variety of programming languages, allowing users to practice and demonstrate their skills in multiple languages.

The aforementioned web pages are used for job



interviews and to improve the logic of junior programmers, these pages use similar logic problems so that when developers arrive to an interview, they are better prepared.

### *B. Programming-focused video game*

In this section three video games are presented, which contains a set of puzzles to be programmed to advance through a set of levels. Three of them aims to evaluate the developers' knowledge and help them improve their logic.

1. CodeCombat is a website that offers a programming game to learn how to code. In this game, you can learn programming languages such as Python, JavaScript, and HTML while solving puzzles and learning to create your own coding games and websites
2. Robocode is a programming environment where the goal is to develop a battle robot to compete against other tanks. These battle robots are programmed in Java. The environment offers a code editor, a debugger, and a compiler, all in one.
3. Screeps is a massive multiplayer online real-time strategy game. The main feature of the game is that it allows players to program the artificial intelligence (AI) of their units using JavaScript. In other words, you control your colony by writing code.

### *C. Video game focused on teaching good practices in the software development process*

Smellware is a game that aims showing distinct types of smells in software code [5] and the best practices to solve them. This proposal aims to help software developers to recognize diverse types of smells and be aware of the effects of these bad practices in software development.

It is based on an educational game called Riskware [6], a game where the board represents the execution of a project in which different risks that affect the progress of the work team may be materialized. In this game, the team members have an initial budget to purchase resources and controls allowing them to avoid the risks that

could be materialized during the development of the game.

## » III. Description of the proposed platform

This section provides technical information regarding the proposed platform.

### *A. Requirements*

The characteristics and requirements considered for this platform are the next listed:

- 1) A web application.
- 2) An intuitive and pleasant interface.
- 3) User registration and authentication system.
- 4) Tracking and likes system.
- 5) A publishing system.
- 6) An educational video game, intuitively playable.
- 7) The system must have also an API [7] to be able to manipulate information between two different systems, video game and web system.
- 8) A section to register/edit questions to the administrator users so that this question can be part of the game.
- 9) And it must have a real time system to display data.

### *B. Development methodology*

Agility is about quickly responding needs and demands to the market and the customer and being able to change direction as the situation requires. Agile methods attempt to maximize the delivery of value to the customer and minimize the risk of creating products that do not or fail, to meet market or customer needs [8].

Due to the specific features of this project, the agile method chosen to carry out this platform was SCRUM [9].

### *C. Framework*

To build the platform we used different frameworks: Laravel for developing the web system; Unity for developing the video game. Besides, we use the MVC pattern for developing the web system [10].

Fig. 1 shows the pattern will help us to organize correctly the code and to communicate with the client.

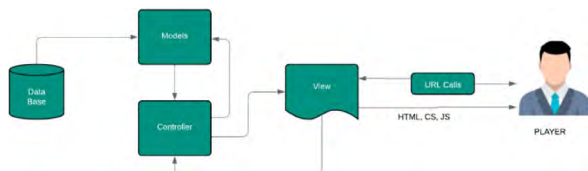


Fig. 1. Architecture Model, View, Controller.

Although it is true that the entire web system can be created using the MVC pattern [11], a small portion is developed using the REST pattern,[12].

For the development of the video game a system called FSM [13] was implemented, which stands for Finite State Machine. It allowed us to develop playable and non-playable characters, with a pattern of states. Besides, it helped us to have a more adequate organization of the platform and to reuse code no matter which sprit or character uses it. This way allowed achieving the possibility that admin users register a question on the page as part of the game; and to be able to create our own API to send data with a satisfactory communication between the web system and the video game.

#### D. Database

For the implementation of the database we use Mysql, so that the database was planned and developed to be relational. The system database has of the following components:

- 1) Users, its function is to store all user data, it has: unique id, name, email, password, image and username.
- 2) Publications, to store all the publications that a user can make, composed of: Unique id, title, description, image, and user id.
- 3) Followers, to store the records of users that follow each other, the table is composed of unique id, user id, and follower id.
- 4) Likes, to store the likes records that a publication has, this record is made up of unique id, user id, and publication id.
- 5) Comments, to store the comments that a publication can have, the table has a unique id, user id, publication id, and comment.

- 6) Levels, which main function is to store the information that a user can have in at a specific moment of finishing a level, this same information will be used to graph the progress of each player. The record is composed of unique id, user id, level number, time, errors, deaths, and extra points.

#### E. Communication

In order to solve an obstacle that was presented at the time of development of the platform between the web system and the video game, we decided to use the REST framework to create our own API to manipulate the information between systems.

#### F. Tests

The system starts with the web system interface, where the user authentication is required, done by email and password. It should be noted that the other views are protected, which means that if a person tries to enter the game without going through the start section, it will not allow that person to play.

In the statistics data interface is where the player can see black graphs of the scores, deaths and errors that the player has made, the average time to pass the level, as well as the best time recorded. Additionally, it provides a list of the best scores of all active players [12].

#### G. Software quality improvement activities

The following section aims to justify how the creation of this video game will help players to implement good software engineering practices in the future. The case proposed in this paper encourages the use of best development practices in coding and testing. Then, the use of these video game can help to meet requirements and deliver value to users [14].

- 1) *Problem decomposition*: the video game is designed to present complex puzzles and challenges that require proper problem decomposition. Players must break challenges into smaller, more manageable tasks, which encourages the practice

of breaking down complex problems into simpler and more reachable goals. This skill is critical in software engineering, as it allows for the development of more structured and maintainable solutions.

2) *Modularity and code reuse*: as players progress through the game, they face increasingly difficult challenges that require more complex solutions. Players learn to break their code into independent, reusable modules, which facilitates future problem solving and promotes greater efficiency in software development.

3) *Testing and debugging*: the video game aims to develop the ability to test and debug players' code by facing errors and bugs in the execution of their solutions, so players gain experience in debugging techniques and learn the importance of testing. By this way, the video game introduces the idea that bug detection and correction are an integral part of the software development process, thus encouraging good practices in terms of code quality and robustness.

4) *Improved code quality*: code review allows identifying and

correcting bugs, inefficiencies and bad practices in the code developed by the players. By submitting their code to peer review or subject matter experts, players can receive valuable feedback on how to improve their coding style, optimize algorithms and apply better programming practices. This leads to improved code quality and the ability of players to develop more efficient and robust solutions.

5) *Code maintenance*: in the video game context, players can learn and practice code maintenance techniques, such as refactoring and bug fixing, as they develop solutions to puzzles. Regular practice of code maintenance helps players understand the importance of readability and modularity, leading to cleaner, easier to understand and maintain code. Also, by fostering the habit of maintaining code in good condition, players develop a quality-oriented mindset and continuous improvement.

The main view of the development platform in Unity [13], which contains all the graphical part of the game, as well as functionalities and animations. This platform has a function called "Platform-dependent compilation" and consists of a way to compile and execute the code to evaluate it on the platform where the video game is presented.

#### *H. Public repository*

The platform was carried out under the Creative Commons license, which means we share the base code and the link to try the game.

- <https://github.com/InventoresDIVTIC/DevGAME-Unity>
- <https://devgame.inventores.org/>

#### *Justification of systems architecture and programming*

To build the 2D design [15], animations, as well as the functionalities of the video game were developed with the Unity engine [16], using the C# programming language for the development of the scripts focused on object-oriented design [17]. For the interactive visualization of the video game, it was necessary to create a web page, exporting it in WebGL format [18], [19] and implemented in the JavaScript programming language to render graphics and thus independently represent them as interactive graphics in the web browser [20].

The web page, created with the Laravel [22] framework, uses the PHP programming language, allowing us to create it in a simpler and more presentable way. Likewise, the CSS framework called Tailwind [23] was used in the design part for its presentation.

For the database management system, we used MySQL, where all the information related to the web page and the users is stored.

#### *Justification of Intelligent Systems*

As mentioned above, we decided to make a more complete implementation for the mechanics performed by the enemies, in order to make the interaction with the users more

dynamic. To achieve this, we developed different implementations focusing on the operation of an intelligent agent developed with state machines, in addition; we implemented a decision tree and a search algorithm called "A\* Pathfinding"[12] when deciding the shortest route in the direction of the player.

In the implementation of the A\* Pathfinding algorithm to find the shortest path to the player, we use the Euclidean distance as a heuristic. The Euclidean distance is calculated in a two-dimensional (x, y) plane and represents the straight distance between two points. In our case, these points are the current position of the enemy and the position of the player.

The A\* Pathfinding algorithm uses this Euclidean distance as an estimation of how far it is to the target, which allows us to select the most favorable nodes in the optimal pathfinding. By combining the Euclidean distance with the distance traveled so far, A\* Pathfinding can make informed decisions about which nodes to explore next, and thus avoiding obstacles in the environment.

Node selection is performed based on the value of an evaluation function called "f(n)". For each candidate node, the value of "f(n)" is calculated by summing the distance traveled from the starting point to the current node (denoted "g(n)") and the estimated distance from the current node to the target (denoted "h(n)").

g(n): It is the cumulative cost from the initial node to the current node. In other words, it represents the distance traveled from the start to the current node on the route currently under consideration.

- h(n): This is the heuristic function that estimates the distance from the current node to the target. In this case, the Euclidean distance is used to calculate it. The specific formula is:  $h(n) = \sqrt{(n.x - goal.x)^2 + (n.y - goal.y)^2}$ . Where "n.x" and "n.y" represent the (x, y) coordinates of the current node, and "goal.x" and "goal.y" represent the (x, y) coordinates of the goal (player's position).

- "f(n)": It is the total value of the evaluation function and is calculated by adding the accumulated cost "g(n)" and the estimated distance "h(n)". Therefore,  $f(n) = g(n) + h(n)$ . The value of "f(n)" represents an estimate of the total cost from the start to the target via the current node.
- The A\* Pathfinding algorithm selects the node with the lowest value of "f(n)" at each step, which means that the nodes that seem most promising in terms of finding an optimal route are explored first. This combination of distance traveled and estimated distance helps guide the enemy to the player in the most efficient way possible, avoiding obstacles and finding the shortest route.

#### *Justification of Distributed Systems*

A programming-related question and answer system was implemented on the web site with the ability to manage and modify questions. There are two user roles: administrator and common user. The administrator can add, delete and edit questions, which will then be displayed in the video game. Different answer options can be added to increase the diversity of the quizzes. Changes made by the administrator are instantly reflected on the website. The video game displays random questions and switches questions when the user answers incorrectly, applying a penalty. This allows feedback and scores to be obtained. After answering each question, users can see their performance on the website, indicating whether it was correct or incorrect.

As result of developing the platform: 1) a presentable, intuitive, and interactive web system was developed to be as accessible to the user without the need to install a third-party program; 2) a web system for enabling an authentication process was developed, with the purpose of having security and a way for users to communicate through a forum; 3) we achieved a successful implementation of microservices for the manipulation of information between the web page and the video game, and 4) an educational video game was developed to measure the level of programming knowledge of

a user and interactively encourage good software development practices.

### *I. Gues-18 methodology*

Integration and validation of the Game User Experience Satisfaction Scale (GUESS-18) was performed within "Logic Driven Learning," an educational video game designed to improve software engineering skills. Our methodology encompasses participant selection, gameplay, data collection and analysis to evaluate the impact of the game on learning outcomes and user satisfaction which are described below.

#### A. Participant Selection

For this study, the participant selection process was meticulously designed to gather a representative sample from a specific academic setting. We recruited a total of 111 students from the University of Guadalajara, specifically from the University Center of Exact Sciences and Engineering. This cohort was chosen based on their background in programming, ensuring that all participants had a fundamental understanding of software engineering principles.

In terms of demographic criteria, the age range of the participants was set between 18 and 28 years. This age bracket was chosen to represent a typical undergraduate and early graduate student population, which is reflective of the primary user base for the "Logic-Driven Learning" game. The inclusion of participants who were already acquainted with programming was crucial. It ensured that the study's findings would be relevant and applicable to students actively engaging with software engineering concepts, thus providing more meaningful insights into the game's effectiveness as an educational tool.

By selecting participants from the University Center of Exact Sciences and Engineering at the University of Guadalajara, the study aimed to maintain a high level of consistency in the educational background of the participants. This consistency was pivotal in assessing the impact of the game on enhancing software engineering skills

and in evaluating the user experience through the GUESS-18 survey with a homogeneous group, ensuring that the results were both relevant and applicable to the field of software engineering education.

#### B. Pre-Game Setup

Before the commencement of the gameplay, a critical stage of preparation, termed the Pre-Game Setup, was diligently conducted to establish a baseline and ensure all participants were adequately oriented. This phase was essential for aligning the participants' expectations and understanding of the study's purpose and the game's objectives.

**Orientation Session:** An integral component of this phase was the Orientation Session. Here, participants were introduced to "Logic-Driven Learning," the educational video game at the heart of the study. This session was not merely a walkthrough of the game mechanics but also an informative briefing on the research objectives. Participants were informed about the importance of their feedback for the study and how it would contribute to enhancing educational tools in software engineering. The session aimed to foster a sense of engagement and responsibility among the participants, encouraging them to be attentive and reflective during their gameplay experience.

**Baseline Data Collection:** Alongside the orientation, a Pre-Game Survey was administered. The purpose of this survey was to gather crucial baseline data regarding the participants' prior knowledge and experience in both software engineering and gaming. This data would later serve as a comparative measure to assess the impact of the game on the participants' learning and user experience. By understanding their initial skill levels and attitudes towards gaming and software engineering, the research team could more accurately evaluate the game's effectiveness and the participants' progress.

The Pre-Game Setup phase was meticulously planned to ensure that all participants started the game with a clear understanding of what to

expect and how their involvement was pivotal to the research. It set the stage for a controlled and informed study environment, crucial for the validity and reliability of the research outcomes.

Integrating the GUESS-18 survey into “Logic Driven Learning” required a thoughtful and personalized approach, ensuring that the survey resonated with the specific context of the game while maintaining its primary purpose of evaluating the user experience.

The process for Integrating the survey into the game was approached with the player experience in mind. We strategically placed the GUESS-18 survey prompts at points in the game where they would be most meaningful and least intrusive. For example, after completing a significant level or after a particularly challenging software engineering problem, the survey was presented to players. This moment was chosen to capture immediate, thoughtful feedback while maintaining a seamless gaming experience.

For final testing before full deployment, this version of GUESS-18 underwent a series of tests with a small group of 4 to 5 players. Their feedback was used to further refine the survey wording and integration points within the game, ensuring clarity and relevance.

Through this process the GUESS-18 survey was effectively integrated into “Logic Driven Learning,” prepared to capture valuable information about players' learning experiences and satisfaction levels. This integration was instrumental in evaluating the effectiveness of the game as an educational tool in software engineering and in understanding the nuances of the user experience in an educational game context.

### C. Data Collection

The Data Collection phase of the study, integral to the evaluation of "Logic-Driven Learning," was meticulously structured to capture a comprehensive understanding of the players' experiences using the GUESS-18 survey, supplemented by additional

demographic and experience-related questions.

**Implementation of GUESS-18 via Google Forms:** To facilitate the collection of survey data, we utilized Google Forms, a reliable and user-friendly platform. This approach not only streamlined the data collection process but also provided a familiar interface for participants, potentially increasing response rates and accuracy. The Google Forms survey embedded the 18 questions from the GUESS-18, carefully adapted and integrated as previously outlined.

**Additional Questions for Contextual Insight:** Alongside the standard GUESS-18 questionnaire, four additional questions were incorporated into the survey. These questions were designed to gather demographic data and information about the participants' programming experience, crucial for contextualizing their feedback on the game. The added questions were:

- **Age:** To understand the age distribution of our participant group and to see if there were any age-related trends in the feedback.
- **Sex:** To capture gender-based perspectives, if any, on the gaming and learning experience.
- **Programming Experience:** To ascertain whether participants had prior experience in programming, which could influence their interaction with the game and their perception of its educational value.
- **Self-Assessment of Programming Skills:** Participants were asked to rate their own programming skills. This self-assessment helped in correlating perceived skill levels with user experience and satisfaction, providing insights into how the game benefits learners at different skill levels.

**Deployment and Response Monitoring:** The survey was deployed at specific intervals during gameplay, as well as at the conclusion of the gaming session. The timing was strategically chosen to capture feedback that was both immediate and reflective of the overall experience. Continuous monitoring

of survey responses was carried out to ensure a high response rate and to address any technical issues promptly.

Through this comprehensive data collection approach, employing Google Forms for ease of use and adding crucial demographic and experience-related questions, the study aimed to gather rich, multi-dimensional data. This data was pivotal in assessing the educational impact and user experience of "Logic-Driven Learning," thereby contributing to the broader understanding of user engagement and satisfaction in educational gaming environments.

#### ► IV. Results and discussions

The Fig. 2 show bar's graph that represents the age distribution of 111 participants who completed a Google Forms survey for the usability study of the "Logic-Driven Learning" game. Ages range from 17 to 40 years, with the majority of responses concentrated in the 25-year-old group, which has 24 participants (21.6%). The next highest concentrations of responses are at ages 23 and 21 years, with 21 (18.9%) and 14 (12.6%) participants, respectively. The distribution indicates that the predominant group of players providing feedback consists of young adults in the early stages of university or postgraduate education.

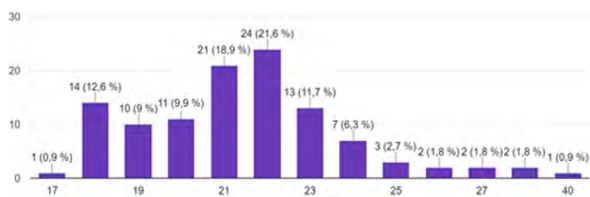


Fig. 2 Graph of number of respondents and their age.

The Fig. 3 displays a pie chart reflecting responses to the question "Do you have programming experience?" from the 111 participants who completed the Google Forms survey for the usability study of the "Logic-Driven Learning" game. The chart shows a significant majority, 93.7%, indicating "Yes" (represented in blue), suggesting they have programming experience. A smaller wedge, 6.3% (depicted in red), represents the participants who answered "No," indicating

they do not have programming experience. This data highlights that the vast majority of respondents are familiar with programming, providing context for the survey responses within the domain of software engineering education.

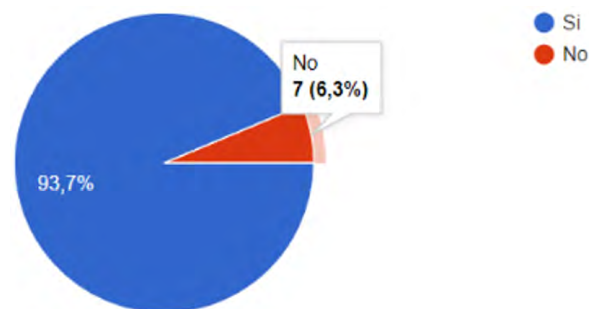


Fig. 3 Pie chart reflecting responses to the question "Do you have programming experience?"

Fig. 4 is a bar graph summarizing responses to a usability question from the GUESS-18 survey about the game "Logic-Based Learning." Most participants found the game's controls easy to use, with 78 of 111 indicating they agreed or strongly agreed that the controls are straightforward and simple. A minority of 18 participants disagreed or strongly disagreed with the statement, while 15 participants were neutral.

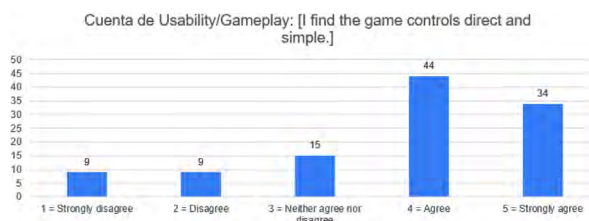


Fig. 4 Graph of the game controls direct and simple.

Fig. 5 shows responses to the GUESS-18 survey question about the usability of the game interface, specifically whether players find the interface easy to navigate. The majority of participants responded positively: 45 agree (4 on scale) and 40 strongly agree (5 on scale), suggesting that they find the interface easy to use. A minority of respondents, 7 strongly disagree (1 on scale) and 6 disagree (2 on scale), did not find the interface easy to navigate. Thirteen participants are neutral (3 on the scale). The results indicate that the majority of players had a favorable experience with the game interface.

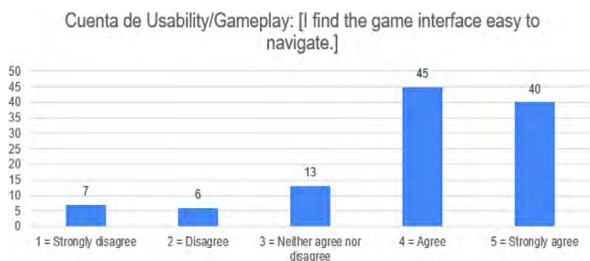


Fig. 5 Graph of the game interface easy to navigate.

Fig. 6 shows responses about the game's narrative, specifically whether players are captivated by the story from the beginning. The most common response is neutral: 43 participants selected “Neither agree nor disagree” (3 on the scale). This is followed by 25 participants who “strongly disagree” (1 on the scale) and 21 who “disagree” (2 on the scale), indicating a lack of engagement with the story. At the other extreme, 16 agree (4 on the scale) and 6 strongly agree (5 on the scale), which shows a certain level of commitment. These results suggest that while some players are interested in the story, a notable portion of them are not enthralled from the start.

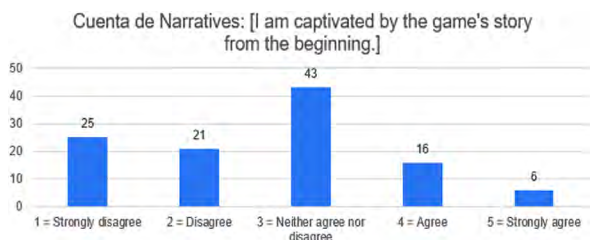


Fig. 6 Graph of captivated by the game's story from the beginning.

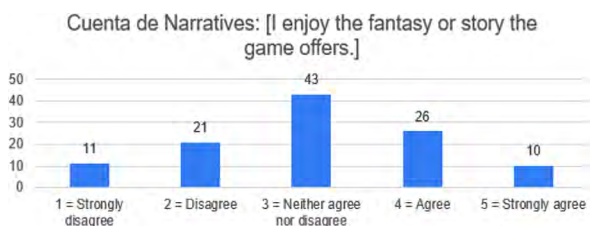


Fig. 7 Graph of the fantasy or story the game offers.

Fig. 7 presents responses to the narrative engagement question from the GUESS-18 survey: “I enjoy the fantasy or story the game offers.” The majority of participants, 43, are neutral, neither agreeing nor disagreeing. Twenty-six participants agree with the statement and 10 strongly agree, indicating that they enjoy the story of the game.

In contrast, 21 disagree and 11 strongly disagree, suggesting that they do not enjoy the narrative as much. This distribution implies a varied reception of the game's story among players.

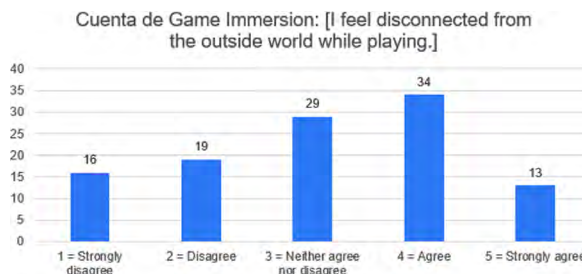


Fig. 8 Graph of the outside world while playing.

Fig. 8 is a graph that describes how players feel disconnected from the outside world while playing. The results show a distribution across the spectrum: 16 participants strongly disagree with feeling disconnected, 19 disagree, 29 neither agree nor disagree, 34 agree, and 13 strongly agree. Most participants feel some level of immersion, with a combined total of 47 of 111 indicating that they agree or strongly agree with the statement, suggesting that the game has a substantial immersive effect on these players.

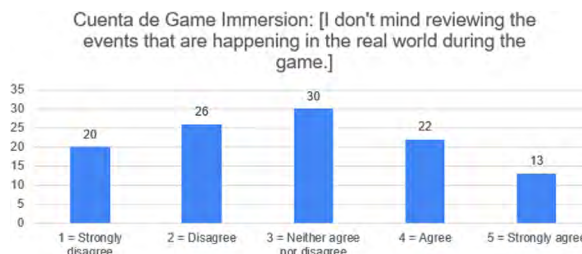


Fig. 9 Graph of the events that are happening in the real world during the game.

Fig. 9 shows responses to a GUESS-18 survey question on game immersion: “I don't mind reviewing events that happen in the real world during the game.” Player opinions appear to vary: 20 strongly disagree with being indifferent to real-world events while playing, suggesting high immersion, while 26 disagree and 30 neither agree nor disagree. . Meanwhile, 22 agree and 13 strongly agree, indicating that they are not as immersed and do not mind being connected to the real world during the game. The data suggest a balanced distribution of immersion levels among participants.



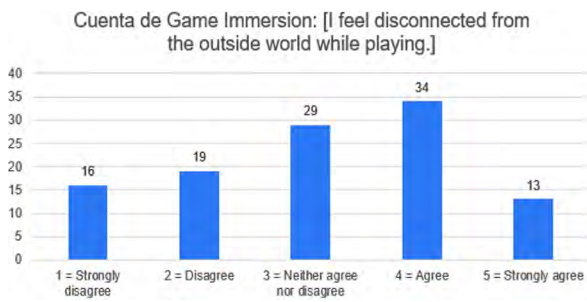


Fig. 10 Graph of the game is fun.

Fig. 10 is the responses to the enjoyment question from the GUESS-18 survey: “I think the game is fun.” A considerable number of participants, 40, agree that the game is fun, while 23 strongly agree, indicating a high level of enjoyment. On the other hand, 7 participants totally disagree and 13 disagree that the game is fun. Twenty-eight participants are neutral, neither agree nor disagree. Overall, most responses lean towards the game being perceived as fun.

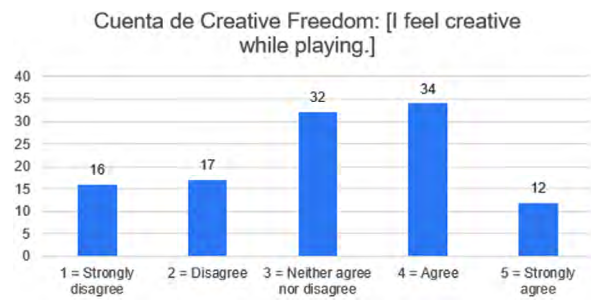


Fig. 13 Graph of the creative while playing.

The Fig. shows a distribution in which most players agree (34 responses) or strongly agree (12 responses) that they feel creative during the game, indicating a sense of creative engagement. A moderate number of participants are neutral (32 responses), while fewer participants disagree (17 responses) or strongly disagree (16 responses) with the statement, suggesting that they feel less creative freedom while playing. Overall, the answers lean towards gameplay, providing a sense of creative freedom to players.

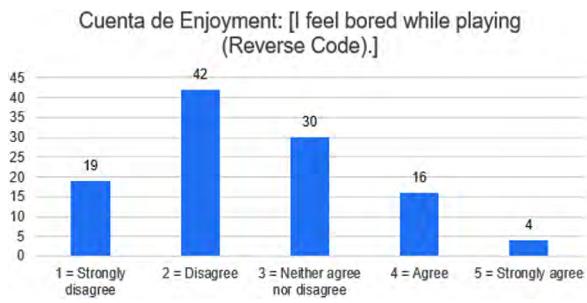


Fig. 11 Graph of the bored while playing.

Fig. 11 indicates that 42 participants disagree and 19 strongly disagree with feeling bored, suggesting that they find the game engaging. Thirty participants adopt a neutral stance, while a minority of 16 agree and 4 strongly agree with feeling bored, indicating a lower degree of enjoyment. The majority sentiment leans towards the game not being boring for the players.

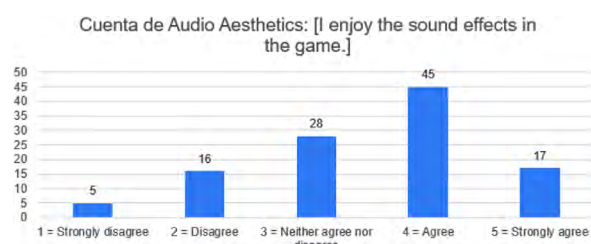


Fig. 14 Graph of the sound effects in the game.

Fig. 14 shows the results of a GUESS-18 survey question on audio aesthetics: “I enjoy the game’s sound effects.” The majority of participants, 45, say they enjoy the sound effects, while 17 strongly agree, indicating a positive reception of the game’s audio. A moderate number, 28, are neutral and neither agree nor disagree. Fewer respondents, 16, disagree, and an even smaller group, 5, strongly disagree with enjoying the sound effects. This suggests that most players are satisfied with the game’s audio experience.

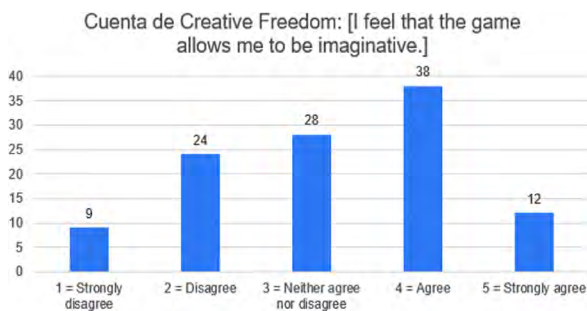


Fig. 12 Graph of the game allows me to be imaginative.

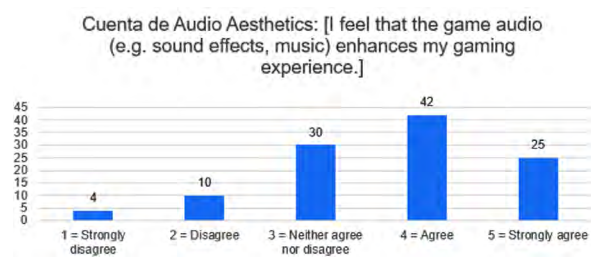


Fig. 15 Graph of the game audio.

Fig. 15 shows a bar graph from the GUESS-18 survey, illustrating responses to the statement “Audio Aesthetics: [I feel that game audio (e.g., sound effects, music) enhances my gaming experience. ]”. Most participants agree (42) or strongly agree (25) that audio contributes positively to their gaming experience. Thirty participants are neutral, while a smaller number disagree (10) or strongly disagree (4). The data suggests that most gamers generally perceive in-game audio as a factor that enhances the gaming experience.

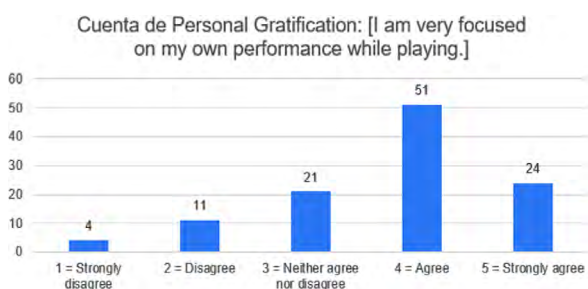


Fig. 16 Graph of the performance while playing.

Fig. 16 shows responses to an item from the GUESS-18 survey on Personal Gratification: “I am very focused on my own performance while playing.” The largest group of respondents, 51, agree with the statement, suggesting a high level of self-focus on performance. Twenty-four strongly agree, further emphasizing this approach. A smaller number of participants, 21, are neutral, while 11 disagree and only 4 strongly disagree, indicating that they are less concerned about their performance. This indicates that most players are attentive to their performance during the game.

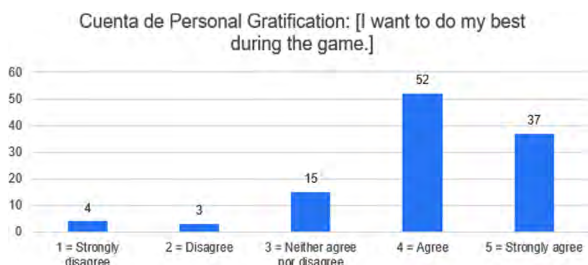


Fig. 17 Graph of the Personal Gratification

Fig. 17 shows responses to the Personal Gratification survey question: "I want to do the

best I can during the game." A significant majority of players agree (52 responses) or strongly agree (37 responses) with the desire to perform well, suggesting a high level of motivation and personal investment in the game. Only a small number of participants disagree (4 responses) or strongly disagree (3 responses), and 15 participants are neutral on the matter. This shows that most players are committed to excelling in the game.

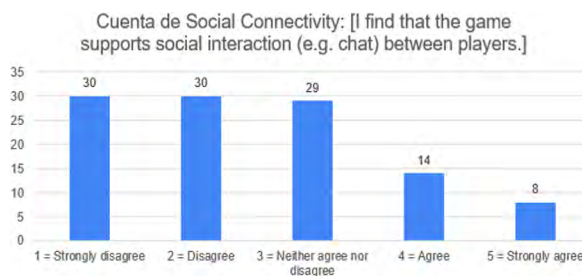


Fig. 18 Graph of the the game supports social interaction.

Fig. 18 shows responses to the question about social connectivity: “I think the game supports social interaction (e.g. chat) between players.” The most common responses are neutral, with 30 participants neither agreeing nor disagreeing. Both those who agree and those who disagree have responses equal to 30 each, indicating a division of opinion on the social characteristics of the game. A minority of participants, 14, agree and 8 strongly agree that the game supports social interaction, suggesting that while some find the social aspects adequate, others do not feel that the game facilitates interaction.

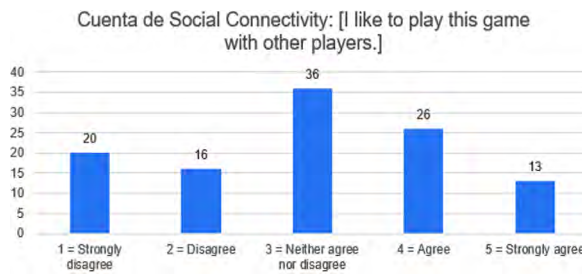


Fig. 19 Graph of the Social Connectivity.

Fig. 19 shows that the largest group of respondents, 36, are neutral. Twenty respondents strongly disagree with the preference of playing with others, while 16 disagree. In contrast, 26 agree and 13 strongly agree that they enjoy the game more when playing with others, indicating a variety of preferences for social play within the game.

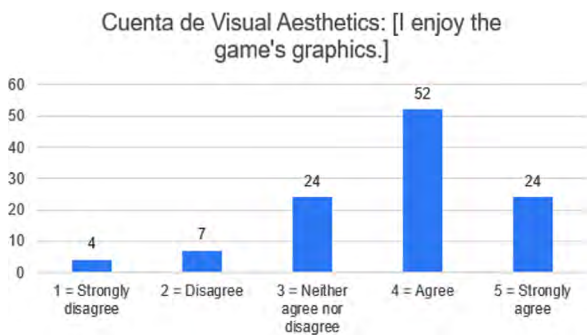


Fig. 20 Graph of the Visual Aesthetics.

Fig. 20 shows responses to the visual aesthetics question from the GUESS-18 survey: "I enjoy the game's graphics." The majority of participants, 52, agree that they enjoy the graphics, while 24 strongly agree, indicating a positive response to the game's visual design. A smaller group of 24 respondents is neutral, neither agreeing nor disagreeing. Very few respondents disagree (7) or strongly disagree (4), suggesting that the graphics are generally well received among gamers.

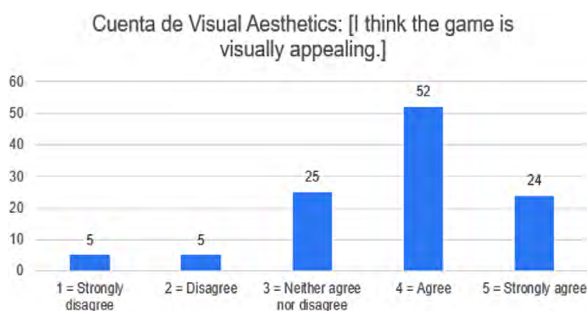


Fig. 21 Graph of the game is visually appealing.

Fig. 21 shows the most common response is agree: 52 participants selected "Agree" and an additional 24 chose "Strongly Agree," indicating a favorable opinion of the game's visual design. A smaller number of participants, 25, are neutral, while very few respondents disagree (5) or strongly disagree (5) with the visual appeal of the game. This suggests that most players find the game's visuals attractive.

## ► V. Conclusions and future work

After reviewing the 18 graphical representations of the GUESS-18 assessment for "Logic-Based Learning: An Interactive Video Game to Encourage Software Engineering," several nuanced

conclusions are drawn that offer a complete picture of user experience and educational effectiveness of the game:

- Usability and control mechanics:**  
The data overwhelmingly indicates that the game's control mechanics are easy to use, and most players find them straightforward and simple. This suggests that the game developers have successfully implemented an intuitive interface, which is crucial to maintaining player engagement and facilitating the learning process within the software engineering context of the game.
- Game Immersion:**  
There seems to be a wide range of experiences related to immersion, and a notable portion of players feel disconnected from the outside world, a hallmark of a compelling gaming experience. However, the split in responses also points to a segment of players who remain less absorbed. This dichotomy suggests there is room to refine game design to consistently improve immersion across the entire player base.
- Social Connectivity:**  
Responses reflect ambivalence regarding the game's facilitation of social interaction. While some players appreciate the social elements, a significant portion remain neutral or dissatisfied. This indicates potential opportunities to deepen the social features of the game, possibly improving communication tools or collaborative challenges within the game.
- Enjoyment and commitment:**  
The game appears to score high on enjoyment, with many players expressing that they find the game fun and are willing to give their best while playing, indicating high levels of engagement and motivation. This positive feedback is crucial as it underlines the potential of gaming as an effective educational tool, harnessing enjoyment to improve learning outcomes.

- **Personal gratification:**  
Responses show that players generally focus on their performance, suggesting that the game successfully encourages players to invest in skill development and mastery, which are key aspects of educational games.
- **Creative freedom:**  
A considerable number of players feel a sense of creative freedom while playing, indicating that the game likely offers elements that encourage creative problem solving, a valuable skill in software engineering.
- **Audiovisual and Visual Aesthetics:**  
The visual and audio elements of the game receive positive feedback and most players enjoy the graphical and audio experience, which contributes positively to their overall gaming experience. High-quality audiovisuals are often associated with higher levels of user satisfaction and can play an important role in facilitating an immersive learning environment.

In summary, “logic-based learning” appears to be a well-received educational tool that stands out for its usability, player enjoyment, and aesthetic appeal. However, there are areas for improvement in immersion and social connectivity. The game seems to have struck a chord with its target audience in terms of providing a compelling educational experience that is both enjoyable and conducive to learning. In the future, these insights could guide iterative improvements, improving the game's strengths and addressing its weaknesses to better serve its educational purpose.

Building on the insights gained from the GUESS-18 evaluation, future work on “Logic-Driven Learning: An Interactive Video Game to Foster Software Engineering” could focus on enhancing player immersion. A detailed analysis of the moments where players felt most and least immersed could inform the development of new game levels or features. Immersion can be deepened through richer storytelling, dynamic challenges that adapt to the player's skill level, and more interactive game mechanics that

closely simulate real-world software engineering problems. Additionally, leveraging virtual reality (VR) or augmented reality (AR) could offer a more immersive experience, potentially increasing the user's sense of presence within the game.

Social connectivity also presents an avenue for expansion. Given the mixed responses about the game's current social features, future iterations could introduce more collaborative projects or pair programming challenges that require communication and cooperation between players. Community-driven features like leaderboards, forums for sharing solutions, and in-game rewards for teamwork could foster a stronger sense of community. Moreover, integrating social media functionalities could allow players to share achievements and insights, thereby extending the game's reach and impact.

Lastly, there is room for further tailoring the game to the educational objectives of software engineering. Continuous user feedback loops could be established to refine the game's content and ensure that it remains up-to-date with the latest industry practices and technologies. Collaborations with educational institutions and software engineering professionals could provide real-world applicability to the game's scenarios. Research into the game's long-term impact on players' learning outcomes, career trajectories, and retention of software engineering concepts could validate and shape the game's role as an effective educational tool.

## » VI. Referencias

- [1] B. Whitfield. "What Is a Game Developer? How to Become One, Salary, Skills. | Built In". built in. <https://builtin.com/learn/careers/game-developer> (Accessed July 12, 2023)
- [2] D. Bonilla Carranza, A. Peña Pérez Negrón y M. Muñoz, “Logic-Driven Learning: An Interactive Video Game to Foster Software Engineering”, Encuentro Nac. Comput. 2023, 2023.
- [3] Keebler, J. R., Shelstad, W. J., Google,

- D. C. S., Chaparro, B. S., & Phan, M. H. (2020). Validation of the GUESS-18: A Short Version of the Game User Experience Satisfaction Scale (GUESS). <http://uxpajournal.org>.
- [4] J. Carrillo, "Los 10 Sitios Web Más Populares para Desafíos de Programación [Actualizado para 2020]", [freecodecamp.org](https://www.freecodecamp.org), 01-dic-2020. [Online]. Available at: <https://www.freecodecamp.org/espanol/news/los-10-sitios-web-mas-populares-para-desafios-de-programacion-actualizado-para-2020/>. [Accessed: 31-ene-2023].
- [5] C. M. Zapata, M. C. Gomez y J. C. Hernandez. "SMELLWARE: un juego para la enseñanza de buenas prácticas en el proceso de desarrollo de software". SciELO - Scientific electronic library online. [https://www.scielo.cl/scielo.php?script=sci\\_ext&pid=S0718-33052020000400645](https://www.scielo.cl/scielo.php?script=sci_ext&pid=S0718-33052020000400645) (accedido el 12 de julio de 2023).
- [6] C. Zapata Jaramillo, M. Gomez Álvarez and G. González Calderón. "Riskware: A Game for Teaching Software Project Risk Management". *Developments in Business Simulation and Experiential Learning*. Vol. 40. 2013.
- [7] "¿Qué es una API REST? | IBM". IBM - Deutschland | IBM. <https://www.ibm.com/mx-es/topics/rest-apis> (Accessed July 12, 2023).
- [8] M. Alqudah and R. Razali, "A Review of Scaling Agile Methods in Large Software Development," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, no. 6, pp. 1-10, Dec. 2016. [Online]. Available: <https://dx.doi.org/10.18517/IJASEIT.6.6.1374>
- [9] "Scrum: qué es, cómo funciona y cómo empezar | Atlassian". Atlassian. <https://www.atlassian.com/es/agile/scrum> (Accessed July 12, 2023).
- [10] "Información general de ASP.NET MVC". Microsoft Learn: Build skills that open doors in your career. <https://learn.microsoft.com/es-es/aspnet/mvc/overview/older-versions-1/overview/asp-net-mvc-overview> (Accessed July 12, 2023).
- [11] Introduction to A\*. Loading... <https://theory.stanford.edu/~amitp/GameProgramming/AStarComparison.html> (accessed July 12, 2023).
- [12] "Qué es MVC", [Desarrolloweb.com](https://desarrolloweb.com), 02-ene-2014. [Online]. Available at: <https://desarrolloweb.com/articulos/que-es-mvc.html>. [Consultado: 2023].
- [13] reeCodeCamp.org. "Finite State Machine Explained". [freeCodeCamp.org](https://www.freecodecamp.org). <https://www.freecodecamp.org/news/finite-state-machines/> (accedido el 12 de julio de 2023).
- [14] Download Archive y Beta Program, "ProgressBar Pack", Unity.com. [Online]. Available at: <https://assetstore.unity.com/packages/tools/gui/progressbar-pack-120981>. : 2023].
- [15] Download Archive y Beta Program, "Medieval pixel art asset FREE", Unity.com. [Online]. Available at: <https://assetstore.unity.com/packages/2d/environments/medieval-pixel-art-asset-free-130131>. [Accessed: 2023].
- [16] E. Román. "Mejores Prácticas de Desarrollo de Software". Centro de Mejora - El blog de Innevo. <https://blog.innevo.com/mejores-practicas-desarrollo-software> (Accessed July 12, 2023).
- [17] "Unity", Microsoft. [Online]. Available at: <https://dotnet.microsoft.com/es-es/apps/games/unity>. [Accessed: 31-ene-2023].
- [18] M. Pêcheux, "How to create a simple behaviour tree in Unity/C#", *Geek Culture*, 02-nov-2021. [Online]. Available at: <https://medium.com/geekculture/how-to-create-a-simple-behaviour-tree-in-unity-c-3964c84c060e>. [Consultado: 2023].
- [19] Unity Technologies, "Empezar con el desarrollo de WebGL", [Unity3d.com](https://unity3d.com). [Online]. Available at: <https://docs.unity3d.com/es/2018.4/Manual/webgl-getting-started.html>. [Accessed: 2023].
- [20] "Primeros pasos con WebGL", [Mozilla.org](https://mozilla.org). [Online]. Available at: [https://developer.mozilla.org/es/docs/Web/API/WebGL\\_API/Tutorial/Getting\\_started\\_with\\_WebGL](https://developer.mozilla.org/es/docs/Web/API/WebGL_API/Tutorial/Getting_started_with_WebGL). [Accessed: 2023].
- [21] Download Archive y Beta Program, "Pixel Adventure 2", Unity.com. [Online].

Available at: <https://assetstore.unity.com/packages/2d/characters/pixel-adventure-2-155418>. [Accessed: 2023].

- [22] "Installation", Laravel.com. [Online]. Available at: <https://laravel.com/docs/9.x>. [Consultado: 2023].
- [23] "Documentation - Tailwind CSS". Tailwind CSS - Rapidly build modern websites without ever leaving your HTML. <https://v2.tailwindcss.com/docs> (Accessed July 12, 2023).

# HACIA UN MODELO DE GOBIERNO DE APIs, MAPEO SISTEMÁTICO DE LA LITERATURA

## Towards an API Governance Model: Systematic Mapping of the Literature

Elizabeth Suescún Monsalve <sup>1</sup>	esuescu1@eafit.edu.co
Marta Tabares Betancur <sup>2</sup>	mtabares@eafit.edu.co
Liliana González Palacio <sup>3</sup>	lgonzalez8@eafit.edu.co
Mariana Vásquez Escobar <sup>4</sup>	mvasqueze@eafit.edu.co

<sup>1, 2, 3, 4</sup> Universidad EAFIT.

### RESUMEN

Las empresas digitales emergentes están provocando una profunda transformación que permea las grandes organizaciones, las sinergias permiten mantener su posición en los mercados modernos. Para comprender plenamente este ciclo de cambio, resulta imprescindible considerar el papel de las Interfaces de Programación de Aplicaciones (APIs), las cuales habilitan la transformación a través de innovadores modelos de negocio. En el ámbito financiero, es un hecho innegable que las Fintech utilizan de manera activa las APIs para impulsar los negocios digitales. No obstante, los bancos de gran envergadura se enfrentan a nuevos y desafiantes obstáculos para generar nuevo valor en colaboración con estos innovadores socios digitales, o bien, para adquirir nuevas capacidades de forma individual mediante la implementación estratégica de las APIs. En ese contexto, se presenta un mapeo sistemático de la literatura con el objetivo de mostrar una visión amplia y estructurada sobre el estado actual del conocimiento relacionado al gobierno de APIs en el sector financiero enmarcado en un proyecto de investigación para este sector y específicamente para recaudos en la banca. Del estudio se concluyó que las APIs son indispensables en los ecosistemas actuales, ellas permiten que consumidores, funcionarios, desarrolladores de terceros, Fintech, minoristas y otros componentes como algoritmos, datos, procesos y funcionalidades se comuniquen, Por

tanto, los negocios necesitan ser rediseñados para definir reglas que les permita competir y ampliar sus márgenes, sin las APIs no es posible abrazar el mundo digital.

**Palabras Clave:** APIs, Transformación Digital, Fintech

### ABSTRACT

Digital startups are driving a profound transformation that permeates large organizations, synergies that allow them to maintain their position in modern markets. To fully understand this cycle of change, it is imperative to consider the role of Application Programming Interfaces (APIs), which enable transformation through innovative business models. In finance, it is an undeniable fact that Fintechs are actively using APIs to drive digital business. However, large banks face new and challenging obstacles to generate new value in collaboration with these innovative digital partners, or to acquire new capabilities individually through the strategic implementation of APIs. In this context, a systematic mapping of the literature is presented with the aim of showing a broad and structured view on the current state of knowledge related to API governance in the financial sector framed in a research project for this sector and specifically for banking collections.

From the study it was concluded that APIs are indispensable in today's ecosystems, they allow consumers, officials, third-party developers, Fintech, retailers and other components such as algorithms, data, processes and functionalities to communicate, therefore, businesses need to be redesigned to define rules that allow them to compete and expand their margins, without APIs it is not possible to embrace the digital world.

**Keywords:** APIs, Digital Transformation, Fintech

## ► I. Introducción

Un mapeo sistemático de la literatura en la Ingeniería de Software sirve para inventariar y sistematizar la producción en determinada área del conocimiento. Este tipo de investigaciones son importantes para obtener comprensión y un estado de un dominio del tema a través de un proceso metodológico, adicional a eso, los resultados permiten evidenciar tendencias y vertientes. En ese sentido para el presente trabajo fue usado el protocolo propuesto por Petersen et al. [1] el cual es utilizado en investigaciones en la Ingeniería de Software por servir como guía metodológica a paso para realizar mapeos sistemáticos de la literatura.

En el entendimiento de la adopción de APIs en el sector financiero este proceso favoreció la organización de las fuentes encontradas, las perspectivas de investigación, los estudios recurrentes, y la identificación fuentes que nos permitieran tener una visión holística sobre el tema. Por tanto, la principal motivación para realizar este proceso fue la posibilidad de analizar y evaluar el conocimiento que está siendo generado en el contexto de APIs para el sector financiero. Los hallazgos permitieron entender lo que está siendo producido para acelerar la transformación digital de los bancos, y consecuentemente, la competitividad y la innovación, por tal motivo y como se explicará en este trabajo, el rediseño de la base tecnológica en el sector bancario hizo de las APIs un medio para adoptar el llamado modelo *Open Banking*, es así como, se mostrará la evolución de este concepto, las características, enfoques, así como

también vacíos y oportunidades de investigación, como también que aportaciones puede hacer esta información al recaudo en la banca.

Este documento está organizado de la siguiente manera: Trabajos Relacionados, Proceso de Búsqueda, Plan de la Revisión, Discusión, Trabajos Futuros, Conclusiones, Agradecimientos y Referencias.

## ► II. Trabajos Relacionados

Una interfaz de programación de aplicaciones o API por la abreviatura del término en inglés de *Application Programming Interface* desde la perspectiva del software es un conjunto de subrutinas, funciones y procedimientos, quiere decir, es una aplicación independiente a los lenguajes de programación y admite portabilidad, también expone interfaces para ser consumida o usada por otro(s) software(s), dicha exposición se conoce como -capa de abstracción- y su principal propósito es ser usada por desarrolladores de aplicaciones o por implementadores de servicios [2]. Para entender las APIs es necesario entender que las aplicaciones en el mundo de las APIs son diversas en su construcción y tecnologías, sin embargo, siguen patrones y especificaciones que luego son proveídas por las APIs para la correcta comunicación y orquestación de servicios. Para ilustrar, si se analiza un sitio de compras online que ofrece diferentes formas de pago, y que incluye, principalmente, tarjetas de crédito y débito en sus formas de pago, podemos deducir que ese site dentro de su funcionamiento involucra APIs. Dicha afirmación la podemos sustentar al ver que dicho sitio permite el uso de servicios que involucran bancos y operadores para ofrecer los servicios de tarjeta de crédito y débito. Internamente el sitio trabajará para recibir datos, realizar validaciones con seguridad y tranquilidad sin que el usuario final lo perciba.

Es por esto por lo que desde la perspectiva de una organización una API es un mecanismo que logra conectar el negocio para innovar, automatizar y sirve como instrumento para transformarse digitalmente, en pocas palabras, permite entrar al mundo de los negocios digitales [3]. Uno de los



objetivos de una organización es actualizarse y utilizar cada vez más la tecnología en su beneficio, la tecnología permite optimizar sus procesos y transformar sus servicios en actividades que ha hagan más competitiva, también permite llegar a nuevos mercados y audiencias de manera más efectiva al ofrecer servicios externos y canales digitales. Esto puede ayudar a diversificar los ingresos y aumentar la presencia en el mercado. Por ejemplo, Google ofrece una parte de su código para que se incorpore *Google Maps* en una aplicación de terceros puede ser en el ramo de la hotelería o transporte, con esto se proporciona una experiencia de ubicación robusta y confiable para los usuarios, lo que puede mejorar la usabilidad y la funcionalidad de la aplicación, así como aumentar su atractivo [4].

Si bien las APIs inicialmente fueron concebidas para revolucionar la manera cómo los usuarios y las organizaciones mantenían sus relaciones en el ambiente digital. Haciendo más fácil la gestión de las informaciones y acceso a otros negocios. Ahora las APIs van más allá, pueden mejorar increíblemente el negocio ya que permiten gerenciar, analizar estadísticas, proveer informes, optimizar recursos, manejar diferentes áreas, inclusive es posible medir el nivel de satisfacción de usuarios y consumidores, con las APIs es más fácil implementar valor a la experiencia de usuario [5].

### ► III. Proceso de Búsqueda

Como guía para realizar el presente trabajo fue usado el protocolo propuesto por Petersen et al. [6]. Y fueron adaptados los siguientes pasos: (1) Definir una pregunta de investigación, (2) Conducir la búsqueda de estudios primarios, (3) Seleccionar los trabajos basados en los criterios de inclusión/exclusión, (4) Clasificar los trabajos, y por último, (5) Extraer y mapear los datos recolectados.

### ► IV. Plan de la Revisión

La clasificación final obedece a la selección de artículos que conforme al criterio de los investigadores podían proveer información relevante en el entendimiento de APIs como

interfaces fundamentales para la integración de sistemas y el intercambio de información en instituciones financieras, empresas Fintech y otras entidades relacionadas con servicios financieros. En ese sentido, las palabras claves y la lectura de los resúmenes, resultados y conclusiones fueron un criterio importante para la selección de los trabajos.

#### A. *Objetivos y Preguntas de Investigación*

##### *Objetivo Principal*

Proporcionar una visión general, estructurada del estado actual que ilustre como las APIs intervienen en el contexto de Cash Management, específicamente para el área de recaudos, y qué beneficios y retos conlleva su adopción.

##### *Objetivos Específicos*

G1: Identificar la literatura relevante para la integración de APIs en la banca.

G2: Reconocer el contexto actual alrededor de las APIs en la banca y cuáles servicios dentro de Cash Management/Recaudos pueden ser implementados en este contexto.

G3: Determinar los principales beneficios y riesgos en la adopción de APIs.

G4: Reconocer nuevas tecnologías que podrían disminuir los riesgos o aumentar los beneficios en la adopción de APIs.

G5: Reconocer cómo se integran las APIs al modelo de negocio actual de Cash Management/recaudos en la banca.

G6: Identificar los modelos de negocio Fintech que existen alrededor de Cash Management/recaudos.

G7: Determinar los principales aspectos para tener en cuenta en el desarrollo de una API financiera.

G8: Identificar posibles limitantes dentro de las regulaciones a la hora de adoptar APIs en la banca.

Tabla 1  
Preguntas de investigación y motivación derivadas de los objetivos de búsqueda.

Objetivos	Id	Preguntas de investigación	Motivación
G1	RQ1	¿Cuál es el tiempo de distribución de los trabajos?	La distribución del tiempo representará una tendencia macro de la literatura a lo largo del tiempo marcada por los últimos años.
G1	RQ2	¿En dónde se publicaron los trabajos?	Los lugares relevantes, como las conferencias, revistas y sitios especializados que contienen información relacionada con nuestro tema de interés en esta investigación.
G1	RQ3	¿Cuál es la distribución geográfica de los trabajos?	La distribución geográfica presenta las regiones, los países, las universidades y los grupos que están trabajando en el tema y que están logrando una gran contribución a este campo de investigación.
G2	RQ4	¿Cuál es la clasificación de los trabajos seleccionados?	Los trabajos se clasificaron principalmente por la relevancia en el trabajo realizado, principalmente fue analizado el contenido y la inclusión de las palabras claves definidas.
G2	RQ5	¿Qué tipos de APIs se utilizan actualmente en operaciones de Cash Management?	Conocer las APIs que se usan en la actualidad en cash management para comprender el entorno global.
G2	RQ6	¿Qué métodos o procesos se han propuesto para la adopción de APIs en la banca?	Identificar cuáles métodos de adopción de APIs pueden ser útiles para Recaudos Bancolombia.
G2	RQ7	¿Cuáles son los dominios en los cuales son usadas las APIs en las Fintech?	Identificar los dominios en los cuales las APIs podrían ser aplicadas en el Banco. Para Bancolombia es útil explorar este tipo de soluciones y encontrar oportunidades para su negocio.
G3	RQ8	¿Qué oportunidades con APIs fueron identificadas y que puedan ser adoptadas por el Banco?	Identificar oportunidades que permitan potencializar el sistema de Recaudos y así captar mejor el flujo de capital asociado.
G3	RQ9	¿Cuál es el efecto financiero que la banca identifica con el uso de APIs?	Entender el impacto de las APIs en la banca.
G3	RQ10	¿Contribuye la utilización de APIs en el Costo de oportunidad (TCO) en el área de negocio Cash Management?	Medir los beneficios que se hubiesen logrado con la implementación de las APIs de negocio para el área de Cash Management.
G3	RQ11	¿Cuál es el impacto en términos del Time to Market (TTM) cuando se usan APIs de negocio para construir productos en el área de negocio de Cash Management?	Determinar el impacto positivo o negativo de las APIs de negocio sobre la velocidad de salida de los productos de Cash Management.
G4	RQ12	¿Qué nuevas tecnologías podrían fortalecer la implementación de las APIs?	Conocer tecnologías emergentes que podrían afectar favorablemente la implementación de las APIs.
G5	RQ13	¿Qué servicios del negocio de recaudos podrían ser implementados por una API?	Identificar servicios que tienen el potencial de ser implementados a través de APIs para el negocio de recaudos, para posteriormente poder hacer su evaluación desde el punto de vista de costo/beneficio.
G5	RQ14	¿Qué modelos de negocio o de gestión alrededor de las APIs existen en la banca?	Identificar los modelos de negocio o de gestión en torno a las APIs que existen en la banca.
G6	RQ15	¿Cómo integran las APIs las diferentes áreas de la banca, con otros bancos y con las Fintech?	Reconocer el potencial que tienen las APIs como integrador tanto en el área interna como externa de la banca.
G7	RQ16	¿Cuáles son los aspectos de mayor importancia para un banco en el momento de definir y diseñar una API?	Identificar los esquemas de definición de APIs, tanto para la vista del negocio como la vista técnica.
G8	RQ17	¿Cuáles son las principales restricciones que se tienen en cuenta desde el punto de vista legal, gobierno de datos y gobierno TI para la implementación de APIs en el sector Bancario, en específico Cash Management?	Identificar cuáles son las restricciones más comunes con las cuales se enfrenta la banca a la hora de construir APIs de negocio para el área de Cash Management.

Derivado de los objetivos de investigación formulados, se fueron definiendo unas preguntas de investigación y su respectiva motivación con el propósito acotar la búsqueda y los aportes que cada trabajo pudiera realizar al presente estudio, La Tabla I presenta el detalle de este paso del proceso.

### *B. Estrategia de Búsqueda*

Como estrategia para la búsqueda de trabajos relacionados y poder estructurar un estado del arte, se utilizó la siguiente estrategia, así:

- Inicialmente se realizó un levantamiento de información la cual se ha detallado en los trabajos relacionados, y específicamente el proceso previo fue descrito en [24], esa etapa sirvió de contextualización inicial para entender el universo de las APIs.
- La lectura individual de esos documentos sirvió para realizar discusiones grupales, aclarar dudas, talleres de ideación, así como también nuevas iteraciones a los documentos para establecer categorías.
- Los documentos iniciales también sentaron las bases para identificar y seleccionar las palabras claves y las cadenas de búsqueda definidos la búsqueda automática del presente trabajo.
- Síntesis, inferencias y consideraciones.

### *C. Criterios de Inclusión y de Exclusión*

El objetivo general fue analizar y clasificar trabajos que trataban el tema de las APIs, cómo ellas funcionan y cómo sirven de integrador digital, o sea, su utilidad para unir y ofrecer productos de una organización. Adicional a eso, fueron usados como criterio de inclusión la selección de artículos que pertenecieran al enfoque principal, es decir, APIs como punto crítico de aquellas organizaciones que se transformaron digitalmente. Aquellos trabajos que no contenían información relacionada a APIs fueron excluidos de los estudios pertinentes. Luego, los criterios de inclusión se aplicarían a los documentos restantes.

### *Criterios de inclusión*

(i) Trabajos cuyos dominios se encontraban en el ámbito de las APIs, (ii) Trabajos publicados entre el 2010 y el 2023, (iii) Trabajos que presentaban la transformación digital de una organización a través de APIs. (iv) Trabajos que permitían entender la economía de las APIs. (vi) Documentos disponibles en medios digitales.

### *Criterios de exclusión*

(i) Trabajos que no han sido publicados, (ii) Trabajos que no presentaban un enfoque correcto, (iii) Trabajos en los cuales su tratamiento no se refería al contexto de APIs. (iv) Enfoques sin un detalle adecuado, explicación, diagnóstico, tratamiento del tema entre otros.

### *D. Proceso de Selección*

Para seleccionar los documentos relevantes, se siguió un proceso que tuvo tres escaneos: escanear por título, escanear el resumen y escanear texto completo. Para cada exploración se usó los criterios de inclusión y exclusión definidos anteriormente. Se validaron los resultados obtenidos. (i) Etapa 1: Primero se eliminaron todos los documentos duplicados. Luego, se analizaron los documentos por título y palabras clave, con el objetivo de eliminar cualquier documento irrelevante. Cualquier documento considerado para incluirse o en duda debió permanecer en el conjunto de documentos candidatos para la etapa 2. (ii) Fueron examinados los resúmenes de los documentos candidatos de la etapa 1 y fueron evaluados cada uno con los criterios de inclusión/exclusión. Cualquier documento considerado a incluirse o en duda debió permanecer en el conjunto de documentos candidatos para la etapa 3. (iii) Etapa 3: Fueron revisados los textos completos de los documentos candidatos de la etapa 2 y se evaluaron según los criterios de inclusión/exclusión. Cualquier documento considerado a incluirse o excluirse permaneció en el conjunto de trabajos incluidos o en el conjunto de documentos excluidos.

E. Trabajos duplicados, estudios repetidos, y

estudios secundarios

Un documento duplicado, o sea, el que se recuperó de diferentes fuentes de búsqueda (es decir, bibliotecas digitales), para el presente trabajo, fue excluido en la primera etapa del escaneo.

Fueron identificados como estudios repetidos aquellos sobre el mismo estudio que se publicaron en diferentes lugares, generalmente con los mismos autores o la lista de autores con algunas variaciones. Al encontrar este tipo de estudios en el proceso de búsqueda fueron examinados para incluir el más completo o el más reciente.

Los estudios secundarios (es decir, revisiones sistemáticas de la literatura o mapeos sistemáticos de literatura) no fueron encontrados dentro de la búsqueda.

#### F. *Búsqueda, Bibliotecas digitales y motores de búsqueda*

Para la búsqueda automática se realizó a través de motor Descubridor de la Universidad EAFIT y las bibliotecas digitales que él agrupa. La Biblioteca de la Universidad EAFIT ofrece el servicio de esta búsqueda automática inicial (Servicio de Búsqueda de la Universidad EAFIT). Para el uso de este servicio se describe el propósito de la búsqueda, la cadena de búsqueda, el proceso sigue con una entrevista para definir el alcance y resumir lo que se necesita. Con base en estas informaciones el personal de la Biblioteca realizó la ronda con la cadena de búsqueda propuesta, que fue ejecutado en: *ACM Digital Library, British Standards Online, Directorio de Open Access Journals, EBSCO Host, Emerald, IEEE Computer, Society Digital Library, ISI Web of Science, ProQuest, Safari* entre otros. Adicionalmente, se utilizaron el *Repositorio de la CEPAL* y el *Google Scholar*. Finalmente, los documentos obtenidos se almacenaron en formato digital.

#### G. *Cadena de búsqueda*

Teniendo en cuenta que la mayoría de los artículos de avances científico-tecnológicos se encuentran en inglés, las palabras claves se definieron en ese

idioma. Éstas se dividieron en primarias, mínimo una de estas debía estar siempre presente en la búsqueda, y secundarias. Quiere esto decir, en el proceso se definieron unas palabras claves que a la vez fueron clasificadas en primarias por su importancia en el presente estudio y secundarias por el aporte que pudieran realizar. Luego de realizar los respectivos refinamientos se llegó a una cadena de búsqueda específica.

**Palabras claves primarias:** *Cash Management Innovation, Collections Factory, Collections On Behalf Of, Third Party Payment Service Providers, Collections APIs, Enterprise Payments, Payments Hubs, Payment Processing.*

**Palabras claves secundarias:** *Virtual account management, Payment Factory, Fintech APIs, Digital Bank, Digital Economy, API Economy, Payments On Behalf Of, API management, The Banking Industry, Deploy APIs, Open Apis, Bank co-creates APIs, Business APIs, Digital Business models, Value-added Services, Integration of IT and APIs, Financial Services Industry, Financial Services Banks, Legacy Bank Evolution, APIs from other Industries, Technology on Payments, Growing Capabilities of Fintech, Corporate Payments, API Market, Mobile Wallets, Bank as a Service, Banking as a Platform, Fintech-Startups, APIs and E-commerce, Blockchain, Bitcoin, KYC/AML, APIs and Analytics.*

**Cadena de búsqueda genérica:** *"Cash Management Innovation" OR "Collections Factory" OR "Collections On Behalf Of" OR "Third Party Payment Service Providers" OR "Collections APIs" OR "Enterprise Payments" OR "Payments Hubs" OR "Payment Processing" OR "Virtual account management" OR "Payment Factory" OR "Fintech APIs" OR "Digital Bank" OR "Digital Economy" OR "API Economy" OR "Payments On Behalf Of" OR "API management" OR "The Banking Industry" OR "Deploy APIs" OR "Open APIs" OR "Bank co-creates APIs" OR "Business APIs" OR "Digital Business models" OR "Value-added Services" OR "Integration of IT and APIs" OR "Financial Services Industry" OR "Financial Services Banks" OR "Legacy*

*Bank Evolution" OR "APIs from other Industries" OR "Technology on Payments" OR "Growing Capabilities of Fintech" OR "Corporate Payments" OR "API Market" OR "Mobile Wallets" OR "Bank as a Service" OR "Banking as a Platform" OR "Fintech-Startups" OR "APIs and E-commerce" OR "Blockchain" OR "Bitcoin" OR "KYC/AML" OR "APIs and Analytics".*

**Cadena de búsqueda específica:**  
*(TITLE-ABS-KEY("Cash Management Innovation") OR TITLE-ABS-KEY("Collections Factory") OR TITLE-ABS-KEY("Collections On Behalf Of") OR TITLE-ABS-KEY("Third Party Payment Service Providers") OR TITLE-ABS-KEY("Collections APIs") OR TITLE-ABS-KEY("Enterprise Payments") OR TITLE-ABS-KEY("Payments Hubs") OR TITLE-ABS-KEY("Payment Processing") OR TITLE-ABS-KEY("Virtual account management") OR TITLE-ABS-KEY("Payment Factory") OR TITLE-ABS-KEY("Fintech APIs") OR TITLE-ABS-KEY("Digital Bank") OR TITLE-ABS-KEY("Digital Economy") OR TITLE-ABS-KEY("API Economy") OR TITLE-ABS-KEY("Payments On Behalf Of") OR TITLE-ABS-KEY("API management") OR TITLE-ABS-KEY("The Banking Industry") OR TITLE-ABS-KEY("Deploy APIs") OR TITLE-ABS-KEY("Open APIs") OR TITLE-ABS-KEY("Bank co-creates APIs") OR TITLE-ABS-KEY("Business APIs") OR TITLE-ABS-KEY("Digital Business models") OR TITLE-ABS-KEY("Value-added Services") OR TITLE-ABS-KEY("Integration of IT and APIs") OR TITLE-ABS-KEY("Financial Services Industry") OR TITLE-ABS-KEY("Financial Services Banks") OR TITLE-ABS-KEY("Legacy Bank Evolution") OR TITLE-ABS-KEY("APIs from other Industries") OR TITLE-ABS-KEY("Technology on Payments") OR TITLE-ABS-KEY("Growing Capabilities of Fintech") OR TITLE-ABS-KEY("Corporate Payments") OR TITLE-ABS-KEY("API Market") OR TITLE-ABS-KEY("Mobile Wallets") OR TITLE-ABS-KEY("Bank as a Service") OR TITLE-ABS-KEY("Banking as a Platform") OR TITLE-ABS-KEY("Fintech-Startups") OR TITLE-ABS-KEY("APIs and E-commerce") OR TITLE-ABS-KEY("Blockchain") OR TITLE-*

*ABS-KEY("Bitcoin") OR TITLE-ABS-KEY("KYC/AML") OR TITLE-ABS-KEY("APIs and Analytics")) AND PUBYEAR > 2018*

Al aplicar las cadenas de búsqueda, se establecieron criterios de aceptación e inclusión que condujeron a un total de 53 artículos. Estos artículos se encuentran detallados en el siguiente enlace: [<https://n9.cl/946ig>]. Con esta lista como punto de partida, el equipo de trabajo procedió a llevar a cabo una revisión exhaustiva de los títulos de los artículos, con la expectativa de que contuvieran las palabras clave relevantes o estuvieran relacionados con el tema de interés. Además, se realizaron lecturas de los resúmenes para asegurarse de que proporcionaran información pertinente sobre el tema abordado en cada artículo.

Posteriormente, se procedió a la lectura completa de los trabajos seleccionados, y se generó una síntesis de estos que se encuentra disponible en el siguiente enlace: [<https://n9.cl/leqbc>]. Finalmente, se identificó el aporte de cada artículo a la pregunta planteada. Este proceso se puede ver en la Tabla II y la Figura 1, la primera muestra la frecuencia de ocurrencia del tema en el total de los artículos y la segunda muestra de forma gráfica la contribución de estos, cabe anunciar que algunas preguntas todavía no se han respondido y se requiere construir una nueva cadena de búsqueda encaminada a la segunda fase del trabajo y una nueva iteración, que permita, conforme el alcance del proyecto poder mostrar la dimensión completa del tema que se está tratando. En ese sentido, presentamos una sesión de discusión que nos ofrece el panorama inicial sobre este tema.

Tabla II  
 Preguntas de investigación y motivación derivadas de los objetivos de búsqueda.

Pregunta	Artículos por Pregunta	Porcentaje de Artículos
RQ5	3	5.66%
RQ6	7	13.21%
RQ7	10	18.87%
RQ8	15	28.3%
RQ9	6	11.32%
RQ10	4	7.55%
RQ11	4	7.55%
RQ12	13	24.53%
RQ13	1	1.89%
RQ14	11	20.75%
RQ15	3	5.66%
RQ16	9	16.98%
RQ17	7	13.21%



**Fig. 1.** Relación de las preguntas y trabajos que contribuyeron en la respuesta de la pregunta.

## ► V. Discusión

A partir de los trabajos identificados como relevantes para el presente estudio se dice: los servicios financieros están pasando por una dramática transformación, cada vez las economías son más independientes para entregar información digital. Por este motivo los bancos tienen el desafío de ofrecer servicios innovadores y modernos. Es un hecho, globalmente los bancos han disminuido sus canales físicos y se están enfocando a una experiencia digital centrada en el cliente. Como expuesto en [6] estamos en la era de las Fintechs, empresas de tecnología que están expandiendo sus servicios y muestran cómo la sobrevivencia depende del éxito en el ámbito digital, conservando una relación con el surgimiento de nuevas necesidades por parte del usuario, por ejemplo, que su consumo sea sostenible y amigable con el medio ambiente, tal como se menciona en [19], o aquellas que van acorde a la presencia cada vez mayor de nuevas generaciones en el mercado financiero [7]. Adicional a eso, la competencia también surge a través de los bancos de nueva generación, aquellos que nacen 100% digitales, los cuales permiten que sus clientes realicen todas las operaciones exclusivamente a través de internet, y donde como cliente se puede recibir atención vía telefónica, plataformas en línea tipo web o móvil y/o videoconferencias.

Para ilustrar, los bancos ya pueden acceder a la versión de pagos de una tienda de aplicaciones conocidas como tiendas en líneas o *app store* para ofrecer una gama de productos de consumo, tales como administración financiera personal,

presupuestos o pagos móviles, así como una serie de servicios más sofisticados que incluyen seguros, inversiones, o modelos de financiamiento colaborativo en el cual los préstamos son otorgados directamente de persona a persona a través de una plataforma en línea, o hacer alianzas con cualquier Fintech. Se abren muchas posibilidades en el contexto de los pagos, también se puede realizar alianzas bancarias para que los startups de bancos digitales o entidades no bancarias como aquellas de venta de productos directamente a los consumidores finales o por su nombre en inglés *retail* puedan conectarse y ofrecer servicios bancarios a gran escala. Resaltando que dichos pagos utilizan la plataforma de computación en la nube cómo su tecnología base [8].

Las empresas en internet han encontrado oportunidades en el mundo de la banca. Como es señalado en [9] ya en el mercado financiero hay muchos productos y servicios financieros basados en la web, productos y servicios que actualmente los clientes no pueden obtener de su banco o de un proveedor similar. Esto da lugar a un nuevo entorno competitivo. Los proveedores no bancarios, principalmente impulsados por la tecnología, ingresan a los mercados a ofrecer servicios financieros simples. Las diferencias regulatorias, por supuesto, son un factor importante que todavía no se ha explorado lo suficiente.

En [10] se expone el siguiente escenario, clientes que “viven sus vidas a través de sus móviles”, ellos descargan aplicaciones, se registran, hacen sus propios videos, envían fotografías o documentos, o gestionan documentos que contienen imágenes, información de pasaportes y licencias de conducir. Todo gracias a la modernidad de sus teléfonos inteligentes, los datos y la seguridad de las tecnologías que son desarrolladas por *startups* y tecnologías financieras, el sector Fintech. En [11] se afirma algo similar, los autores describen cómo los bancos en Estados Unidos están innovando y colaborando con las nuevas empresas (Fintech) para ofrecer los servicios que los clientes desean. Se vislumbra una nueva ola de tecnologías que podrían dar forma al futuro de la industria. Se describen las tecnologías que impulsan la innovación bancaria en la actualidad, incluso ellos

afirman que la actividad bancaria cambiará como la conocemos ahora en los próximos 10 años, para ilustrar y en orden cronológico ellos exponen, como escenarios disruptivos, los préstamos digitales, la biometría, los datos de los clientes, la inteligencia artificial y el internet de las cosas.

Sin embargo, en [12] se hace el siguiente cuestionamiento, las innovaciones del sector financiero han reducido los costos y mejorado la conveniencia, pero los banqueros y empresas tradicionales no han sido las transformadoras, más bien han sido reacios al riesgo y a pensar en un avance revolucionario. Hasta el momento, la mayoría de los cambios han sido impulsados por *startups* y Fintech que desarrollan soluciones para dispositivos móviles/tabletas, datos biométricos para autenticación de usuarios, tokens de seguridad multibanca, entre otros. Estas empresas proponen nuevas formas de hacer las cosas que otros insisten en continuar haciendo de la misma forma. Esto se da por la falta de liquidez de los bancos globales la cual está reprimiendo la inversión, y hasta que eso no cambie, muchos bancos se centrarán más en el ahorro de costos que en el crecimiento. Las innovaciones digitales se han centrado principalmente en brindar sofisticación de canales en línea y facilidad de uso. Patricia Hines, analista senior de banca corporativa en *Celent* afirma en [12] que la mayoría de la "innovación" en la banca a la fecha, simplemente está enfocada en hacer que un proceso manual sea electrónico en lugar de ser transformador y disruptivo.

No obstante, los bancos deben cambiar porque, como se afirma en [13], lejos están los días cuando los bancos dictaban cuando las personas podían tener acceso a su dinero, actualmente los papeles se han invertido, son las personas las que deciden cómo administran sus finanzas, por este motivo la banca debe convertirse en un "*life partner*" para sus clientes. Esto no sólo significa tener buenas aplicaciones, sino también estar presente en cualquier sitio, aplicaciones o dispositivos que los clientes usen regularmente, como las redes sociales.

Lo anterior se facilita gracias a los beneficios que conlleva el uso de APIs, componentes que le

permiten a los bancos innovar mucho más fácil y rápido. Estas ganancias aumentan al exponer las APIs a otros bancos y a las Fintech, donde estas últimas no están reemplazando a los primeros, sino que sirven a mercados desatendidos [14]. La potencial alianza de los bancos con el sector Fintech garantizaría la continua innovación y satisfacción de nuevas necesidades que surjan en los usuarios, como la sostenibilidad [15] o la inclusión financiera [16], y aumentaría su alcance. Este enfoque también abordaría la preocupación común de que las nuevas tecnologías empleadas en las Fintech y son riesgosas e inseguras [17] debido a su aparente falta de regulación en comparación con los bancos y evitaría o mitigaría la pérdida del dominio de éstos sobre mercado [18].

Un buen ejemplo de esto es el presentado en [19], que consiste en la colaboración entre el banco BBVA (*BBVA Compass*) y la Fintech *Dwolla*, las cuales en conjunto hicieron disponible un grupo de APIs de autenticación y pagos; el banco y *Dwolla* trabajaron juntos para desarrollar una forma de autenticar a los clientes y poner en común sus credenciales de modo que no se pasa información personal identificable a *Dwolla*. Sin embargo, *Dwolla* puede vincular cuentas de BBVA y mover dinero en tiempo real. Lo anterior es un excelente ejemplo de colaboración entre un banco robusto y una Fintech, en donde se brinda un beneficio a los usuarios, el cual a su vez atrae nuevos clientes, ya que se pasa de tener una verificación y transferencia realizada entre tres y seis días, a ser en tiempo real.

En [19] también hablan de cómo algunos consideran a las APIs como la clave para la supervivencia de los bancos y la relevancia en el mundo del "*first-mobile*" o primero el móvil. Lo cual, hace referencia a proceso de diseño que se enfoca en dar prioridad al desarrollo y visualización para dispositivos móviles y no a soluciones de escritorio. Afirmación corroborada por [20], donde se precisa que la clave para la banca del siglo XXI es estar dentro de lo móvil, lo social y en los canales online. Por este motivo, en [19] se concluye que el desarrollo de APIs juegan un rol importante en el cambio de los bancos a la transformación digital. En realidad, como se indica en [21], la

idea de bancos usando plataformas como la Apple App Store para ofrecer servicios de Fintech como parte de su propia suite de productos, es para muchos el camino que está tomando la industria bancaria en este momento. "Si piensas en un iPad, las aplicaciones de Apple son solo algunas de las muchas que podrían estar en él", dice Steve Van Wyk, jefe de tecnología de *PNC Financial Services Group* "¿Por qué no convertirse en el 'iBank' con una tienda de aplicaciones donde las personas pueden realizar operaciones bancarias de muchas maneras diferentes?" [20]. "Los bancos que tienen éxito en el futuro se convertirán en plataformas", menciona Suresh Ramamurthi, presidente y director de tecnología de CBW Bank en Weir, Kan. "Pero no se trata solo de conectar nuevas tecnologías, sino también se trata de transformar toda la cultura y los procesos " [21], manteniendo en la mira el objetivo de proveer al usuario la misma atención por diferentes medios [17]. Estas palabras tuvieron bastante eco en junio del 2016, cuando la CBW publicó más de 500 APIs que proveían oportunidades a otros bancos y desarrolladores, con el fin de presentar un modelo de servicio de "Banking-as-a-Service". Aunque escalar al modelo requirió de un consumo alto de tiempo y dinero, facilitando la asociación de las Fintech con los bancos [21]. La cantidad de APIs que pueda ofrecer un banco es indicador de su esfuerzo por responder de forma proactiva a la inminente digitalización y nuevas regulaciones, y permite que establezca más acuerdos comerciales [18].

Otros dos ejemplos de cómo se busca la integración de las APIs, es el de *Capital One* y *PNC Financial Services Group*, tratados en [22]. En el caso de *Capital One*, se lanzó un portal para desarrolladores en el 2015, llamado *DevExchange*, junto con tres nuevas APIs abiertas. El objetivo es que el portal actuará como base de operaciones para todos los proyectos de fuente abierta del banco. Por su parte, PNC celebró en el 2016 un *API Fest* en el que pidió a todos sus empleados unirse para presentar ideas de innovación orientada al cliente. El evento, que tuvo lugar a principios de abril, contó con la participación de 37 equipos compuestos por 184 empleados. Cada equipo debía tener un desarrollador, pero se alentaba a los

empleados no técnicos a participar.

Sin embargo, no se debe olvidar ser selectivos con el software y los servicios que se comparten con terceros, como se recuerda en [19]. En [22], Van Wyk menciona como PNC es consciente de las preocupaciones de seguridad sobre exponerse con las APIs. Por lo tanto, cualquier aplicación de terceros deberá superar un proceso de revisión exhaustivo para garantizar que nada puede poner en peligro los datos del banco o de los clientes.

Por último, en [23] son analizados el uso de los servicios suministrados por las Fintech y muestra cómo el modelo TAM puede ser útil para analizar el comportamiento y la aceptación de la tecnología por parte de los consumidores, percepción de utilidad y facilidad de uso son algunas de las variables que pueden ser analizadas con esta propuesta.

## » VI. Trabajos Futuros

Se propone la ampliación del estudio para abordar aquellas interrogantes que previamente quedaron sin respuesta debido a la limitada disponibilidad de información y alcance del proyecto. Se enfocaría en la búsqueda de evidencia científica adicional que pueda enriquecer el trabajo, ya que las respuestas a estas cuestiones se consideran de relevancia significativa para obtener una comprensión más completa de la estrategia de adopción de APIs en el sector financiero tradicional y específicamente en recaudos en banca. Concretamente, se pretende explorar cómo estas entidades pueden concebir y ejecutar estrategias efectivas para la integración de APIs en su infraestructura y, al mismo tiempo, establecer colaboraciones eficaces con empresas Fintech y otros actores del ecosistema digital.

También se podría complementar el presente estudio con la evaluación del impacto de APIs en la transformación digital en diferentes sectores además del financiero, como la salud, la educación o la industria manufacturera. Esto proporcionaría una visión más completa de su papel en la economía digital.

Otros temas a abordar podrían ser: desarrollo



de modelos de negocio innovadores con uso de APIs, seguridad y privacidad en entornos *API-Driven*, análisis de tendencias futuras en *API Economy*, El papel de las APIs en la innovación de productos financieros, adopción de APIs en países en desarrollo, y por último, y no menos importante, la gobernanza de APIs y normativas, estas son algunas áreas de investigación que podrían ser exploradas en el contexto de la transformación digital impulsada por las APIs en empresas digitales emergentes y organizaciones tradicionales. Cada una de estas ideas podría llevar a investigaciones valiosas que contribuyan a la comprensión y mejora de esta dinámica en constante evolución.

Al realizar este tipo de trabajo nos encontramos con posibles sesgos de idioma, sesgos de acceso a la literatura, sesgos de publicación y sesgos en la búsqueda.

Se identificaron las limitaciones en la cantidad de artículos incluidos, la evolución del campo y la posibilidad de no tratar temas emergentes.

### » VII. Conclusiones

En síntesis, el impacto transformador de las empresas digitales emergentes en las estructuras organizativas es innegable en el actual panorama de los mercados modernos. Para una comprensión integral de este ciclo evolutivo, el análisis de APIs se presenta como esencial, ya que son la fuerza motriz detrás de modelos de negocio innovadores. En el sector financiero, la activa adopción de APIs por parte de las Fintech subraya su papel central en el fomento de operaciones digitales.

No obstante, en el contexto de negocios de banca de gran envergadura se enfrentan a desafíos significativos al intentar generar valor en asociación con estos socios digitales o al buscar adquirir capacidades innovadoras por sí mismas a través de la implementación estratégica de APIs. Este estudio se propuso analizar el entorno actual de las APIs en el ámbito bancario. A través de un mapeo sistemático de la literatura, se logró sintetizar y discutir cómo las APIs están impulsando la transformación digital en el sector

bancario.

La premisa clave que se destaca en este análisis es el valor fundamental que las APIs brindan a las instituciones bancarias. Estas interfaces no solo habilitan la colaboración y la innovación en un ecosistema digital en constante cambio, sino que también empoderan a consumidores, desarrolladores de terceros, Fintech y otros actores al permitirles redescubrir y reconfigurar componentes críticos como algoritmos, datos y procesos. Esta capacidad de rediseño es esencial para definir las reglas del juego que permiten a las organizaciones competir en un entorno digital y expandir sus horizontes.

### » VIII. Agradecimientos

Este artículo es resultado del proyecto de investigación “Esquematizar un modelo de integración dinámico basado APIs de negocio que potencialice la renovación del CORE bancario de recaudos en su interacción con los clientes a través de diferentes canales digitales”, el cual ha sido apoyado por la Universidad EAFIT y financiado por un banco de la Ciudad de Medellín, Colombia. Los profesores Ph.D. Elizabeth Suescún, Ph.D. Marta Tabares y Ph.D. Liliana González agradecen el aporte de la Universidad, donde se desempeñan como Profesores Titulares. Asimismo, los autores desean expresar su sincero agradecimiento Nelson Moreno, Catalina Patiño y Alex Montoya por su valiosa contribución y apoyo en la realización de este estudio.

### » IX. Referencias

- [1] Kai Petersen, Robert Feldt, Shahid Mujtaba, Michael Mattson. (2008). Systematic Mapping Studies in Software Engineering. Systematic MapProceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, 17.
- [2] Google Code. (s/f). Recuperado el 10 de septiembre de 2023, de Google.com website: <https://code.google.com/>
- [3] Qué es una API y qué puede hacer por mi negocio. (2016, junio 7). Recuperado el

- 10 de septiembre de 2023, de BBVA API\_Market website: <https://bbvaopen4u.com/es/actualidad/que-es-una-api-y-que-puede-hacer-por-mi-negocio>
- [4] Google Maps Platform. (s/f). Recuperado el 10 de septiembre de 2023, de Google for Developers website: <https://developers.google.com/maps/>
- [5] Vértigo. (s/f). Recuperado el 10 de septiembre de 2023, de Vértigo website: <http://vertigo.com.br/tag/api-banco-do-brasil/>
- [6] Bruggink, M. L. A. (2016). How FinTech is transforming the way money moves around the world: An interview with Mike Laven. *ournal of Payments Strategy & Systems*, 10(1), 6–12.
- [7] Varma, P., Nijjer, S., Sood, K., Grima, S., & Rupeika-Apoga, R. (2022). Thematic analysis of financial technology (Fintech) influence on the banking industry. *Risks*, 10(10), 186. doi:10.3390/risks10100186
- [8] Yurcan, B. (2016b, agosto 3). Fintech vet aims to make banking as easy as an app download. Recuperado el 10 de septiembre de 2023, de American Banker website: <https://www.americanbanker.com/news/fintech-vet-aims-to-make-banking-as-easy-as-an-app-download>
- [9] Dapp, T. F. (2014). Fintech – The digital (r)evolution in the financial sector. Deutsche Bank Research. Recuperado de <https://www.dbresearch.com>
- [10] Flinders, K. (s/f). Starling poised for growth as customers flock to the app-based challenger bank: Thousands of people a week are installing the Starling app as the bank builds its customer base. *Computer Weekly*, 7–9.
- [11] Morgan, R. (2017, septiembre 5). The top fintech trends driving the next decade. Recuperado el 10 de septiembre de 2023, de ABA Banking Journal website: <https://bankingjournal.aba.com/2017/09/the-top-fintech-trends-driving-the-next-decade/>
- [12] Wright, G. (2016, junio 10). Building toward A breakthrough. Recuperado el 10 de septiembre de 2023, de Global Finance Magazine website: <https://www.gfmag.com/magazine/june-2016/building-toward-breakthrough>
- [13] Crosman, P. (2017, septiembre 29). How Citi aims to become a ‘life partner’ for customers. Recuperado el 10 de septiembre de 2023, de American Banker website: <https://www.americanbanker.com/news/how-citi-aims-to-become-a-life-partner-for-customers>
- [14] Skinner, C. (2017, julio 27). The battle between banks and disruptors is just beginning. Recuperado el 10 de septiembre de 2023, de American Banker website: <https://www.americanbanker.com/opinion/the-battle-between-banks-and-disruptors-is-just-beginning>
- [15] Hopalı, E., Vayvay, Ö., Kalender, Z. T., Turhan, D., & Aysuna, C. (2022). How do mobile wallets improve sustainability in payment services? A comprehensive literature review. *Sustainability*, 14(24), 16541. doi:10.3390/su142416541
- [16] Lagna, A., & Ravishankar, M. N. (2022). Making the world a better place with fintech research. *Information Systems Journal*, 32(1), 61–102. doi:10.1111/isj.12333
- [17] Rosli, M. S., Saleh, N. S., Md. Ali, A., & Abu Bakar, S. (2023). Factors determining the acceptance of E-wallet among Gen Z from the lens of the extended Technology Acceptance Model. *Sustainability*, 15(7), 5752. doi:10.3390/su15075752
- [18] Stefanelli, V., & Manta, F. (2023). Digital financial services and open banking innovation: Are banks becoming ‘invisible’? *Global Business Review*, 097215092311514. doi:10.1177/09721509231151491
- [19] Crosman, P. (2015, julio 8). Fintech Glasnost: Why U.S. Banks Are Opening Up APIs to Outsiders. Recuperado el 10 de septiembre de 2023, de Americanbanker.com website: <https://www.americanbanker.com/news/fintech-glasnost-why-us-banks-are-opening-up-apis-to-outsiders>
- [20] Reosti, J., Kline, A., Crosman, P., Sanneh, E. S., & Adams, J. (2015, enero 7).

- Creating a Bank that's Both Cool and Fair. Recuperado el 10 de septiembre de 2023, de American Banker website: <https://www.americanbanker.com/opinion/creating-a-bank-thats-both-cool-and-fair>
- [21] Yurcan, B. (2017, mayo 19). Can vendors bridge gap between legacy tech and open banking? Recuperado el 10 de septiembre de 2023, de American Banker website: <https://www.americanbanker.com/news/can-vendors-bridge-gap-between-legacy-tech-and-open-banking>
- [22] Yurcan, B. (2016a, abril 29). How will APIs change banking? PNC looks within for answers. Recuperado el 10 de septiembre de 2023, de American Banker website: <https://www.americanbanker.com/news/how-will-apis-change-banking-pnc-looks-within-for-answers>
- [23] Li-Min Chuang, C.-C. L. & H.-K. K. (2016). The Adoption of Fintech Service: TAM perspective. *International Journal of Management and Administrative Sciences (IJMAS)*, 3(07), 01–15.
- [24] Tabares, M. S., & Suescun, E. (2020). Towards an APIs Adoption Agile Model in Large Banks. En *Trends and Innovations in Information Systems and Technologies* (pp. 302–311). Cham: Springer International Publishing.





## REVISTA PERSPECTIVAS

REVISTA TÉCNICA CIENTÍFICA DE LA FIE

---

- ELECTROMAGNETISMO Y ÓPTICA APLICADA.
  - AUTOMATIZACIÓN Y CONTROL.
  - SOFTWARE Y APLICACIONES.
- REDES Y COMUNICACIÓN DE DATOS.
  - SISTEMAS ELECTRÓNICOS.
  - INFORMÁTICA EDUCATIVA.
    - SISTEMAS DIGITALES.
  - TELECOMUNICACIONES.
    - DISEÑO GRÁFICO.
    - TIC`S.



Facultad de  
Informática y  
Electrónica

<http://perspectivas.esPOCH.edu.ec/>