Interserrana:
Ciencia, Cultura y Territorio
Rev. Interserrana Cienc. Cult. Territ. · Vol. 1, Núm. 1 · 2026

ARTÍCULO ORIGINAL · CIENCIA, TECNOLOGÍA E INNOVACIÓN TERRITORIAL

Repositorio institucional inteligente aplicado a comunidades rurales

Iván Vicente-Anaya[1], Edison Milton Balvín-Sánchez[1*]

[1] Licenciatura en Ingeniería en Tecnología Informática, Universidad Interserrana del Estado de Puebla-Ahuacatlán (UIEPA), Puebla, México

* Autor(a) por correspondencia: edison.balvin@uiepa.edu.mx · ORCID: 0009-0006-5307-6342

RESUMEN

Las instituciones de educación superior generan una cantidad importante de conocimiento académico mediante tesis y proyectos de investigación; sin embargo, en contextos universitarios rurales parte de esta información permanece almacenada en formatos físicos o sistemas con capacidades limitadas de búsqueda, reduciendo su accesibilidad y aprovechamiento. En este trabajo se desarrolló e implementó Neurorepo, un repositorio institucional inteligente orientado a la preservación, organización y recuperación de tesis académicas mediante herramientas de Inteligencia Artificial Generativa. Durante el desarrollo del sistema se identificó que uno de los principales retos fue adaptar las capacidades de los modelos generativos a documentos académicos con diferentes estructuras y formatos de digitalización. Para desarrollar Neurorepo se trabajó mediante ciclos de prototipado, donde cada versión permitió probar funciones, corregir errores e incorporar mejoras antes de la evaluación final. Neurorepo fue construido mediante una arquitectura Modelo-Vista-Controlador (MVC) utilizando PHP y MySQL, integrando técnicas de Procesamiento de Lenguaje Natural para la extracción automática de metadatos desde documentos PDF y un asistente conversacional para facilitar la consulta de información académica. La evaluación incluyó a 20 participantes relacionados con actividades de investigación, revisión y gestión documental. Sus respuestas permitieron conocer la experiencia de uso de Neurorepo dentro del contexto institucional donde fue implementado, mostrando una aceptación favorable del sistema, obteniendo un 80 % de valoraciones positivas en la extracción automática de metadatos, 90 % en el desempeño del chatbot académico y una media general de 4.5 sobre 5 en la evaluación de funcionalidad y utilidad. Los resultados obtenidos permitieron observar que Neurorepo contribuyó a mejorar la organización de tesis académicas y facilitó la recuperación de información dentro de la institución evaluada. Se concluye que Neurorepo constituye una solución adaptable para apoyar la preservación y recuperación de información científica generada en comunidades universitarias rurales.

Palabras clave: *repositorio institucional; inteligencia artificial generativa; procesamiento de lenguaje natural; preservación digital; gestión documental*

ABSTRACT

Higher education institutions generate a significant amount of academic knowledge through theses and research projects; however, in rural university contexts, part of this information remains stored in physical formats or systems with limited search capabilities, reducing its accessibility and use. In this work, Neurorepo was developed and implemented — an intelligent institutional repository aimed at the preservation, organization, and retrieval of academic theses using Generative Artificial Intelligence tools. During the development of the system, one of the main challenges identified was adapting the capabilities of generative models to academic documents with varying structures and digitization formats. Neurorepo was developed through prototyping cycles, where each version allowed functions to be tested, errors to be corrected, and improvements to be incorporated before the final evaluation. Neurorepo was built using a Model-View-Controller (MVC) architecture with PHP and MySQL, integrating Natural Language Processing techniques for the automatic extraction of metadata from PDF documents, and a conversational assistant to facilitate the retrieval of academic information. The evaluation conducted with 20 participants involved in research, review, and document management processes made it possible to analyze real users' responses to the platform, showing favorable acceptance of the system — achieving 80% positive ratings for automatic metadata extraction, 90% for the academic chatbot's performance, and an overall mean score of 4.5 out of 5 in the assessment of functionality and usability. The evaluation revealed that Neurorepo contributed to improving the organization of academic theses and facilitated information retrieval within the evaluated institution. It is concluded that Neurorepo constitutes an adaptable solution to support the preservation and retrieval of scientific information generated in rural university communities.

Keywords: *institutional repository; generative artificial intelligence; natural language processing; digital preservation; document management*

Cómo citar: Vicente-Anaya, I. & Balvín-Sánchez, E. M. (2026). Repositorio institucional inteligente aplicado a comunidades rurales. *Interserrana: Ciencia, Cultura y Territorio, 1*(1), 25–34.

Recibido: mayo 2026 · **Aceptado:** junio 2026 · **Publicado:** 2026

1. INTRODUCCIÓN

Cada año, las instituciones de educación superior generan tesis y proyectos que contienen conocimiento valioso; conservar este material y facilitar su consulta continúa siendo un desafío. Su impacto depende no solamente de su creación, sino también de la capacidad institucional para organizarlo y ponerlo a disposición de nuevas generaciones de estudiantes e investigadores. Uno de los principales desafíos de las universidades es evitar que tesis, proyectos y documentos académicos permanezcan únicamente como archivos almacenados. Para atender esta necesidad se requieren estrategias que faciliten la preservación, consulta y aprovechamiento de estos documentos dentro de la comunidad académica.

Los repositorios institucionales surgieron como una respuesta tecnológica ante esta necesidad, al proporcionar espacios digitales destinados a preservar, clasificar y difundir la producción científica generada dentro de las universidades, favoreciendo la construcción de infraestructuras académicas digitales y el acceso abierto al conocimiento científico. Estas plataformas han permitido fortalecer el acceso abierto al conocimiento y mejorar los procesos de recuperación documental mediante esquemas organizados de almacenamiento de información académica (Borgman, 2010; Suber, 2012; Umaña-Alpízar, 2015; Voutssás-Márquez, 2014).

A pesar de estos avances, la adopción de soluciones digitales no ocurre bajo las mismas condiciones en todos los contextos educativos. En instituciones ubicadas en regiones rurales o con infraestructura tecnológica limitada, todavía es frecuente encontrar procesos documentales basados principalmente en archivos físicos, registros manuales y mecanismos tradicionales de búsqueda. Esta situación reduce la visibilidad del conocimiento generado localmente y dificulta que estudiantes, docentes e investigadores puedan utilizar trabajos previos como base para nuevas investigaciones. Estudios desarrollados en instituciones de la Sierra Norte de Puebla han evidenciado que las condiciones tecnológicas y contextuales pueden influir en los procesos académicos y en el acceso a recursos digitales (Balvín-Sánchez et al., 2026; Rodríguez-Abitia & Correa, 2021).

Esta problemática se presenta en la Universidad Interserrana del Estado de Puebla Ahuacatlán, donde durante varios años se ha generado un acervo importante de tesis pertenecientes a diferentes programas académicos. Aunque estos documentos forman parte de la memoria científica institucional, una proporción considerable permanece disponible únicamente mediante consulta física, lo que limita su alcance y aumenta el riesgo de pérdida, deterioro o subutilización de la información.

Frente a este escenario, los avances recientes en Inteligencia Artificial Generativa (IAG) y Procesamiento de Lenguaje Natural (PLN) ofrecen nuevas posibilidades para transformar los repositorios tradicionales en sistemas capaces de analizar y organizar información de manera automatizada. Estas tecnologías permiten realizar tareas como extracción de metadatos, identificación de contenido relevante y recuperación eficiente de información, procesos fundamentales dentro de los sistemas de búsqueda documental y organización de recursos digitales (Weibel et al., 1998; Manning et al., 2008). Estas capacidades adquieren importancia debido al aumento constante de información digital, donde los métodos tradicionales de revisión y clasificación pueden ser insuficientes para manejar grandes volúmenes de documentos (Gandomi & Haider, 2015).

La aplicación de estas tecnologías depende del contexto donde serán implementadas. En universidades con recursos tecnológicos limitados, no basta con incorporar herramientas inteligentes; también deben analizarse las condiciones técnicas del lugar donde serán implementadas, incluyendo infraestructura disponible, mantenimiento y recursos institucionales.

A partir del diagnóstico realizado en la institución, se desarrolló Neurorepo, un repositorio institucional inteligente orientado a la gestión digital de tesis académicas en la Universidad Interserrana del Estado de Puebla Ahuacatlán. Su diseño integra una arquitectura web basada en PHP y MySQL con herramientas de Inteligencia Artificial Generativa orientadas a la extracción automática de metadatos, asistencia académica mediante chatbot y recuperación eficiente de información, buscando mejorar la preservación, organización y acceso al conocimiento generado en contextos universitarios rurales.

2. METODOLOGÍA

El estudio se planteó desde un enfoque experimental-computacional porque fue necesario desarrollar el sistema y, posteriormente, evaluar su funcionamiento en un escenario real. Este enfoque permitió analizar el funcionamiento técnico del sistema y evaluar su interacción con los usuarios dentro del entorno donde sería utilizado.

Para el cumplimiento de los objetivos planteados, se diseñó una metodología propia de carácter iterativo y prototipado. Durante el desarrollo se generaron diferentes versiones del sistema. Cada etapa permitió revisar el funcionamiento de los módulos, detectar errores y realizar ajustes antes de avanzar a una nueva versión. Este esquema permitió ajustar el desarrollo conforme aparecieron nuevos requerimientos y necesidades detectadas durante la construcción del sistema.

Los prototipos funcionales ayudaron a revisar desde etapas tempranas la arquitectura MVC, la experiencia de usuario y la integración de herramientas de IAG antes de la versión final del sistema. Esta forma de trabajo permitió realizar modificaciones conforme avanzaba el desarrollo, incorporando observaciones obtenidas durante las pruebas del sistema.

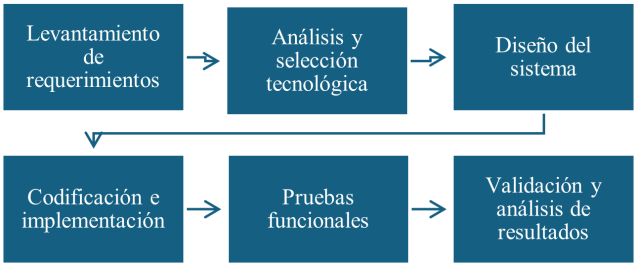

Figura 1
Metodología general del proyecto.
Nota. Elaboración propia (2026).

2.1 Diseño de la muestra y sujetos de estudio

La validación funcional del sistema se realizó con una muestra no probabilística por conveniencia, seleccionando participantes relacionados directamente con procesos de gestión de documentos académicos. Este tipo de muestreo es común en investigaciones aplicadas donde se requiere la participación de sujetos con perfiles específicos relacionados con el fenómeno de estudio (Hernández-Sampieri & Mendoza, 2018).

La población estuvo conformada por 20 participantes pertenecientes a la Universidad Interserrana del Estado de Puebla Ahuacatlán, integrados por docentes con grado de Doctorado, profesores investigadores y personal académico vinculado a procesos de revisión y titulación. La participación de estos usuarios permitió evaluar el sistema con personas que realizan actividades reales de consulta, revisión y gestión de documentos académicos dentro de la institución.

Tabla 1
Nivel académico de los participantes

Nivel académico	Frecuencia
Doctorado	8
Maestría	7
Licenciatura	5
Total	20

Nota. Elaboración propia (2026).

Tabla 2
Rol institucional de los participantes

Rol institucional principal	Frecuencia
Docentes	9
Investigadores	6
Revisores o asesores de tesis	5
Total	20

Nota. Elaboración propia (2026).

Los expertos fueron seleccionados porque su experiencia permitía valorar si el repositorio respondía a los problemas de consulta, organización y preservación documental identificados en la institución. Investigaciones previas señalan que las universidades localizadas en regiones rurales enfrentan mayores limitaciones en infraestructura tecnológica y acceso a plataformas digitales especializadas (Rodríguez-Abitia & Correa, 2021).

2.2 Instrumento de recolección de datos

Para el levantamiento de información se utilizó la plataforma Google Forms, debido a que facilitaba la aplicación del instrumento y permitía reunir las respuestas de los participantes en un solo espacio para su posterior análisis. El instrumento fue diseñado bajo una estructura mixta compuesta por preguntas de opción múltiple y reactivos basados en escala de Likert.

Se utilizó una escala de Likert debido a que permite cuantificar la percepción de los usuarios respecto a la usabilidad, satisfacción y desempeño tecnológico del sistema. Este tipo de escala es ampliamente utilizada en investigaciones relacionadas con evaluación de sistemas interactivos y análisis de experiencia de usuario.

El instrumento estuvo integrado por 15 reactivos distribuidos de la siguiente manera:

Cinco preguntas de opción múltiple orientadas a identificar perfiles de usuario, hábitos de consulta y dispositivos utilizados.

Diez reactivos basados en una escala de Likert de cinco niveles: Malo, Regular, Bueno, Muy Bueno y Excelente.

Los reactivos permitieron evaluar aspectos relacionados con: precisión de la extracción automatizada de metadatos, claridad de las respuestas generadas por el chatbot, facilidad de navegación, accesibilidad de la plataforma, eficiencia de búsqueda documental, desempeño general del sistema, y utilidad del repositorio institucional en procesos académicos.

El análisis de los datos se realizó mediante estadística descriptiva, utilizando frecuencias y valores promedio para interpretar la percepción de los usuarios sobre el sistema. Este análisis fue seleccionado porque permitió interpretar la aceptación inicial del sistema durante la evaluación realizada (Hernández-Sampieri & Mendoza, 2018).

2.2.1 Levantamiento de requerimientos

La primera fase estuvo orientada a identificar las necesidades institucionales relacionadas con preservación, organización y consulta de tesis académicas dentro de la universidad. El objetivo principal consistió en transformar problemáticas administrativas y documentales en requerimientos funcionales y técnicos para el desarrollo del repositorio inteligente.

Para ello se emplearon distintas técnicas de recopilación de información:

Lluvia de ideas, utilizada para identificar funciones prioritarias del sistema y analizar mecanismos de automatización mediante Inteligencia Artificial.

Entrevistas semiestructuradas, aplicadas a coordinadores académicos y personal administrativo para comprender el flujo de trabajo relacionado con recepción, almacenamiento y consulta de tesis.

Encuestas de necesidades, dirigidas a docentes y estudiantes para detectar dificultades asociadas con búsqueda de información académica y acceso documental.

Análisis documental, realizado sobre el acervo físico y digital existente para identificar estructuras comunes en los archivos PDF y determinar los metadatos requeridos para la clasificación de documentos.

A partir de estas técnicas se identificaron problemas relacionados con duplicidad de registros, pérdida de trazabilidad documental y dificultades en la consulta de antecedentes académicos. Estas dificultades reflejan la importancia de fortalecer los procesos de gestión documental universitaria mediante estrategias que permitan organizar y preservar adecuadamente la información institucional (Quindemil et al., 2021). Investigaciones previas han señalado que la falta de digitalización y organización documental puede afectar significativamente los procesos de preservación y acceso abierto en instituciones de educación superior (Umaña-Alpízar, 2015; Voutssás-Márquez, 2014).

Como resultado de esta fase se elaboró un documento de especificación de requerimientos funcionales y técnicos, definiendo roles de usuario, niveles de acceso, mecanismos de seguridad y funciones automatizadas basadas en Inteligencia Artificial.

2.2.2 Análisis de requerimientos y selección tecnológica

Una vez definidos los requerimientos funcionales, se realizó un análisis técnico orientado a seleccionar herramientas y tecnologías que ofrecieran estabilidad y fueran compatibles con la infraestructura disponible en la institución.

La elección de PHP y MySQL respondió a las condiciones reales de implementación identificadas durante el análisis inicial del proyecto, donde se buscó priorizar compatibilidad con la infraestructura institucional existente, facilidad de mantenimiento y bajo consumo de recursos computacionales. Para la organización interna del sistema se adoptó el patrón Modelo-Vista-Controlador (MVC), el cual facilita la separación entre lógica de negocio, administración de datos e interfaz de usuario, favoreciendo la modularidad y mantenimiento del software. Este tipo de patrones arquitectónicos permite organizar sistemas complejos mediante estructuras reutilizables y mantenibles durante el ciclo de vida del software (Buschmann et al., 1996).

Para las funciones inteligentes se integró la API Gemini 1.5 Flash. Su selección respondió a la necesidad de incorporar un modelo capaz de procesar documentos académicos extensos y generar respuestas contextuales manteniendo tiempos de respuesta adecuados para una plataforma institucional. La incorporación de modelos de lenguaje de gran escala ha demostrado ser útil en tareas relacionadas con extracción automática de información, clasificación documental y generación de respuestas contextuales (Brown et al., 2020).

Como parte del entorno de desarrollo, se incorporaron herramientas complementarias como Composer para la gestión de dependencias y PHPMailer para el envío automatizado de notificaciones institucionales. Además, se implementaron mecanismos de seguridad mediante archivos .htaccess y variables de entorno para proteger credenciales y restringir accesos no autorizados.

2.2.3 Diseño del sistema

En esta fase se definió la estructura lógica, funcional y visual del repositorio institucional. El diseño contempló tres componentes principales: (a) arquitectura de software basada en MVC, (b) diseño de base de datos relacional, y (c) diseño de interfaz centrada en experiencia de usuario (UX/UI).

La base de datos fue estructurada bajo el motor InnoDB de MySQL, permitiendo mantener integridad referencial mediante relaciones entre tablas correspondientes a usuarios, tesis, carreras, asesores y registros de actividad. La estructura también incorporó mecanismos de trazabilidad y auditoría para registrar las operaciones realizadas dentro del sistema.

Respecto al diseño visual, se utilizó Bootstrap para desarrollar una interfaz responsiva adaptable a dispositivos móviles y computadoras de escritorio. El diseño consideró principios centrados en el usuario, debido a que la facilidad de interacción influye directamente en la aceptación y uso efectivo de los sistemas tecnológicos (Norman, 2013). Esta característica fue considerada durante el desarrollo de Neurorepo, ya que en contextos con limitaciones de conectividad la accesibilidad de las plataformas digitales influye directamente en su adopción y uso (Rodríguez-Abitia & Correa, 2021).

2.2.4 Codificación e implementación

En esta etapa se desarrollaron los módulos funcionales del sistema utilizando PHP nativo, JavaScript y MySQL. La implementación incluyó procesos de autenticación, administración de usuarios, carga documental, visualización de tesis y automatización inteligente mediante IA generativa.

Uno de los componentes principales fue el desarrollo del módulo de extracción automatizada de metadatos. Durante las primeras pruebas del sistema se identificó que la captura manual de información representaba uno de los principales puntos de demora en la organización documental; por ello, el módulo fue orientado a automatizar la extracción de títulos, autores, resúmenes y palabras clave mediante técnicas de Procesamiento de Lenguaje Natural (PLN). La literatura especializada muestra que las técnicas de minería de texto y PLN permiten optimizar tareas relacionadas con análisis, clasificación y recuperación documental automatizada (Liu, 2012; Miner et al., 2012; Boukhers & Bouabdallah, 2022).

A partir de las necesidades observadas durante el diseño del repositorio, se incorporó un chatbot asistencial como un mecanismo de apoyo para reducir la dependencia de búsquedas manuales y facilitar la interacción del usuario con los documentos almacenados. La comunicación con el modelo de IA se realizó mediante endpoints dinámicos y respuestas estructuradas en formato JSON.

Para reforzar la seguridad del sistema, se implementaron mecanismos orientados a restringir accesos no autorizados, proteger la integridad de los documentos y reducir vulnerabilidades relacionadas con inyección SQL y exposición de credenciales, siguiendo recomendaciones generales de desarrollo seguro en aplicaciones web (Pressman & Maxim, 2020).

3. RESULTADOS

Con Neurorepo se evaluó la aplicación de herramientas de IAG en procesos reales de organización y recuperación documental dentro de la universidad analizada. Durante las pruebas, el sistema funcionó de manera estable y los participantes mostraron una valoración positiva de las herramientas implementadas.

Durante las pruebas funcionales se verificó el correcto desempeño de los módulos relacionados con autenticación de usuarios, almacenamiento de tesis, visualización documental, generación automatizada de metadatos y asistencia conversacional mediante IA. Las pruebas también permitieron comprobar la compatibilidad del sistema con distintos dispositivos y navegadores, manteniendo estabilidad en los procesos de carga, consulta y administración de información académica sin registrar fallos críticos durante la evaluación.

Las respuestas obtenidas por parte de los expertos académicos mostraron una buena aceptación de las funciones automatizadas implementadas en Neurorepo. Particularmente, la identificación automática de títulos académicos mediante Inteligencia Artificial presentó predominancia de respuestas en las categorías “Excelente” y “Muy Bueno”, alcanzando conjuntamente el 80 % de las valoraciones obtenidas. La evaluación refleja una aceptación positiva del sistema para apoyar tareas de extracción documental automatizada.

Tabla 3
Percepción sobre la precisión de extracción automática de metadatos

Categoría	Frecuencia	Porcentaje
Excelente	10	50 %
Muy Bueno	6	30 %
Bueno	4	20 %
Regular	0	0 %
Total	20	100 %

Nota. Datos obtenidos mediante el instrumento de evaluación aplicado a los participantes.

Durante la evaluación se observó que uno de los principales beneficios percibidos por los usuarios fue la reducción del trabajo manual asociado con la búsqueda, clasificación y consulta de tesis académicas mediante la generación automática de resúmenes y metadatos. Este comportamiento coincide con investigaciones recientes relacionadas con Procesamiento de Lenguaje Natural y automatización documental, donde se destaca que los modelos de lenguaje pueden optimizar procesos administrativos y mejorar la recuperación de información académica (Kowsari et al., 2019; Brown et al., 2020).

Respecto al chatbot institucional, los participantes indicaron que las respuestas generadas mostraron niveles adecuados de claridad, coherencia y utilidad durante los procesos de consulta. Este resultado coincide con el desarrollo reciente de agentes conversacionales, los cuales buscan facilitar la interacción humano-computadora mediante interfaces basadas en lenguaje natural (McTear, 2020). La evaluación de este módulo mostró nuevamente predominancia de categorías positivas, donde el 90 % de los participantes calificó el desempeño general entre “Excelente” y “Muy Bueno”, lo que refleja una buena recepción de los usuarios hacia el uso de asistentes conversacionales dentro de plataformas universitarias.

Tabla 4
Evaluación del desempeño del chatbot institucional

Categoría	Frecuencia	Porcentaje
Excelente	12	60 %
Muy Bueno	6	30 %
Bueno	2	10 %
Regular	0	0 %
Total	20	100 %

Nota. Resultados obtenidos durante la evaluación funcional del sistema.

Los participantes valoraron positivamente el uso del repositorio dentro del entorno institucional evaluado. Además, se consideró que la plataforma contribuye al fortalecimiento del acceso digital, la preservación académica y la organización documental dentro de la universidad. Este punto es especialmente importante en instituciones rurales, donde la digitalización documental y el acceso a herramientas especializadas todavía representan retos cotidianos (Rodríguez-Abitia & Correa, 2021).

Durante las pruebas técnicas se observó que el sistema conservó funciones esenciales de almacenamiento y consulta documental aún bajo condiciones limitadas de conectividad. Esto indica que la arquitectura utilizada logró mantener las funciones principales del repositorio aun cuando existieron limitaciones técnicas durante las pruebas.

El análisis descriptivo de las respuestas permitió identificar una valoración global favorable del sistema. La media general obtenida en los reactivos tipo Likert fue de 4.5 sobre 5, lo que refleja una valoración positiva de los usuarios sobre la funcionalidad, accesibilidad y utilidad académica del repositorio institucional.

Tabla 5
Estadísticos descriptivos de evaluación del sistema

Variable	Media	DE
Precisión de extracción automática	4.6	0.52
Claridad del chatbot	4.4	0.61
Facilidad de navegación	4.5	0.49
Utilidad académica	4.7	0.44

Nota. Elaboración propia (2026).

La evaluación también permitió identificar áreas que pueden fortalecerse, principalmente en los tiempos de respuesta y en algunos procesos automatizados del sistema. Aunque todavía existen aspectos por mejorar, los participantes evaluaron positivamente el funcionamiento general del repositorio, mostrando que este tipo de sistemas puede aplicarse en instituciones educativas rurales.

4. DISCUSIÓN

Los resultados obtenidos con Neurorepo muestran que la inteligencia artificial puede incorporarse a procesos reales de gestión documental universitaria, especialmente cuando existe una necesidad concreta de organizar grandes cantidades de información académica. El propósito del sistema no se limitó a digitalizar archivos existentes, sino a explorar una forma diferente de organizar y consultar la producción académica generada dentro de la universidad mediante técnicas de Inteligencia Artificial Generativa.

Durante la evaluación destacó principalmente el comportamiento del módulo de extracción automática de metadatos, donde el 80 % de los participantes evaluó su funcionamiento dentro de las categorías “Excelente” y “Muy Bueno”. Este resultado fue consistente con la problemática observada al inicio del proyecto, donde el registro y organización de tesis requería una intervención manual considerable. La incorporación de técnicas de Procesamiento de Lenguaje Natural redujo parte de estas actividades y facilitó la clasificación inicial de los documentos almacenados.

Esto permitió que el sistema atendiera una necesidad detectada durante el diagnóstico inicial del proyecto: reducir el tiempo requerido para registrar y clasificar documentos académicos de forma manual.

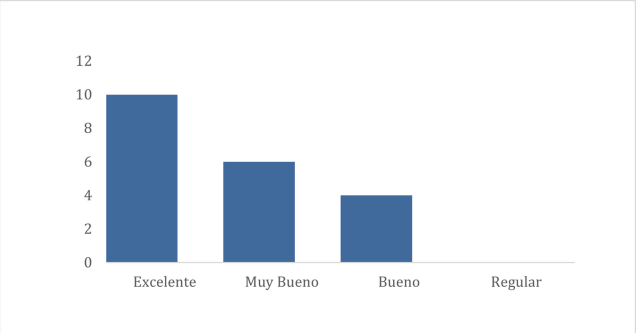

Figura 2
Precisión percibida en la extracción automática de metadatos.
Nota. Elaboración propia con datos obtenidos mediante Google Forms (2026).

Una explicación de este comportamiento puede relacionarse con lo reportado por Kowsari et al. (2019), quienes describen que las técnicas de PLN han mejorado los procesos de clasificación textual mediante la identificación automática de patrones en grandes conjuntos de datos. Sin embargo, en el contexto de esta investigación, el principal aporte no radica únicamente en aplicar modelos inteligentes, sino en adaptarlos a una necesidad institucional específica relacionada con la preservación y reutilización de producción académica local.

En el caso del chatbot académico, los usuarios evaluaron positivamente la claridad y utilidad de las respuestas generadas,

obteniendo un 90 % de valoraciones dentro de las categorías más altas de aceptación. Este resultado evidencia que los asistentes conversacionales pueden representar un mecanismo complementario para mejorar la interacción entre los usuarios y los repositorios digitales, al permitir consultas más cercanas al lenguaje natural en lugar de depender exclusivamente de búsquedas mediante palabras clave.

Estos resultados mantienen relación con los avances reportados en modelos de lenguaje contextualizados, donde se ha demostrado que las arquitecturas basadas en aprendizaje profundo pueden mejorar la interpretación semántica del texto y la generación de respuestas asociadas a una consulta específica (Devlin et al., 2019; Brown et al., 2020). A pesar de estos avances, estos sistemas todavía requieren procesos académicos de revisión y validación documental, por lo que su función principal se orienta al apoyo en la recuperación y organización de información.

La decisión de utilizar PHP y MySQL respondió principalmente a las condiciones tecnológicas encontradas durante el desarrollo del proyecto. Aunque existen plataformas especializadas para repositorios digitales, se buscó una alternativa que pudiera mantenerse con los recursos disponibles dentro de la institución.

Aunque actualmente existen plataformas especializadas para repositorios digitales, muchas requieren infraestructura técnica, recursos económicos o personal especializado para su administración. Bajo estas condiciones, Neurorepo muestra que es posible implementar soluciones institucionales adaptadas a escenarios donde la disponibilidad tecnológica puede representar una limitante. En escenarios con restricciones de infraestructura, el diseño de sistemas eficientes y de bajo consumo computacional resulta relevante para acercar capacidades digitales avanzadas a los usuarios finales (Satyanarayanan, 2017). Esta decisión tecnológica buscó equilibrar la incorporación de herramientas avanzadas de IA con las restricciones propias del entorno donde sería utilizado el sistema.

La media general obtenida en la escala de Likert fue de 4.5 sobre 5. Este valor muestra una aceptación favorable del sistema por parte de los usuarios evaluados; sin embargo, debe interpretarse como una validación funcional inicial, ya que la muestra estuvo conformada por un grupo específico de usuarios con experiencia académica y no representa una evaluación generalizada de toda la comunidad universitaria.

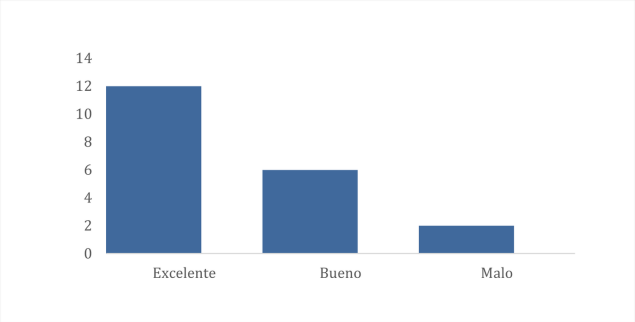

Figura 3

Percepción del impacto institucional de Neurorepo.

Nota. Elaboración propia con datos obtenidos mediante Google Forms (2026).

Durante el desarrollo de Neurorepo se identificó una limitación relacionada con el uso de servicios externos de Inteligencia Artificial para algunas funciones de análisis documental. Esta experiencia mostró la necesidad de conservar una etapa de revisión humana, debido a que las respuestas generadas por IA pueden requerir validación académica (Bender et al., 2021). Asimismo, la calidad de los documentos digitalizados influye directamente en la precisión de extracción automática, especialmente cuando existen archivos con baja resolución o formatos poco estructurados.

La experiencia obtenida durante la implementación de Neurorepo mostró que el uso de Inteligencia Artificial en repositorios institucionales puede apoyar la reducción de brechas tecnológicas en comunidades educativas rurales. Este enfoque coincide con la visión de desarrollo digital orientado a reducir desigualdades mediante soluciones tecnológicas adaptadas al contexto social donde serán implementadas (Heeks, 2018). La relevancia de este enfoque no se limita únicamente a mejorar procesos administrativos, sino a evitar que el conocimiento generado dentro de estas instituciones permanezca aislado o con poca visibilidad debido a la falta de herramientas adecuadas para su gestión.

4.1 Limitaciones de la investigación

Aunque las pruebas realizadas mostraron un funcionamiento adecuado del sistema, durante el proceso se identificaron algunas limitaciones técnicas y metodológicas. Durante la implementación de Neurorepo se identificó que la dependencia parcial de servicios externos de Inteligencia Artificial representa una limitación para ejecutar funciones avanzadas de extracción automática de metadatos y asistencia conversacional. En escenarios donde la conectividad a internet es limitada o inestable, el sistema mantiene únicamente funciones básicas relacionadas con almacenamiento y consulta documental.

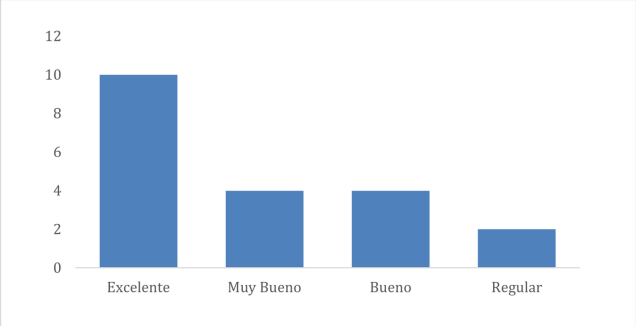

Figura 4
Evaluación de la coherencia y calidad del resumen generado por IA.
Nota. Elaboración propia con datos obtenidos mediante Google Forms (2026).

La mayoría de los evaluadores calificó favorablemente la claridad y coherencia de los resúmenes generados automáticamente por el sistema.

La calidad de los documentos procesados también influyó en el desempeño del módulo de extracción automatizada, ya que problemas como baja resolución, formatos irregulares o errores de escaneo pueden afectar la identificación correcta de autores, títulos y resúmenes académicos.

El alcance de la validación también estuvo condicionado por el tamaño de la muestra utilizada durante la evaluación funcional del sistema. Aunque los participantes contaban con experiencia académica y perfiles especializados, la evaluación se enfocó principalmente en analizar el funcionamiento inicial del sistema, por lo que será necesario ampliar la muestra en futuros estudios antes de establecer una generalización estadística de los resultados.

4.2 Comparación con otros trabajos

Al contrastar Neurorepo con investigaciones relacionadas se encontraron similitudes en el uso de Inteligencia Artificial y PLN dentro de procesos documentales. La diferencia principal del sistema desarrollado se encuentra en su aplicación directa dentro de una institución con necesidades específicas de preservación y consulta de tesis académicas.

Las mejoras observadas en los procesos de extracción automática de metadatos se relacionan con los planteamientos de Kowsari et al. (2019), quienes destacan la capacidad de los modelos basados en PLN para identificar patrones y clasificar información textual. En el caso de Neurorepo, esta capacidad fue aplicada para disminuir la dependencia de registros manuales y facilitar la organización de documentos académicos generados dentro de la universidad.

La incorporación de un asistente conversacional también refleja la evolución de los sistemas tradicionales de búsqueda hacia mecanismos de interacción basados en lenguaje natural. Investigaciones como las de Devlin et al. (2019) y Brown et al. (2020) han demostrado que los modelos de lenguaje pueden mejorar la interpretación contextual de consultas textuales. Los resultados obtenidos durante la evaluación del chatbot muestran una aplicación práctica de estas capacidades dentro de un repositorio académico, permitiendo que los usuarios interactúen con la información mediante consultas más flexibles.

La arquitectura web ligera utilizada en Neurorepo representó una diferencia importante frente a plataformas que requieren mayor infraestructura especializada. Aunque existen soluciones robustas para gestión documental, el desarrollo de Neurorepo buscó equilibrar funcionalidades inteligentes con las condiciones tecnológicas disponibles en una institución rural. Este enfoque coincide con los planteamientos de Rodríguez-Abitia y Correa (2021), quienes destacan la importancia de adaptar los procesos de transformación digital a las características particulares de las instituciones educativas latinoamericanas.

Estos resultados muestran que el aporte principal de Neurorepo va más allá de incorporar herramientas de Inteligencia Artificial, ya que busca resolver una necesidad concreta de organización y acceso al conocimiento académico.

5. CONCLUSIONES

El desarrollo y evaluación de Neurorepo mostró que las herramientas de IA generativa pueden integrarse en repositorios académicos incluso en escenarios con limitaciones tecnológicas. Las pruebas realizadas indican que una arquitectura web ligera, combinada con herramientas de Procesamiento de Lenguaje Natural, puede apoyar la organización, consulta y recuperación de tesis digitales sin requerir infraestructura compleja.

La evaluación realizada con usuarios vinculados a procesos académicos evidenció una percepción favorable respecto al funcionamiento del sistema, principalmente en las tareas relacionadas con extracción automática de metadatos y asistencia mediante chatbot. Los resultados obtenidos muestran que los modelos inteligentes pueden disminuir tareas repetitivas relacionadas con la clasificación documental y facilitar nuevas formas de interacción con repositorios académicos.

La implementación de Neurorepo permitió observar que uno de los desafíos en instituciones rurales no se encuentra únicamente en digitalizar documentos, sino en lograr que la información almacenada pueda ser localizada y utilizada de manera efectiva. Por ello, el desarrollo del repositorio se orientó hacia una solución ajustada a las condiciones reales de la institución, priorizando el acceso y aprovechamiento de la producción académica generada localmente. Bajo este enfoque, Neurorepo representa una propuesta orientada a recuperar, organizar y hacer más accesible la producción académica generada dentro de la propia institución. La elección de una arquitectura basada en PHP, MySQL e integración con servicios de Inteligencia Artificial respondió a la necesidad de equilibrar funcionalidades

inteligentes con una implementación viable en un entorno institucional con recursos tecnológicos limitados. Durante el desarrollo del proyecto se observó que la aplicación de IA en un entorno educativo rural no depende únicamente de utilizar tecnologías avanzadas, sino de adaptar dichas herramientas a problemas concretos y a las condiciones disponibles de la institución. Este aspecto puede beneficiar principalmente a instituciones que buscan implementar herramientas digitales sin depender de infraestructura tecnológica compleja.

Los resultados deben interpretarse considerando el alcance del estudio y las condiciones en las que se evaluó el Sistema. La validación realizada corresponde a una evaluación funcional inicial con una muestra específica de participantes, por lo que futuras investigaciones deberán ampliar el número de usuarios evaluados, incorporar métricas de rendimiento computacional y analizar el comportamiento del sistema durante periodos prolongados de uso institucional.

Como siguiente etapa del proyecto, se plantea fortalecer Neurorepo mediante modelos de búsqueda semántica avanzada, recomendación automática de documentos y análisis inteligente de producción académica. Los próximos desarrollos permitirán evaluar el crecimiento de Neurorepo como una herramienta de apoyo para la gestión del conocimiento académico generado dentro de la universidad. Con ello, se busca fortalecer el acceso al conocimiento académico y apoyar el desarrollo de herramientas digitales más inclusivas dentro de los entornos educativos (UNESCO, 2021).

ANEXO A: INSTRUMENTO DE EVALUACIÓN APLICADO A EXPERTOS

Reactivo	Escala
Precisión de identificación de metadatos	Malo–Excelente
Coherencia del resumen generado	Malo–Excelente
Calidad de palabras clave sugeridas	Malo–Excelente
Normalización APA automática	Malo–Excelente
Precisión de extracción textual	Malo–Excelente

Figura A1
Instrumento de evaluación aplicado a expertos.
Nota. Elaboración propia (2026).

Reactivo	Escala
Utilidad del flujo de trabajo	Malo–Excelente
Facilidad del panel administrativo	Malo–Excelente
Utilidad institucional	Malo–Excelente
Impacto académico esperado	Malo–Excelente
Recomendación del sistema	Malo–Excelente

Figura A2
Evaluación de utilidad e impacto del sistema.
Nota. Elaboración propia (2026).

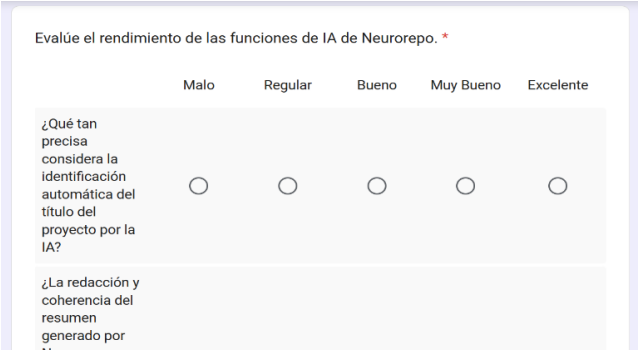

Figura A3
Instrumento de evaluación implementado mediante Google Forms para validar Neurorepo.
Nota. Captura del instrumento utilizado para la validación funcional (2026).

ANEXO B: INTERFACES PRINCIPALES DE NEUROREPO

Figura B1
Pantalla principal del repositorio institucional.
Nota. Interfaz principal de Neurorepo (2026).

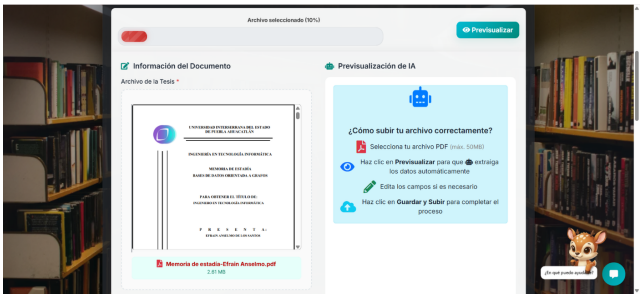

Figura B2
Módulo de carga de tesis PDF.
Nota. Módulo de carga documental implementado en el sistema (2026).

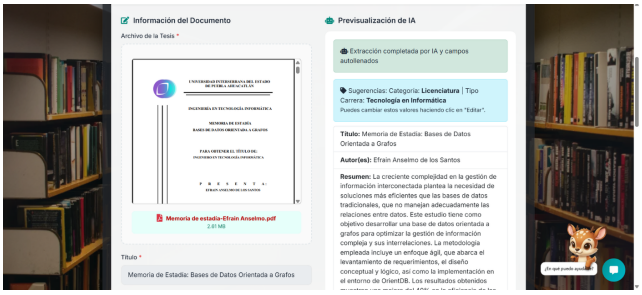

Figura B3
Extracción automática de metadatos.
Nota. Proceso automatizado de identificación de autores, títulos y palabras clave (2026).

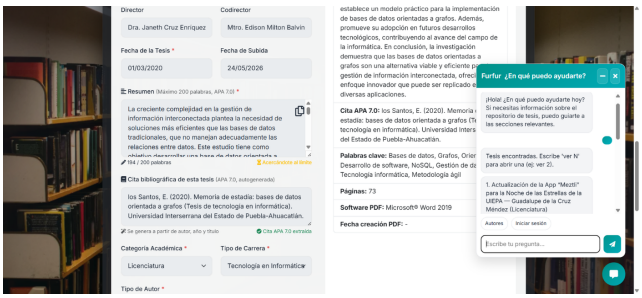

Figura B4

Chatbot institucional.

Nota. Asistente conversacional implementado para consulta académica (2026).

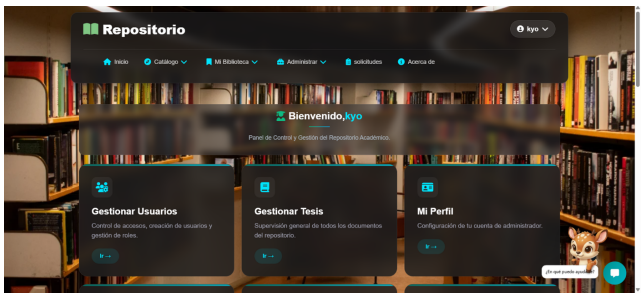

Figura B5

Panel administrativo.

Nota. Módulo administrativo para la organización y control de tesis académicas (2026).

REFERENCIAS

Balvín-Sánchez, E. M., Luis-Rojas, S., & Anselmo-De los Santos, E. (2026). Retos y desafíos en el aprendizaje de la programación: un análisis en estudiantes de ciencias computacionales de la Sierra Norte del estado de Puebla, México. Acta Universitaria, 36, e4517. https://doi.org/10.15174/au.2026.4517

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Borgman, C. L. (2010). Scholarship in the digital age: Information, infrastructure, and the Internet. MIT Press. https://doi.org/10.7551/mitpress/7434.001.0001

Boukhers, Z., & Bouabdallah, A. (2022). Vision and natural language for metadata extraction from PDF documents: A multimodal approach. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, 1–5. https://doi.org/10.1145/3529372.3533295

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165

Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P., & Stal, M. (1996). Pattern-oriented software architecture: A system of patterns. Wiley. https://doi.org/10.5555/249013

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Heeks, R. (2018). ICT4D 3.0? Part 1 – The components of an emerging "digital-for-development" paradigm. The Electronic Journal of Information Systems in Developing Countries, 86(3), 1–14. https://doi.org/10.1002/isd2.12124

Hernández-Sampieri, R., & Mendoza, C. P. (2018). Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta. McGraw-Hill.

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. Information, 10(4), 150. https://doi.org/10.3390/info10040150

Liu, B. (2012). Sentiment analysis and opinion mining. Morgan & Claypool Publishers. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

McTear, M. F. (2020). Conversational AI: Dialogue systems, conversational agents, and chatbots. Morgan & Claypool Publishers. https://doi.org/10.2200/S01060ED1V01Y202010HLT048

Miner, G., Elder, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). Practical text mining and statistical analysis for non-structured text data applications. Academic Press. https://doi.org/10.1016/C2010-0-66188-8

Norman, D. (2013). The design of everyday things (Revised and expanded edition). Basic Books. https://jnd.org/books/the-design-of-everyday-things-revised-and-expanded-edition/

Pressman, R. S., & Maxim, B. R. (2020). Software engineering: A practitioner's approach (9th ed.). McGraw-Hill.

Quindemil Torrijo, E. M., Zambrano Plúa, I. E., & Rumbaut León, F. (2021). Gestión documental en universidades: Una mirada desde Latinoamérica. Revista Científica Multidisciplinaria, 6(Especial), 108–119. https://doi.org/10.33936/rehuso.v6iEspecial.3779

Rodríguez-Abitia, G., & Correa, F. (2021). Digital divide in higher education institutions in Latin America. Education and Information Technologies, 26(3), 3177–3199. https://doi.org/10.1007/s10639-020-10371-7

Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39. https://doi.org/10.1109/MC.2017.9

Suber, P. (2012). Open access. MIT Press. https://doi.org/10.7551/mitpress/9286.001.0001

Umaña-Alpízar, J. (2015). Los repositorios institucionales y el acceso abierto al conocimiento científico en universidades latinoamericanas. Revista e-Ciencias de la Información, 5(2), 1–15. https://doi.org/10.15517/eci.v5i2.19052

UNESCO. (2021). Reimagining our futures together: A new social contract for education. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf000037970

Voutssás-Márquez, J. (2014). Estrategias de conservación y preservación de archivos digitales: elementos técnicos. Benemérita Universidad Autónoma de Puebla. https://iibi.unam.mx/voutssasmt/documentos/archivos_digitales_3_corto.pdf

Weibel, S., Kunze, J., Lagoze, C., & Wolf, M. (1998). Dublin Core metadata for resource discovery. Internet Engineering Task Force (IETF). https://doi.org/10.17487/RFC2413