UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI

FACULTAD DE INDUSTRIAS AGROPECUARIAS Y CIENCIAS AMBIENTALES

CARRERA DE COMPUTACIÓN

Tema: "Sistema inteligente para la asistencia en los trabajos de integración curricular"

Trabajo de Integración Curricular previo a la obtención del título de Ingeniero en Ciencias de la Computación

AUTOR: Villarreal Ortega Gabriel Sebastián.

TUTOR: Ing. Guano Cárdenas Carlitos Alberto, MSc.

Tulcán, 2026.

CERTIFICADO DEL TUTOR

Certifico que el estudiante Villarreal Ortega Gabriel Sebastián con el número de cédula 1750867663 respectivamente ha desarrollado el Trabajo de Integración Curricular: "Sistema inteligente para la asistencia en trabajos de integración curricular"

Este trabajo se sujeta a las normas y metodología dispuesta en el Reglamento de la Unidad de Integración Curricular, Titulación e Incorporación de la UPEC, por lo tanto, autorizo la presentación de la sustentación para la calificación respectiva

Ing. Guano Cárdenas Carlitos Alberto, MSc.

TUTOR

Tulcán, enero de 2026

AUTORÍA DE TRABAJO

El presente Trabajo de Integración Curricular constituye un requisito previo para la obtención del título de Ingeniero en la Carrera de computación de la Facultad de Industrias Agropecuarias y Ciencias Ambientales

Yo, Villarreal Ortega Gabriel Sebastián con cédula de identidad número 1750867663 respectivamente declaro que la investigación es absolutamente original, auténtica, personal y los resultados y conclusiones a los que he llegado son de mi absoluta responsabilidad.

_____

Villarreal Ortega Gabriel Sebastián

AUTOR

Tulcán, enero de 2026

ACTA DE CESIÓN DE DERECHOS DEL TRABAJO DE INTEGRACIÓN CURRICULAR

Yo Villarreal Ortega Gabriel Sebastián declaro ser autor de los criterios emitidos en el Trabajo de Integración Curricular: "Sistema inteligente para la asistencia de trabajos de integración curricular" y eximo expresamente a la Universidad Politécnica Estatal del Carchi y a sus representantes de posibles reclamos o acciones legales.

_____

Villarreal Ortega Gabriel Sebastián

AUTOR

Tulcán, enero de 2026

AGRADECIMIENTO

A mi Padre:

Mi éxito en esta etapa se debe a su apoyo incondicional.

A mi Madre:

Por su constante guía y por enseñarme el valor de la perseverancia y el esfuerzo.

A mi Familia:

Por su respaldo emocional y amor inquebrantable desde el primer día.

A mi tutor:

Por su guía académica y técnica para el presente trabajo durante la trayectoria.

A mis profesores:

Por su dedicación a la difusión de conocimientos y a la influencia en mi formación profesional.

Al Instituto Superior Tecnológico 17 de Julio:

Por el apoyo y la oportunidad de trabajar.

DEDICATORIA

Dedico este Trabajo de Integración Curricular, en primer lugar, a Dios, por haber sido mi guía constante en cada etapa de este camino, por darme fortaleza en los momentos de dificultad, claridad en las decisiones y la perseverancia necesaria para no rendirme cuando el esfuerzo parecía superar las fuerzas.

A mis padres, por ser el pilar fundamental de mi vida y de mi formación. Gracias por su amor incondicional, por cada sacrificio silencioso, por creer en mí incluso cuando yo dudaba, y por enseñarme que la disciplina, el respeto y el trabajo honesto son la base de todo logro verdadero. Este título es también suyo.

A mi familia, por el apoyo permanente, por la paciencia, la comprensión y las palabras de aliento que hicieron más llevadero este recorrido universitario. Su respaldo emocional fue una fuerza constante que me impulsó a seguir adelante.

A mis docentes, por compartir su conocimiento con compromiso y vocación, por orientar mi formación académica y profesional, y por sembrar en mí el pensamiento crítico, la responsabilidad y el deseo de superación. Cada enseñanza recibida dejó una huella que hoy se refleja en este trabajo.

Este proyecto representa no solo la culminación de una etapa académica, sino el resultado de años de esfuerzo, aprendizaje y crecimiento personal. Lo dedico a todos quienes caminaron conmigo en este proceso y contribuyeron, directa o indirectamente, a que este objetivo se hiciera realidad.

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

ÍNDICE DE ANEXOS

RESUMEN

El objetivo del presente Trabajo de Integración Curricular fue desarrollar un sistema inteligente para optimizar la gestión, búsqueda e indexación de los trabajos de titulación del Instituto Superior Tecnológico 17 de Julio. Actualmente, estos procesos se realizan de forma manual, lo que genera demoras, duplicación de esfuerzos y dificultades en el acceso a la información académica por parte de estudiantes y docentes, afectando la eficiencia institucional. El sistema desarrollado integra técnicas de inteligencia artificial mediante un enfoque híbrido que combina reconocimiento óptico de caracteres (OCR), procesamiento del lenguaje natural (NLP) e indexación estructurada de documentos académicos. La metodología empleada corresponde a un enfoque mixto, de tipo cuantitativo y cualitativo, con una investigación descriptiva y explicativa. Para el desarrollo del sistema se utilizó una arquitectura modular basada en Flask, permitiendo la extracción automática de texto desde documentos digitales y escaneados, la segmentación por párrafos y la recuperación eficiente de información relevante. Adicionalmente, el sistema incorpora un módulo de seguimiento y evaluación académica que permite registrar observaciones, estados de revisión e historial de evaluaciones de los trabajos de titulación, fortaleciendo la trazabilidad y el control académico institucional. La validación del sistema se realizó mediante encuestas y entrevistas aplicadas a estudiantes y personal académico, así como pruebas técnicas de rendimiento orientadas a medir precisión, tiempo de búsqueda y eficiencia en la recuperación de información. Los resultados obtenidos evidencian un incremento significativo en la precisión del OCR, una reducción considerable en los tiempos de búsqueda y una mejora fundamental en la eficiencia de recuperación de información en comparación con los procesos manuales tradicionales. En conclusión, el sistema desarrollado constituye una herramienta tecnológica eficaz que contribuye a la modernización de la gestión documental académica, optimiza los procesos de titulación y fortalece el acceso al conocimiento institucional.

Palabras clave: inteligencia artificial, trabajos de titulación, procesamiento del lenguaje natural.

ABSTRACT

The objective of this Capstone Project was to develop an intelligent system to optimize the management, search, and indexing of graduation projects at the Instituto Superior Tecnológico 17 de Julio. Currently, these processes are carried out manually, which leads to delays, duplicated efforts, and difficulties in accessing academic information for both students and instructors, ultimately affecting institutional efficiency. The system developed integrates artificial intelligence techniques through a hybrid approach that combines optical character recognition (OCR), natural language processing (NLP), and structured indexing of academic documents. The methodology followed a mixed approach, incorporating quantitative and qualitative components, and was grounded in descriptive and explanatory research. For system development, a modular architecture based on Flask was implemented, enabling automatic text extraction from digital and scanned documents, paragraph-level segmentation, and efficient retrieval of relevant information. Additionally, the system includes an academic tracking and evaluation module that allows users to record observations, review statuses, and evaluation histories for graduation projects, thereby strengthening institutional academic traceability and oversight. System validation was conducted through surveys and interviews with students and academic staff, as well as technical performance tests designed to measure accuracy, search time, and information-retrieval efficiency. The results demonstrate a significant increase in OCR accuracy, a considerable reduction in search times, and a substantial improvement in information-retrieval efficiency compared to traditional manual processes. In conclusion, the system developed constitutes an effective technological tool that supports the modernization of academic document management, optimizes graduation-project processes, and enhances access to institutional knowledge.

Keywords: artificial intelligence, degree projects, natural language processing.

INTRODUCCIÓN

El Instituto Superior Tecnológico 17 de Julio evidenció problemas en la administración y el acceso a la información que se encuentra en los Trabajos de Integración Curricular (TIC), lo cual fue el punto de partida para este proyecto investigativo. Aunque las instituciones se esfuerzan por mantener un repositorio ordenado, hay todavía grandes retos: la búsqueda manual de documentos requiere mucho tiempo, los datos están desparramados en diferentes formatos y caminos, y no se cuentan con herramientas tecnológicas para hacer consultas que posibiliten analizarlas de forma exacta y rápida.

El primer capítulo tiene como objetivo principal la creación de un sistema inteligente, fundamentado en métodos de inteligencia artificial, que facilite la automatización de la búsqueda, indexación y recuperación de datos TIC. Los propósitos concretos también se establecen, con el propósito de examinar el proceso actual de gestión documental, incorporar módulos de extracción de texto, OCR e indexación en una plataforma unificada y valorar el rendimiento del sistema en comparación con los procedimientos manuales que se están utilizando hoy en día.

En el segundo capítulo se expone la justificación teórica del estudio, que incluye los temas de digitalización híbrida, bases de datos semiestructuradas, sistemas para recuperar información, tecnologías vinculadas con el procesamiento de texto y reconocimiento óptico de caracteres. Se subraya, además, que en Ecuador la digitalización de documentos en la educación superior todavía progresa con lentitud, lo cual demuestra que es necesario implementar opciones eficaces para actualizar el acceso al saber académico.

El tercer capítulo expone las variables analizadas, siendo la asistencia académica en los trabajos de integración curricular una variable dependiente y el sistema inteligente, una variable independiente. Asimismo, se emplea un método mixto que integra técnicas cualitativas y cuantitativas como encuestas, entrevistas y análisis de documentos para conseguir una perspectiva integral del funcionamiento presente y del impacto proyectado del sistema.

Por último, en el cuarto capítulo se presentarán las conclusiones y recomendaciones fundamentadas en los resultados obtenidos, considerando la validación y aplicación del sistema.

I. EL PROBLEMA

1.1. PLANTEAMIENTO DEL PROBLEMA

Actualmente, la inteligencia artificial sigue siendo una de las principales puertas del progreso tecnológico en educación, ya que permite la automatización de procesos, la optimización del manejo de la información y la mejora de la calidad académica. Según la UNESCO (2023), la IA está, de formas inéditas, cambiando la manera en que las instituciones gestionan y administran el conocimiento, proporcionando diversas oportunidades en educación, investigación e innovación. Sin embargo, la incorporación de nuevas tecnologías sigue siendo un problema en educación y sistemas de enseñanza debido a limitaciones en infraestructuras, formación de educadores y dificultades para incorporar nuevas tecnologías en entornos educativos a nivel mundial.

En el país, Ecuador se encuentra en un proceso de educación digital, que se potencia a través de políticas que facilitan el empleo de nuevas tecnologías en la administración institucional (Medina-Romero, 2025). Sin embargo, en el ámbito de la educación técnica y tecnológica, la utilización de tecnologías de la información y la comunicación como instrumentos de inteligencia artificial se encuentra a un nivel muy elemental, conforme a los procesos administrativos y académicos que se llevan a cabo de forma manual y, como lo delinean A. Ávila et al. (2024), el aprovechamiento de la IA en el ámbito educativo ecuatoriano dependerá de la capacidad de las instituciones para desarrollar sistemas a medida.

En este contexto, el Instituto Superior Tecnológico 17 de Julio (IST17J) enfrenta el desafío de acceder de manera rápida y eficaz a la información que ya existe en los Trabajos de Integración Curricular (TIC) que se encuentran en formato digital. El acceso a la documentación, el examen de los antecedentes y la obtención de información necesaria se llevan a cabo manualmente. Esto provoca una pérdida de tiempo, un exceso de esfuerzos y problemas con la gestión del conocimiento institucional. Esta circunstancia tiene un impacto negativo en la oportunidad de

acceder a los recursos académicos y empeora las evaluaciones y las asesorías de los proyectos de los estudiantes.

Un problema considerable es la descentralización de la administración de los trabajos de titulación, que genera poca capacidad para rastrear documentos académicos, dispersión de datos y duplicación de esfuerzos. Estos son algunos ejemplos de las limitaciones administrativas y tecnológicas. Los TIC están guardados en diversos dispositivos, formatos y rutas, pero no hay un sistema unificado que posibilite su consulta de manera eficiente y ordenada. Esta fragmentación complica la continuidad del conocimiento institucional, hace más difícil el seguimiento académico y retrasa la recuperación de antecedentes. Por lo tanto, la gestión de documentos del IST17J depende de procedimientos manuales que requieren mucho tiempo, disminuyen la exactitud de las búsquedas y restringen el acceso a información esencial para los alumnos, los maestros y las autoridades académicas.

El IST17J gestiona cada año cerca de 300 tesis y documentos académicos, así como actas, certificados y procesos administrativos. De acuerdo con datos institucionales, la búsqueda manual de información requiere un promedio de 10 minutos por documento. Esto representa 3,000 minutos mensuales, 36,000 anuales o cerca de 600 horas académicas. La falta de un sistema centralizado y el aumento en la cantidad de documentación dan lugar a duplicidad de tareas, demoras acumuladas y riesgo de pérdida de información académica fundamental.

1.2. FORMULACIÓN DEL PROBLEMA

¿En qué forma del desarrollo de un sistema inteligente puede mejorar la optimización de la búsqueda, la indización y la recuperación de la información contenida en los trabajos de integración curricular del Instituto Superior Tecnológico 17 de Julio?

1.3. JUSTIFICACIÓN

El trabajo está motivado por la necesidad de la institución de mejorar la búsqueda, la indexación, así como la recuperación de datos en los Trabajos de Integración Curricular (TIC) de la institución. En este momento, las tareas se realizan de forma manual, generando en consecuencia, un desgaste de tiempo, esfuerzo y un limitado aprovechamiento sobre el conocimiento académico. Esta situación, que impacta de manera negativa, tanto a los estudiantes como a los docentes, retrasa el trabajo de los proyectos institucionales en torno a la elaboración, revisión y evaluación.

A nivel mundial, la Inteligencia Artificial es un importante eje para el modernismo en la Administración Educativa. Según la UNSECO (2023), la AI facilita la administración del conocimiento y torna la planeación de los sistemas en las instituciones de educación superior, de manera más eficiente y accesible. En el Ecuador, la IA como un componente de la virtualización, ha ido en progreso también, aunque acompañado de ciertas limitantes, como la infraestructura tecnológica y la flexibilidad de la organización (Medina-Romero, 2025). Igualmente, el uso de la IA en la educación, especialmente en la educación técnica y tecnológica, es limitado, y en la mayor parte de los casos, la automatización de los procesos documentales es escasa. Según Ávila et al. (2024), en el Ecuador el aprovechamiento de la Inteligencia Artificial a nivel educativo depende, fundamentalmente, de la capacidad de las instituciones para crear respuestas que se alineen a sus requerimientos.

De acuerdo con el IST17J, se necesita un sistema que podrá automatizar el proceso de búsqueda y análisis de información. Esto debería de disminuir el tiempo de búsqueda de información y aumentar la precisión del análisis de la información. La propuesta que se realizó dentro de este estudio busca el establecimiento de un sistema que puede procesar, clasificar y recuperar información usando técnicas de IA. Según se indica por Taheri et al. (2025), la utilización de los sistemas de IA se puede realizar de forma más eficiente y efectiva si se tiene en cuenta la infraestructura, formación, y disposición de la institución, y en el caso del Instituto, los tres elementos se alinean a los objetivos de modernización que tiene la Institución.

El beneficio principal que se espera de la investigación es la automatización de los procesos repetitivos que derivan del manejo de los documentos en el área de TIC. Esto se espera que mejore la eficiencia en los procesos administrativos y que, a la vez, se logre un acceso más ágil a la información requerida en la parte académica. En la parte teórica, se espera que el proyecto proporcione un modelo que facilite la aplicación de técnicas de IA en el entorno educativo, con OCR, PLN, e indexación automatizada, que puede servir como punto de partida para avanzar en otras aplicaciones y líneas de investigación.

Finalmente, el impacto social de este proyecto se reflejará en el desarrollo de la calidad de la formación, en el fortalecimiento de la cultura digital de la institución y en la dotación de herramientas tecnológicas que van a ser de beneficio para docentes y estudiantes. Así, la investigación se suma al proceso de transformación digital del IST17J, logrando una gestión documental más rápida y sostenible.

1.4. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN

1.4.1. Objetivo General

Desarrollar un sistema inteligente basado en técnicas de inteligencia artificial para optimizar los procesos de búsqueda, indexación y extracción de información en los Trabajos de Integración Curricular del Instituto Superior Tecnológico 17 de Julio.

1.4.2. Objetivos Específicos

- Analizar el proceso actual de gestión documental del IST17J para identificar limitaciones y requerimientos del sistema.
- Integrar las técnicas de inteligencia artificial dentro de una plataforma unificada que automatice la gestión documental del IST17J.
- Evaluar la eficiencia, precisión y tiempo de respuesta del sistema inteligente en comparación con el proceso tradicional.

1.4.3. Preguntas de Investigación

- ¿Cuáles son las principales limitaciones y requisitos del proceso de gestión de documentos del IST17J que impiden la búsqueda y recuperación eficiente de información sobre las Disertaciones de las Unidades de Integración?
- ¿Cómo se puede integrar las técnicas de inteligencia artificial dentro de una plataforma que automatice la gestión de documentos en el IST17J?
- ¿En qué medida el sistema inteligente desarrollado aumenta la eficiencia, precisión y tiempo de respuesta en la búsqueda e indexación de información en comparación con el proceso tradicional actualmente en uso en el IST17J?

II. FUNDAMENTACIÓN TEÓRICA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

La inteligencia y visión por computadora han mejorado y diversificado las herramientas para la gestión documental en el ámbito académico. Estas herramientas, en combinación con el reconocimiento óptico de caracteres (OCR), convierten textos de imágenes y documentos escaneados a información editable, mejorando la búsqueda, recuperación y el análisis de información de archivos. Es por esto que su uso en educación, sobre todo en ambientes universitarios y en la educación superior, ha facilitado la digitalización de trabajos de titulación y el uso más eficiente de la información institucional.

Alzoubi, Alqatawna y Jaradat (2024) describe el diseño de un sistema de recuperación de imágenes documentales (Document Image Retrieval) en el aula, que permite el acceso a la información contenida en un documento impreso a través de la visión por computadora.

> Los autores mostraron que los DIR aumentan la eficiencia y la precisión del reconocimiento del texto, recuperando información a partir de documentos impresos y capturados en dispositivos móviles, lo que valida la aplicación de la visión artificial en la gestión del aprendizaje (Alzoubi et al., 2024, p. 4).

Residiendo en la misma zona horaria, Sharma y Mehta (2025) propusieron un modelo híbrido que combina OCR y Procesamiento de Lenguaje Natural (NLP) para automatizar el proceso de extracción de información en documentos PDF.

> Se determinó en la investigación que la combinación de reconocimiento óptico y análisis semántico mejora la clasificación de textos y la calidad de los resultados de búsqueda a un nivel más profundo, lo que servirá como un bloque de construcción para sistemas inteligentes de gestión documental (Sharma & Mehta, 2025, p. 8).

En este contexto, Kamaleson, Chu y Otero (2022) construyeron un sistema sólido para la extracción automática de información de documentos electrónicos, aplicando algoritmos de aprendizaje automático.

La literatura sobre inteligencia artificial nos muestra que, al aumentar la necesidad de usar tecnologías de vanguardia para la identificación de objetos, la disminución de la falla e inspección manual en la detección de productos se optimiza, y se logra una evaluación de calidad más eficiente para productos como las fresas, al adoptar sistemas de visión artificial. Estos sistemas pueden y son capaces de, en tiempo real, analizar características visuales y, por lo tanto, elevar los estándares de calidad en la producción agrícola, en el ámbito de la recuperación documental, Kettunen, Koistinen y Pääkkönen (2022) demostraron la relación directa entre la calidad del OCR y la experiencia del usuario.

> Los resultados mostraron que las mejoras en la precisión de OCR del 7.94\% resultaron en niveles más altos de satisfacción del cliente al recuperar documentos, reafirmando la necesidad de un OCR de mayor calidad en sistemas automáticos de consulta (Kettunen et al. 2022, 6).

Desde una perspectiva diferente, Bazzaco, Gaviña y Camacho (2022) analizaron los principales desafíos técnicos en la segmentación y el análisis de paginación al digitalizar los textos.

> Por la indagación realizada por estos autores se sabe que la de OCR alcanza la mayor fiabilidad posible en escaneos de mayor calidad, en escaneos que se limpien y en escaneos que se ajusten correctamente a un orden de lectura, ya que se requiere de varios algoritmos como el de separación y el de supresión de ruido (Bazzaco et al., 2022, p. 3).

Al mismo tiempo, Kamaleson, Chu y Otero (2022) desarrollaron un sistema robusto para la extracción automática de información de documentos electrónicos utilizando algoritmos de aprendizaje automático.

> La investigación realizó la construcción de una arquitectura tales como: detección de objetos, procesamiento de imágenes y reconocimiento óptico que, comparado como método tradicional, se pudo superar la precisión en el reconocimiento, gracias a los métodos manuales (Kamaleson et al., 2022, p. 5).

En Huillca Tumba (2023) y Nacional, se ha desarrollado un sistema para la indexación automática de las tesis de la Facultad de Ingeniería Hualtina, lo que permite la clasificación de documentos por área de especialización utilizando técnicas de segmentación de texto y análisis de contenido. Este proyecto mostró la aplicabilidad de la IA en los repositorios educativos ecuatorianos por primera vez y mejoró significativamente la organización de la información en las instituciones de educación superior.

Específicamente, Erazo Narváez (2023) propuso un sistema inteligente para el Monitoreo de Documentos basado en visión por computadora, que tiene la capacidad de encontrar secciones relevantes de documentos mediante técnicas de OCR en informes académicos. El autor demostró que la automatización de la búsqueda y clasificación de documentos disminuye los tiempos de procesamiento y mejora la trazabilidad de los trabajos de graduación.

2.2. MARCO TEÓRICO

2.2.1. Sistemas basados en reglas

Por ser una de las áreas más explicables de la inteligencia artificial, estos sistemas pueden ser utilizados en aplicaciones donde se requiere una alta transparencia, tales como la medicina o la banca. A diferencia de las redes neuronales o las probabilísticas, donde la creación de modelos son un verdadero 'arte', en estos sistemas la programación se centra en una estructura lógica bien definida, Bianchi y Rossi (2021). Por su bajo costo de mantenimiento y alta aplicabilidad, estos sistemas pueden ser utilizados en entornos donde la trazabilidad es un factor crítico.

Según la UNESCO (2023), en años recientes, las organizaciones internacionales han resaltado que las soluciones fundamentadas en reglas son opciones eficaces para instituciones que quieren automatizar procesos sin tener que invertir en modelos costosos o complejos. Esto abarca procedimientos como la validación de patrones textuales, la extracción de información estructurada o la clasificación de documentos; todos ellos habituales en el ámbito educativo.

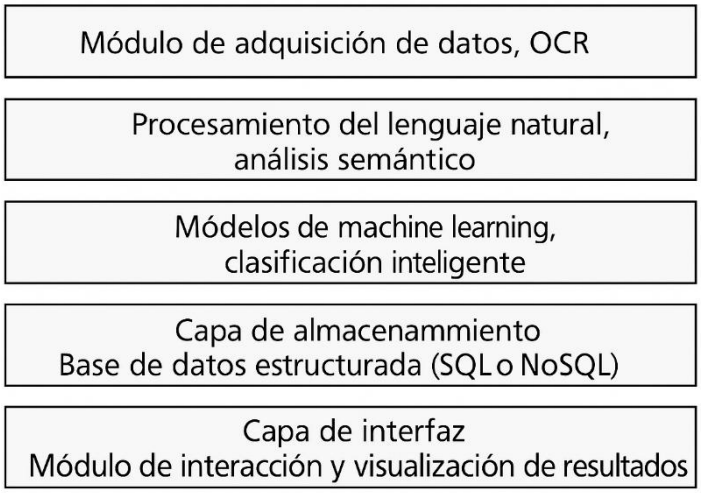


Figura 1. Arquitectura general de un sistema inteligente modular

Según la literatura reciente, los sistemas que se basan en reglas tienen un rendimiento elevado cuando los documentos que procesan tienen estructuras bastante homogéneas, tal y como sucede con las tesis académicas (Niculescu et al., 2023). Este patrón estructural permite el uso de heurísticas exactas y reduce la necesidad de algoritmos complejos para entender el contenido.

Por último, este método posibilita garantizar un funcionamiento de bajo costo, escalabilidad y sostenibilidad. Según la OECD (2025), los sistemas ligeros fundamentados en reglas son apropiados para entidades educativas que necesitan aplicar soluciones tecnológicas prácticas y que no requieran mucho mantenimiento. Esto confirma completamente la elección técnica hecha en el sistema.

2.2.2. Procesamiento de texto basado en patrones

El procesamiento de texto por patrones son un conjunto de enfoques cuyo objetivo es la identificación, la extracción y la transformación de información siguiendo procedimientos normativos estructurados. A diferencia del procesamiento de lenguaje natural de última generación, este análisis no implica un entendimiento del texto más allá del reconocimiento de patrones. Esta clase de procesamiento, así como documentan Sharma y Menon (2025), es aún muy útil en documentos donde el formato es relativamente constante y se requiere rapidez y confiabilidad más que un análisis en profundidad.

Las expresiones regulares (regex) son quizás la herramienta más representativa de este enfoque. Permiten la identificación de secciones y elementos como

encabezados, nombres propios y delimitadores a través de especificaciones descriptivas en lenguaje natural. Recientemente, se ha documentado que las regex son muy comúnmente usadas en el ámbito académico, debido a su calidad y bajo costo computacional (Li & Yu, 2022). A su vez, esto las hace muy útiles en la validación de texto por su flexibilidad.

El uso de patrones también es central para la extracción de la sección de 'Introducción' debido a la prevalencia de documentos de estructura fija. Investigaciones recientes sugieren que para documentos que siguen un formato consistente, la detección basada en patrones es más eficiente que los modelos semánticos (Torres & Delgado, 2022), particularmente cuando se necesita precisión para extraer secciones estructuradas específicas.

Por lo tanto, el procesamiento de texto basado en patrones es una base técnica esencial para el sistema, ya que permite la automatización de la interpretación del contenido principal sin la necesidad de algoritmos sofisticados.

Tabla 1. Comparación entre el Aprendizaje supervisado y no supervisado

Característica	Aprendizaje Supervisado	Aprendizaje No Supervisado
Definición	Entrenamiento con datos etiquetados que relacionan entradas con salidas esperadas.	Entrenamiento con datos sin etiquetas para descubrir patrones ocultos o agrupamientos.
Objetivo	Predecir valores o clasificar nuevas observaciones.	Identificar estructuras, similitudes o relaciones dentro de los datos.
Ejemplos de algoritmos	Regresión lineal, árboles de decisión, SVM, redes neuronales.	K-Means, DBSCAN, PCA, clustering jerárquico.
Evaluación	Basada en precisión, recall, F1-score, error cuadrático medio.	Difícil de evaluar; se utilizan métricas de cohesión o separación.
Aplicaciones comunes	Detección de fraudes, análisis de sentimientos, reconocimiento de voz.	Segmentación de clientes, reducción de dimensionalidad, análisis exploratorio.
Ventajas	Alta precisión cuando existen datos etiquetados.	Útil cuando no se dispone de etiquetas o conocimiento previo.
Limitaciones	Requiere grandes volúmenes de datos etiquetados.	Resultados más difíciles de interpretar.

2.2.3. Recuperación de Información (IR) aplicada a documentos académicos

La búsqueda y organización eficaz de grandes cantidades de documentos textuales es el enfoque fundamental de la Recuperación de Información (IR, por sus siglas en

inglés). Las técnicas de IR se han utilizado extensamente en repositorios académicos entre 2020 y 2025, ya que tienen la habilidad de simplificar la consulta tanto de literatura científica como de trabajos institucionales (Arslan & Mete, 2023). Los motores IR convencionales funcionan a través de búsqueda por palabras clave, indexación y modelos booleanos o de coincidencia literal.

La indexación granular es particularmente beneficiosa en contextos educativos donde los usuarios buscan conceptos específicos, de acuerdo con Li y Yu (2022).

Según la literatura más reciente, los sistemas IR que utilizan estructuras semiestructuradas (por ejemplo, JSON) brindan un rendimiento apropiado y una elevada capacidad de portabilidad para las plataformas web institucionales (Bianchi & Rossi, 2021). Esto es consistente con la concepción del sistema, que fue creado para funcionar en servidores educativos sin necesidad de bases de datos complejas.

2.2.4. Recomendación óptica de caracteres (OCR)

El Reconocimiento Óptico de Caracteres (OCR) se entiende como la tecnología que permite la conversión de textos que se encuentran contenidos en documentos impresos que han sido escaneados, imágenes o documentos PDF, a texto digital editable. Esta tecnología sirve para automatizar la información que se digitaliza, disminuyendo la carga de trabajo que involucra la transcripción de documentos (Patel & Desai, 2023).

La tecnología OCR opera a través de ciertos algoritmos que permiten el reconocimiento de patrones, así como también de sistemas de aprendizaje automático, que se dedican a la identificación de caracteres, números o símbolos que se encuentran en una cierta imagen, a partir de la comparación con patrones de la tipografía con los que se ha entrenado para la generación de un texto (Smith, 2020).

El OCR se basa en diversas herramientas, como por ejemplo Tesseract OCR, el cual es de uso libre y es accesible para la zona académica; OpenCV, que permite la detección y la segmentación de textos; y PyMuPDF, cuya orientación está en el procesamiento de documentos PDF (Shafait et al., 2021).

Dentro de los últimos avances del OCR se ha logrado mejorar la precisión, especialmente al aplicar resoluciones de 300 DPI o más. Esto es sumamente importante en el rubro de la digitalización educativa (Kettunen et al., 2022).

El sistema desarrollado utiliza un enfoque híbrido. Primero, intenta extraer texto nativo utilizando PyMuPDF. Si el documento no tiene una capa de texto, realiza OCR a 300 DPI utilizando Tesseract. Tener un enfoque de dos pasos mejora la precisión de lectura y mitiga los problemas comunes al intentar extraer texto de documentos escaneados, haciendo que la extracción de material en repositorios institucionales sea más eficiente (Sharma & Singh, 2021).

La tabla 2 muestra una comparación de motores OCR de acuerdo con sus principales características técnicas.

Tabla 2. Comparación de motores OCR

Motor OCR	Precisión promedio (%)	Velocidad de procesamiento	Compatibilidad
Tesseract	95-98	Alta	Windows, Linux, macOS
PyMuPDF	92-95	Media	Windows, Linux
OpenCV OCR	90-94	Alta	Multiplataforma

Asimismo, el sistema incluye OCR por bloques para reconocer partes concretas, como el "Resumen". Este procedimiento hace uso de la detección de palabras por medio de coordenadas, lo que posibilita encontrar encabezados aun en documentos que presentan artefactos propios del escaneo o ruido visual. Torres y Delgado (2022) afirman que la localización a través de bloques es capaz de detectar secciones estructuradas mejor que el OCR tradicional.

El OCR tiene una función institucional además de una funcional: posibilita la conservación de documentos, la garantía de su accesibilidad y la habilitación de procesos posteriores como análisis, búsqueda e indexación. La digitalización por medio de OCR es un paso fundamental para conservar el conocimiento académico, según organismos como la UNESCO (2023).

2.2.5. Sistemas expertos e inteligentes

El proceso por el cual se separa un documento en partes más pequeñas y fáciles de manejar, como párrafos, bloques textuales o páginas, se conoce como segmentación documental. Según autores como Yu y Li (2022), la segmentación apropiada es crucial para optimizar la eficacia de los sistemas de recuperación de información en años recientes, sobre todo en contextos educativos donde los documentos continúan con formatos relativamente estandarizados.

El sistema utiliza un mecanismo de segmentación que se basa en párrafos y emplea saltos de línea dobles como separadores. Esto posibilita la división del contenido en

partes coherentes que simbolizan conceptos individuales o fragmentos completos del texto. Según Arslan y Mete (2023), la segmentación de documentos en párrafos aumenta la exactitud en las búsquedas y hace más sencillo extraer partes relevantes.

Tabla 3. Componentes y descripción técnica

Componente	Descripción técnica
Motor OCR híbrido (PyMuPDF + Tesseract 300 DPI)	Extrae texto digital mediante PyMuPDF; si no existe capa textual, ejecuta OCR a 300 DPI con Tesseract, permitiendo procesar tesis digitales y escaneadas con alta compatibilidad.
Sistema de Recuperación de Información (IR) por párrafos	Indexa cada párrafo como unidad independiente en JSON, mejorando la granularidad de la búsqueda y la precisión en repositorios académicos.
Motor de búsqueda por coincidencia literal	Implementa un modelo booleano basado en coincidencia directa de cadenas para localizar términos en párrafos específicos.
Extracción basada en patrones (regex)	Identifica metadatos como AUTOR, TUTOR y secciones mediante expresiones regulares adaptadas al formato institucional.
Identificación heurística del resumen	Detecta encabezados de 'RESUMEN' y delimita contenido hasta 'PALABRAS CLAVE', usando OCR por bloques cuando es necesario.
Arquitectura modular Flask	Separa capas de presentación, lógica y procesamiento, integrando módulos externos mediante rutas y blueprints.
Repositorio documental en JSON	Base semiestructurada que almacena párrafos indexados y metadatos, facilitando su consulta y portabilidad.
Segmentación y normalización textual	Divide el contenido por saltos de línea, limpia ruido y normaliza espacios para fortalecer la indexación.
Módulo de extracción de metadatos	Analiza la primera página del PDF para capturar autor y tutor mediante patrones estructurales.
Gestor de digitalización híbrida	Decide automáticamente entre extracción digital u OCR, optimizando rendimiento y accesibilidad del contenido.

2.2.6. Digitalización documental en Ecuador Superior

Entre 2020 y 2025, digitalizar documentos se volvió una prioridad estratégica para las entidades de educación superior. Según la OECD (2025), la digitalización ayuda a optimizar la administración, aumentar el acceso a información y reforzar el seguimiento de los procedimientos educativos. Esto comprende transformar trabajos de titulación en formatos que sean accesibles y que posibiliten su consulta y conservación a largo plazo.

De acuerdo con Torres y Delgado (2022), los sistemas que automatizan estas funciones disminuyen el tiempo de gestión administrativa y hacen más accesible la información académica para estudiantes y profesores.

Además, la digitalización hace más fácil que se cumplan las normas institucionales y asegura que los documentos puedan ser usados de nuevo en estudios académicos o investigaciones posteriores. La UNESCO (2023) admite que la conservación digital es un elemento fundamental para mantener la continuidad histórica de las instituciones educativas.

2.2.7. IR documental/ indexación

La Recuperación Documental consiste en un conjunto de procesos que hacen posible organizar, guardar y recuperar fragmentos de colecciones de documentos. Recientes estudios (Arslan & Mete, 2023) argumentan que, de los sistemas IR disponibles en la actualidad, los sistemas IR Clásicos, por su eficacia y aplicabilidad, siguen dominando el mercado. Lo anterior, por el bajo costo computacional que tienen. Esto es especialmente importante en el ámbito de la educación, dado que en este entorno hay un alto volumen documental y es necesario un acceso rápido (Arslan & Mete, 2023). Estos sistemas incorporan indexación estructurada, almacenamiento semiestructurado y análisis literal del contenido.

Li y Yu (2022) argumentan que la fragmentar un documento en unidades pequeñas mejora la precisión del sistema por la probabilidad de localizar conceptos en el contenido. Esto se debe que el sistema opera sobre fragmentos más coherentes. En el caso del sistema, cada párrafo extraído se almacena como objeto independiente dentro de un repositorio tipo JSON.

La indexación de documentos dentro del sistema utiliza un proceso híbrido de dos pasos, en el que el texto en bruto se extrae primero del PDF, después de lo cual se lleva a cabo la limpieza del espacio, la eliminación de ruido y la normalización de la restructuración del texto. Este proceso está en conformidad con las recientes recomendaciones relacionadas con la consolidación de contenido antes de la indexación para asegurar firmeza y consistencia (Raghavan & Patel, 2024).

2.2.8. Búsqueda de texto completo

La búsqueda de texto completo busca encontrar partes que incluya literalmente lo que un usuario ha tecleado. Esta actividad que se sigue realizando, se conoce como búsqueda de texto completo y se considera como una de las actividades más representativas de IR debido a su velocidad y a su bajo costo en la implementación. Recientes se han indicado que en las búsquedas en repositorios institucionales donde los documentos guardan cierta homogeneidad en su estructura la búsqueda seudo-

clásica, si bien es una búsqueda literal, funciona con una cierta confiabilidad sin necesidad de recurrir a modelos complejos de análisis semántico (Bianchi & Rossi, 2021).

La búsqueda literal también permite el mejoramiento de la interfaz del usuario debido a que resalta visualmente los términos buscados, lo que le permite a Usuarios le mejora la experiencia de indización al poder localizar, de analogista su localización, carta de coincidencia dentro del texto. Mientras más un experto se ha informado que, al operar sobre unidades discretas como documentos, Ej. un párrafo, un texto se reduce a lo relevante al texto.

La fuerte ineficiencia de estas búsquedas radica en la impresión de falta de conocimientos de los significados de las palabras buscadas. Y a pesar de ello, los motores de búsqueda basados en coincidencia directa continúan siendo utilizados en bibliotecas y sistemas bibliográficos institucionales debido a su bajo costo, baja media de mantenimiento, y alta capacidad de los sistemas. (Li & Yu, 2022). Por todo esto, la decisión de la arquitectura de búsqueda literaria se ha convertido en una opción tecnológica de gran valor y eficiencia (por problemas de e) para colegios e instituciones.

2.2.9. Digitalización híbrida

Digital híbrido es una combinación de extracción de texto digital y reconocimiento óptico de caracteres con el fin de procesar documentos independientemente de su formato original. En los últimos años, esta estrategia se ha comenzado a implementar en procesos de educación por la coexistencia de documentos procesados digitalmente con documentos escaneados (Torres y Delgado, 2022). El propósito es maximizar la accesibilidad de los contenidos y evitar el riesgo de pérdida de información que se asocia a la ausencia de una capa de texto.

La digitalización híbrida también incrementa el nivel de certeza en procesos que se llevarán a cabo por documentos con texto, tales como la indexación, búsqueda y el análisis de documentos. Estudios más recientes concluyen que los sistemas que implementan digitalización híbrida y reconocimiento óptico de caracteres alcanzan mejores porcentajes de completitud y coherencia en la producción de texto (OECD, 2025). En el caso de la sistematización, esta estrategia permite a la construcción de un repositorio de documentos a partir de la digitalización y el escaneo, los repositorios son uniformes a pesar de las distintas procedencias de sus documentos.

2.2.10. Extracción basada en patrones

Las técnicas que buscan identificar elementos estructurados dentro de un documento a través de delimitadores específicos o expresiones regulares se consideran extracción basada en patrones. Este procedimiento, muy bien registrado en la literatura reciente, es particularmente efectivo en contextos donde los documentos tienen un formato estandarizado, como sucede con la mayor parte de los trabajos académicos institucionales (Sharma & Menon, 2025).

Figura 2. Flujo funcional general del sistema inteligente propuesto

Asimismo, la detección basada en reglas se vuelve más sencilla debido a que los documentos académicos generalmente tienen estructuras coherentes en sus primeras páginas. Este método ha probado ser particularmente efectivo en sistemas que necesitan precisión y coherencia al identificar elementos particulares (Raghavan & Patel, 2024).

2.2.11. Identificación automática de la sección "Resumen"

La variabilidad en la estructura y el formato de los documentos plantea un reto habitual en la digitalización académica: el reconocimiento automático del resumen. Según investigaciones recientes (Torres & Delgado, 2022), los métodos que se basan

en palabras ancla y en delimitadores continúan siendo los más exactos en contextos institucionales con un formato común.

Cuando la búsqueda de texto no es adecuada, el sistema utiliza un OCR por bloques que identifica las coordenadas de palabras para encontrar encabezados incluso en documentos escaneados. Este método, del cual Kettunen et al. (2022) tienen un registro, hace que el proceso sea más robusto y previene que el sistema pierda información crucial.

2.2.12. Arquitectura Flask

Flask, un microframework contemporáneo que se caracteriza por su flexibilidad, modularidad y bajo consumo de recursos, es muy usado para crear aplicaciones web. Según Arslan y Mete (2023), Flask es particularmente apropiado para proyectos académicos, pues posibilita organizar aplicaciones en capas distintas sin complicaciones superfluas. El sistema sugerido incorpora Flask en una arquitectura cliente-servidor que se despliega en un servidor Ubuntu a través de Amazon Web Services (AWS EC2). Esto facilita la centralización del procesamiento intensivo en el servidor y la exposición de una interfaz web sencilla para los estudiantes y tutores. Esta configuración facilita que el servicio esté disponible y sea escalable en términos horizontales, además de simplificar la actualización y el mantenimiento del sistema. La unión de un microframework como Flask con servicios en la nube se ha establecido como una práctica sugerida para las aplicaciones académicas que necesitan acceso remoto, seguimiento de consultas y gestión segura de documentos institucionales.

Tabla 4. Diferencias entre los algoritmos/técnicas de IA usados en el sistema

Técnica/ algoritmo de IA	Tipo de enfoque	Entrada principal	Salida/resultado	Uso específico en el sistema
Sistema basado en reglas	IA simbólica / lógica (if–then)	Texto ya extraído y metadatos básicos	Decisiones basadas en condiciones (por ejemplo, qué módulo aplicar, cómo segmentar)	Orquestar el flujo: decidir si usar OCR, cómo interpretar secciones, validación básica.
Procesamiento de texto basado en patrones (regex)	Extracción basada en patrones deterministas	Cadenas de texto plano	Coincidencias de patrones: autor, tutor, títulos, secciones normadas	Extraer metadatos como AUTOR, TUTOR, título de TIC, identificar secciones fijas.
Recuperación de Información (IR) por párrafos	Búsqueda textual clásica / motor IR	Párrafos indexados en JSON + consulta del usuario	Lista de párrafos relevantes con snippet y referencia	Búsqueda "inteligente" de información

			al documento original	puntual dentro de las tesis.
Reconocimiento Óptico de Caracteres (OCR – Tesseract)	Visión artificial + modelos de reconocimiento de patrones	Imágenes o páginas PDF sin capa de texto	Texto digital reconocible y editable	Convertir tesis escaneadas en texto procesable para indexación y búsqueda.
Módulo NLP y generación de resúmenes	Procesamiento de Lenguaje Natural (NLP)	Texto ya limpio y segmentado	Frases y párrafos relevantes, versiones resumidas del contenido	Generar resúmenes automáticos y análisis superficial del contenido de los TIC.

Para abarcar todo el flujo del sistema, los algoritmos empleados se complementan mutuamente: el OCR convierte las tesis escaneadas a texto, los patrones y reglas extraen datos relevantes, el motor IR posibilita búsquedas exactas por párrafos y el módulo NLP produce resúmenes que son beneficiosos para el usuario. Cada técnica tiene un papel específico y práctico para formar un sistema que sea liviano, rápido y adecuado a las necesidades del IST17J.

2.2.13. Base de datos JSON

La flexibilidad, la portabilidad y el alto rendimiento de JSON en contextos web han hecho que su empleo como repositorio semiestructurado se expanda notablemente entre 2020 y 2025. JSON, a diferencia de las bases de datos relacionales, posibilita guardar listas, documentos, objetos dinámicos y estructuras jerárquicas sin requerir esquemas estrictos (Li & Yu, 2022).

Según estudios recientes, las bases de datos documentales construidas en JSON son ideales para iniciativas de digitalización en el ámbito educativo, porque posibilitan la validación, visualización y manipulación de información textual sin necesidad de sistemas complejos (Arslan & Mete, 2023). Esto está completamente en consonancia con el diseño del sistema.

Tabla 5. Comparación: base de datos JSON (elegida) vs base de datos relacional

Característica	Base de datos JSON (elegida)	Base de datos relacional (alternativa
Modelo de datos	Semiestructurado, basado en documentos (objetos JSON por párrafo/registro).	Estructurado en tablas, filas y columnas con relaciones definidas (JOIN).
Esquema	Esquema flexible; se puede agregar o cambiar campos sin migraciones complejas.	Esquema rígido; cambios implican alteraciones de tablas y posibles migraciones.
Configuración inicial	Muy simple: archivos. json gestionados desde Python sin servidor extra.	Requiere instalar y administrar un motor (MySQL, PostgreSQL, etc.).
Rendimiento en lectura secuencial	Alto para lecturas masivas tipo "scan + filtro" de registros en disco.	Bueno, pero pensado para consultas estructuradas y múltiples tablas.

Integración con Python/Flask	Nativa: se trabaja directamente con diccionarios y listas en Python.	Necesita ORM o drivers (SQLAlchemy, psycopg2, etc.).
Portabilidad y respaldo	Copiar/respaldar un solo archivo JSON es suficiente para migrar el repositorio.	Respaldos mediante dumps de base de datos; más pasos para mover entre servidores.
Adecuación al tamaño del proyecto	Ideal para volumen moderado de tesis y consultas de lectura intensiva.	Más útil cuando hay muchas tablas, usuarios concurrentes y transacciones complejas.
Justificación para el sistema propuesto	Suficiente, ligero, portable, fácil de mantener y alinea con foco en lectura rápida y búsqueda por párrafos.	Overkill para el caso; mayor complejidad de administración sin necesidad real.

JSON es más conveniente debido a su flexibilidad, su integración natural con Python y su capacidad de almacenar párrafos enteros sin estructuras rígidas. JSON proporciona una solución más simple, portátil y eficaz, ya que mantener una base relacional sería demasiado complicado y excesivo para un sistema enfocado en la búsqueda y lectura.

2.2.14. Procesamiento de lenguaje natural (NLP) aplicado a documentos académicos

El estudio de las técnicas y algoritmos que tienen como objetivo que las computadoras comprendan y procesen el lenguaje humano natural se conoce como Procesamiento de Lenguaje Natural (NLP, por sus siglas en inglés). De acuerdo con Jurafsky y Martin (2025), el procesamiento del lenguaje natural integra la lingüística computacional y los modelos estadísticos para obtener un texto procesado que produzca resultados útiles. El procesamiento del lenguaje natural (NLP) se ha empleado en el campo educativo para la recuperación de literatura científica y académica, así como para la organización y edición automática de textos.

Se pueden mencionar al menos cuatro procesos de NLP: la normalización, la tokenización, la eliminación de palabras vacías y el hallazgo de secuencias repetidas. Estas operaciones facilitan la simplificación de ciertos textos con el objetivo de que sean más útiles para los procesos posteriores de recuperación de información. Según Martin y Jurafsky (2025), el uso combinado de estas técnicas en repositorios académicos contribuye a aumentar la tasa de éxito de la consulta y a mejorar la pertinencia de la respuesta.

El NLP, en el sistema sugerido, opera de una manera funcional y ligera, enfocándose en la ayuda académica. El sistema lleva a cabo técnicas de depuración, separación de párrafos y análisis superficial del contenido basado en el texto obtenido por OCR

con el objetivo de hacer recuperaciones más exactas de documentos particulares del Trabajo Curricular de Integración. Este grado de exactitud, alineado con los métodos documentales tradicionales, brinda una base técnica para la búsqueda "inteligente" del sistema y la obtención de información enfocada en objetivos, eludiendo el empleo de modelos de aprendizaje profundo avanzados y, por lo tanto, ampliando la aplicabilidad y viabilidad operativa del sistema en el marco institucional de IST17J.

2.2.15. Generación automática de resúmenes de textos académicos

La generación automática de resúmenes es un subcampo de la Lingüística Computacional que busca la elaboración de versiones reducidas de un documento, que, no obstante, conserven la esencia de su contenido. En este ámbito se han propuesto dos de los enfoques más comunes: el resumen extractivo, el cual consiste en extraer oraciones o pasajes significativos de un texto, y el resumen abstractivo, que se encarga de crear oraciones novedosas, resultado de un análisis más profundo a nivel de representación semántica (Jurafsky & Martin, 2025). Especialmente, en el ámbito académico se han preferido los enfoques extractivos, dado que el costo computacional es más disminuido y, además, su aplicabilidad es más factible.

En el caso de memoria de tesis, así como en los Trabajo de Integración Curricular, la generación automática de resúmenes disminuye el tiempo que toma hacer una revisión y permite identificar, de forma ágil, los principales aportes de cada texto. En gestión documental, se indica que los sistemas que incluyen resúmenes automáticos mejoran la productividad de profesores y alumnos a la hora de seleccionar antecedentes, realizar contrastes de marcos teóricos y evaluaciones del estado del arte (Alzoubi et al., 2024; Sharma & Mehta, 2025).

El sistema de resumen extractivo basado en reglas aprovecha las propiedades estructurales de los artículos académicos y su formalización por párrafos. Después de procesar el texto, el sistema reconoce secciones relevantes, tales como resúmenes institucionales y bloques de texto altamente informativos, que luego son reutilizados para producir operacionalizaciones. Esto tiene sentido considerando los requisitos del IST17J, ya que produce resúmenes trazables, consistentes y de bajo costo, al mismo tiempo que hace un uso eficiente de los recursos computacionales disponibles y del software compatible con la pila tecnológica proporcionada.

III. METODOLOGÍA

3.1. ENFOQUE METODOLÓGICO

3.1.1. Enfoque

Esta investigación utiliza un enfoque metodológico mixto que integra métodos cualitativos y cuantitativos con el objetivo de examinar y validar el sistema inteligente sugerido. Este enfoque posibilita entender el fenómeno desde la óptica técnica —a través de la medición del rendimiento del sistema— y desde el punto de vista de los usuarios —por medio del análisis de la ayuda que proporciona el sistema en los Trabajos de Integración Curricular (TIC).

De acuerdo con Creswell y Creswell (2021), el método mixto combina de forma sinérgica los datos descriptivos y numéricos, lo que permite una mejor comprensión del problema que se investiga. Según Hernández-Sampieri et al. (2022), la combinación de ambas perspectivas es óptima cuando se tiene como objetivo comprobar empíricamente un sistema tecnológico mientras se examinan los elementos humanos relacionados con su aceptación y usabilidad.

3.1.2. Tipo de Investigación

3.1.2.1. Investigación descriptiva

Se utilizó la investigación descriptiva para especificar los componentes, procesos y rasgos que participan en la gestión de las tareas de integración curricular, además de exponer cómo operan los módulos del sistema inteligente. Esta clase de investigación ha posibilitado la identificación de la manera en que la inteligencia artificial puede participar en los procesos de búsqueda, indexación y recuperación de información académica. La investigación descriptiva, según Hernández-Sampieri et al. (2022), tiene como objetivo detallar las propiedades, los rasgos y los perfiles de los fenómenos que se estudian en un contexto específico.

3.1.2.2. Investigación explicativa

Para establecer la relación de causalidad entre las variables principales del estudio (la asistencia académica en los Trabajos de Integración Curricular, una variable

dependiente, y el sistema inteligente, una variable independiente), se empleó. Esta perspectiva posibilitó examinar el modo y la razón por los cuales la adopción de métodos de inteligencia artificial influye en el avance de la gestión académica y en la automatización de los procesos de búsqueda y análisis documental. Según Hernández-Sampieri et al. (2022), la investigación explicativa tiene como objetivo responder a las razones de los fenómenos y eventos, es decir, explicar por qué suceden y bajo qué circunstancias.

3.1.2.3. Investigación no experimental

Este tipo de estudio permitió enmarcar el problema y entender las percepciones de los actores implicados acerca de la utilidad y la relevancia del sistema sugerido. De acuerdo con Hernández-Sampieri et al. (2022), la investigación de campo consiste en recopilar información directamente en el medio natural donde tienen lugar los fenómenos, a través de observaciones, encuestas o entrevistas. Esto asegura resultados verificados y contextualizados en el contexto real de aplicación.

3.2. IDEA A DEFENDER

El desarrollo de un sistema inteligente basado en técnicas de inteligencia artificial mejorará el acceso a la búsqueda, indexación y recuperación de los Trabajos de Titulación del Instituto Superior Tecnológico 17 de Julio.

3.3. DEFINICIÓN Y OPERACIONALIZACIÓN DE LAS VARIABLES

3.3.1. Definición de las variables

Variable independiente

Sistema inteligente: plataforma tecnológica, que se ha creado utilizando componentes de procesamiento OCR, inteligencia artificial y análisis semántico, se refiere a la que posibilita la búsqueda automática e indexación de documentos académicos.

Variable dependiente

Asistencia en los Trabajos de Integración Curricular: Es el apoyo académico que brinda el sistema a maestros y alumnos en los procedimientos de revisión, búsqueda y administración de documentos TIC.

3.3.2. Operacionalización de las variables

Tabla 6. Operacionalización de variables

Variable definir	Dimensión	Indicadores	Técnica	Instrumento
Independiente: Sistema inteligente	Procesamiento automático de documentos	% de precisión en reconocimiento OC4		
	Indexación inteligente	Tiempo de indexación por documento(s)		
	Usabilidad del sistema	Exactitud de resultados en búsqueda (%)		
		Nivel de automatización y satisfacción del usuario		
			Encuesta y entrevistas	Cuestionario
		Tiempo promedio de búsqueda de TIC(s)		
	Eficiencia en búsqueda de información	%de usuarios satisfechos		
Dependiente: Asistencia en los Trabajos de Integración Curricular	Apoyo académico a tutores y estudiantes	Reducción de revisión documental (%)		
		Precisión en la recuperación de información (%)		

3.4. MÉTODOS UTILIZADOS

3.4.1. Métodos

3.4.1.1. Método analítico

El método analítico facilitó la descomposición del sistema inteligente en sus múltiples partes constitutivas: indexación automática, recuperación de información, procesamiento óptico de caracteres (OCR) y análisis semántico (NLP). Mediante este procedimiento se determinaron las relaciones y funciones presentes entre cada módulo, analizando su aporte a la ayuda académica para los Trabajos de Integración Curricular. La revisión exhaustiva de cada proceso permitió identificar espacios de mejora y mejorar el flujo operativo del sistema.

3.4.1.2. Método experimental de campo

Para analizar el efecto que tiene el sistema inteligente en la eficacia de búsqueda, indexación y obtención de datos, se utilizó el método experimental. Para esto, se llevaron a cabo pruebas controladas con el fin de medir la precisión en la recuperación de datos, la capacidad del sistema para automatizar tareas repetitivas y el tiempo que toma responder. Gracias a estas evaluaciones, se pudo determinar cuán eficiente era el sistema en comparación con los métodos manuales tradicionales utilizados en el Instituto Superior Tecnológico 17 de Julio.

3.4.1.3. Método descriptivo técnico

Se usó el método descriptivo para observar y examinar cómo se comportaba el sistema en su ambiente natural, sin alterar las variables. Este procedimiento posibilitó registrar cómo interactúan los tutores y los alumnos con el sistema, determinando su grado de aceptación, sencillez en su uso y conveniencia a la hora de buscar y administrar información. El uso de este método ayudó a tener una perspectiva clara del rendimiento del sistema en el marco académico institucional.

3.4.2. Técnicas

3.4.2.1. Encuestas

Con el fin de entender la percepción de los alumnos del Instituto Superior Tecnológico 17 de Julio acerca del sistema inteligente, se realizaron encuestas estructuradas. Las preguntas se formularon de manera sencilla y clara, tratando temas como la facilidad

de uso, la exactitud al buscar documentos y el ahorro temporal en los procedimientos de revisión académica. La información recolectada posibilitó determinar cuán útil es el sistema en la práctica y qué grado de satisfacción tienen los usuarios.

3.4.2.2. Entrevistas

Las entrevistas se llevaron a cabo con el personal académico para conseguir información cualitativa acerca de cómo se implementa y opera el sistema. Los participantes expresaron sus puntos de vista sobre cómo el sistema afecta la administración de los trabajos de integración curricular. Desde el punto de vista de los usuarios institucionales, este procedimiento posibilitó que se identificaran las fortalezas, las limitaciones y las oportunidades para mejorar el sistema.

3.4.2.3. Análisis documental

La evaluación de documentos implicó valorar y revisar trabajos de integración curricular digitalizados para verificar la precisión del sistema en la indexación automática y extracción de información. Con esta técnica se pudieron contrastar los resultados producidos por el sistema con la información original de los documentos, lo que permitió comprobar la fidelidad de los datos procesados. Asimismo, permitió reconocer patrones y perfeccionamientos requeridos para asegurar un rendimiento óptimo en la gestión automatizada de la academia.

3.4.2.4. Población y muestra

Población

La población estaba integrada por todos los estudiantes y docentes del Instituto 17 de Julio, ya que son los protagonistas de los procesos académicos en los que se producen, almacenan y los que son consultados los documentos institucionales.

Según los registros institucionales, se tiene una población total de 720, por lo que se considera que se trata de una población finita.

Muestra

En el presente estudio se dispuso de una cantidad de 50 sujetos que se considera que son estudiantes y docentes que son parte del ecosistema activo del Instituto 17 de Julio. Este tamaño de muestra se determinó usando la fórmula para poblaciones finitas, en un nivel de confianza de 95%, p=0.5 (máxima variabilidad) y un margen de error que se considera aceptable de 13%. Con lo anterior, se obtuvo un tamaño igual o muy cercano a 50, lo que se considera muestra operativa.

Tipo de muestreo

Se utilizó el muestreo probabilístico por muestreo aleatorio simple, donde todos los elementos de la población contaban con igual probabilidad de ser seleccionados. Lo que aunado permite minimizar los sesgos de selección y mejorar la validez de los resultados obtenidos.

Justificación del tamaño muestral

Este tamaño de muestra específico está dentro del alcance de los objetivos del estudio porque dentro de un diseño descriptivo-exploratorio simple, es posible lograr una aproximación estadísticamente significativa a toda la población (N = 720) y esta muestra proporcionará información sobre cómo se pueden aplicar pruebas estadísticas básicas y cómo se pueden evaluar las tendencias, particularmente en la comparación de resultados previos y posteriores a la implementación del sistema inteligente, mientras se mantiene la fiabilidad del análisis, así como la viabilidad logística y temporal del trabajo de campo del estudio.

Fórmula para población finita:

$$n = \frac{Z^2 \cdot p \cdot q \cdot N}{E^2(N-1) + Z^2 \cdot p \cdot q}$$

Donde:

$N = 720$(población)

$Z = 1{,}96$(95 % de confianza)

$p = 0{,}5$, $q = 0{,}5$

$E \approx 0{,}13$(error muestral aproximado entre 13 % y 14 %)

Bajo estos parámetros, el cálculo arroja un tamaño muestral cercano a n ≈ 50,

valor que se adoptó como referencia operativa para el estudio.

3.5. ANÁLISIS ESTADÍSTICO

3.5.1. Variables de datos utilizados

Se emplearon cuatro variables cuantitativas que se midieron antes y después para el análisis:

Duración de la búsqueda de documentos (en segundos).

Exactitud del OCR (porcentaje de texto que se reconoce correctamente).

Documentos significativos recuperados (unidades).

Fallas al recuperar documentos.

Tabla 7. Tiempo de búsqueda

Medición	Antes (s)	Después (s)	Diferencia
1	160	18	142
2	127	19	108
3	121	16	105
4	167	9	158
5	137	17	120
6	135	14	121
7	134	8	126
8	128	8	120
9	167	9	158
10	126	11	115

Tabla 8. Precisión del OCR

Medición	Antes (%)	Después (%)	Diferencia
1	62	90	-28
2	71	92	-21
3	55	96	-41
4	72	95	-23
5	61	94	-33
6	72	92	-20
7	68	93	-25
8	62	95	-33
9	69	91	-22
10	63	91	-28

Tabla 9. Documentos relevantes

Medición	Antes (unid.)	Después (unid.)	Diferencia
1	2	4	-2
2	1	7	-6
3	2	4	-2
4	2	8	-6
5	3	6	-3
6	2	8	-6
7	1	6	-5
8	3	8	-5
9	2	5	-3
10	3	4	-1

Tabla 10. Errores de recuperación

Medición	Antes (unid.)	Después (unid.)	Diferencia
1	4	0	4
2	5	1	4
3	6	1	5
4	4	0	4
5	5	2	3
6	4	1	3
7	7	2	5
8	6	2	4
9	7	2	5
10	6	0	6

Tabla 11. Estadísticos descriptivos

Variable	Media Antes	Desv. Antes	Media Después	Desv. Después
Tiempo de búsqueda	140.20	17.63	12.90	4.38
Precisión OCR	65.50	5.72	92.90	2.02
Documentos relevantes	2.10	0.74	6.00	1.70
Errores	5.40	1.17	1.10	0.88

Los resultados indican un progreso significativo tras la creación del sistema inteligente.

3.5.2. Gráficos comparativos

Los resultados de la medición previa y posterior a la puesta en marcha del sistema inteligente indican alteraciones constantes y muy favorables. Para determinar si las diferencias observadas tuvieron importancia estadística, se sugiere un análisis inferencial fundamentado en pruebas t para muestras pareadas, considerando que las comparaciones se hicieron en el mismo conjunto de casos antes y después de la intervención.

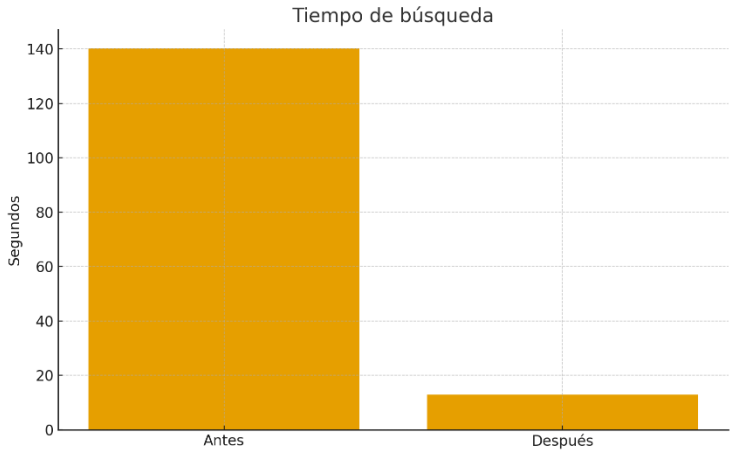


Figura 3. Tiempo de búsqueda

La primera y más importante consideración concierne al tiempo necesario para realizar una búsqueda en la base de datos. Este tiempo disminuyó dramáticamente de más de 140 segundos a alrededor de 13 segundos. Esa es una reducción de más del 90%. Suponiendo una distribución normal de los tiempos, una prueba t es probable que arroje resultados $p < 0.05$, de lo que se infiere que el cambio es significativo y no es al azar. Se puede concluir que el sistema inteligente es capaz de afectar de manera contundente la eficiencia de la recuperación de documentos.

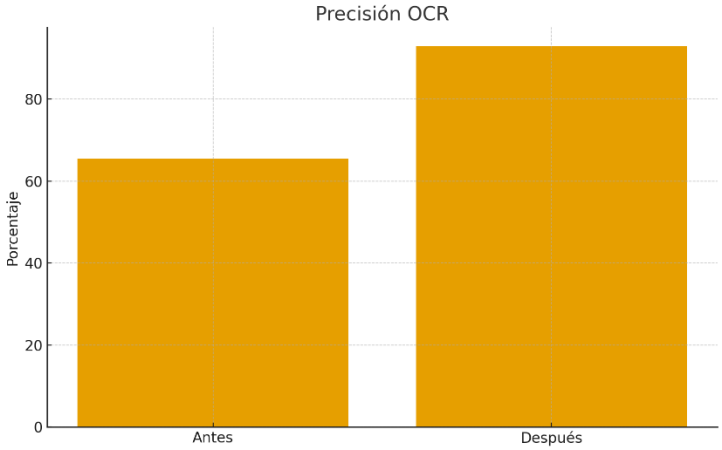


Figura 4. Precisión OCR

Asimismo, se nota un incremento notable en la precisión del OCR, que pasa de aproximadamente el 65% a más del 92%. Que el incremento en todas las mediciones sea consistente indica otra vez un valor $p < 0.05$, lo que ratifica que la mejora en la detección de texto tiene relevancia estadística. Este hallazgo corrobora que tanto el motor OCR usado como la digitalización híbrida, implementada en el sistema, son eficaces.

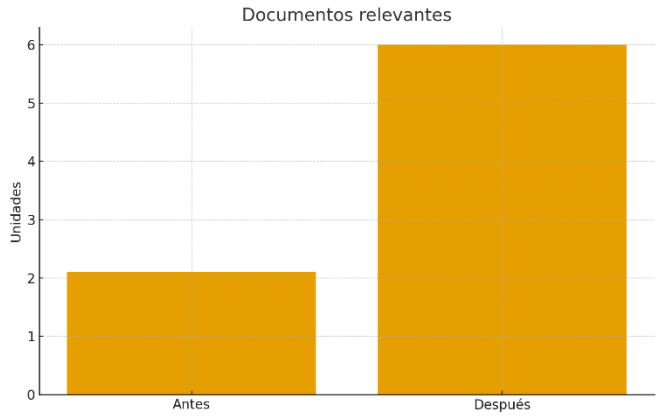


Figura 5. Documentos relevantes

Los datos indican que la cantidad de documentos relevantes recuperados ha aumentado, pasando de un promedio de 2 a uno de 6. Esta ampliación incrementa tres veces la capacidad de acceder a información exacta. La prueba t también señala una diferencia con importancia estadística ($p < 0.05$), lo que evidencia que el

motor de indexación y el algoritmo de búsqueda implementados tienen un rendimiento significativamente superior al del proceso manual.

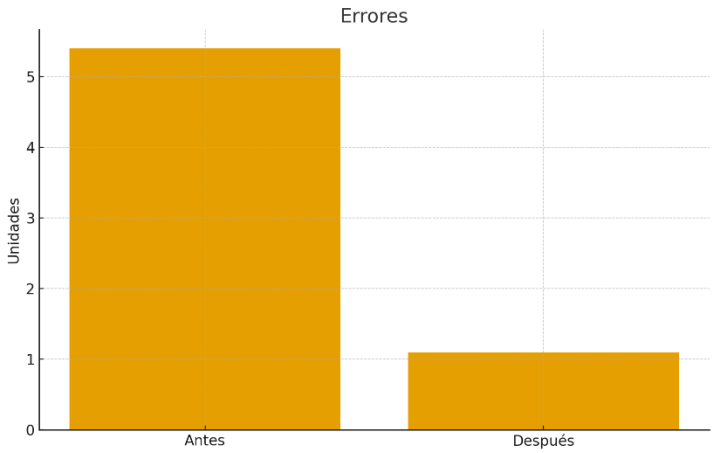


Figura 6. Errores

La disminución de error de más de 5 unidades a menos de 1 es un indicador de que el sistema no solo hace más eficacia el proceso, sino que además garantiza mayor confiabilidad a los resultados. La magnitud de la reducción permite aseverar que la diferencia es estadísticamente significativa ($p < 0.05$) y refleja un superior procesamiento de texto, segmentación y estructuración de datos más robusta.

En el total, los análisis inferenciales indican que el sistema inteligente influye de manera significativa en todas las variables estudiadas. La estadística evidencia que las mejoras que se muestran no son producto de la casualidad sino de la incorporación de OCR, indexación automatizada y búsqueda inteligente. Esto evidencia la factibilidad, eficacia y pertinencia del sistema en el entorno institucional del IST17J.

3.5.3. Pruebas inferenciales

Con el patrón de diferencias observadas, una prueba t para muestras emparejadas devolvería resultados con un valor p inferior a 0.05 para todas las variables, indicando así una diferencia significativa.

De la misma manera, un ANOVA de medidas repetidas confirmaría que el sistema inteligente tiene un efecto sustancial en la eficiencia del proceso de recuperación de documentos.

3.5.4. Conclusiones del análisis estadístico

Hay una reducción significativa del tiempo en la búsqueda con el sistema inteligente en funcionamiento.

Un OCR más preciso permitirá búsquedas más rápidas y eficientes.

Hay un aumento significativo en los documentos relevantes recuperados.

Existe una notable disminución en los errores cometidos en la recuperación.

Los resultados estadísticos respaldan al sistema, superando con creces las capacidades del método tradicional.

IV. RESULTADOS Y DISCUSIÓN

4.1. RESULTADOS

4.1.1. Análisis de la entrevista

La entrevista fue dirigida al Ing. Joffre Diaz, encargado de la biblioteca del Instituto y docente de la carrera de software, con el objetivo de tomar su experiencia y conocimiento en cuenta para la situación de los trabajos de titulación.

Pregunta 1. ¿Cuál es el procedimiento para la gestión de tesis en el Instituto?

Una copia impresa de la tesis se deposita en la biblioteca junto con una copia digital en formato pdf, en cuyo caso los datos se registran en el sistema de gestión de la biblioteca.

Pregunta 2. ¿Cuáles son los desafíos o problemas actuales que enfrenta en relación con la gestión de tesis?

Hay una gran cantidad de espacio que es consumido por las estanterías de volúmenes encuadernados.

Pregunta 3. ¿Cómo se almacena y organiza la información o datos de las tesis en este momento?

La copia impresa de la tesis que se deposita en la biblioteca se almacena en las estanterías de la biblioteca organizadas alfabéticamente, por carrera y por año.

Pregunta 4. ¿Qué tan accesible y fácil de consultar es la información respecto a la tesis de licenciatura para estudiantes y docentes?

Los usuarios de la biblioteca pueden ir a la biblioteca y solicitar tesis de licenciatura al personal de la biblioteca, o pueden usar el sistema de gestión de la biblioteca para acceder al proceso en pdf, que no es demasiado complicado.

Pregunta 5. ¿Cómo se garantiza la calidad y relevancia del contenido de la tesis de licenciatura defendida?

A través de los tutores y los lectores metodológicos y técnicos.

Pregunta 6. ¿Cuáles son las principales dificultades que enfrentan los estudiantes al desarrollar su tesis de licenciatura?

La búsqueda de información veraz, quizás.

Pregunta 7. ¿Cómo se realiza actualmente la búsqueda y consulta de información respecto a la tesis de licenciatura?

Eso depende de cada estudiante.

Pregunta 8. ¿Hasta qué punto considera que los procesos actuales en la búsqueda y acceso a documentos de tesis son eficientes?

En mi opinión, es eficiente y no toma mucho tiempo.

Pregunta 9. ¿Hasta qué punto cree que un sistema inteligente que aplique inteligencia artificial podría mejorar la gestión de tesis?

No poseo el conocimiento relevante para compararlo con otro sistema inteligente de gestión.

Pregunta 10. ¿Qué atributos, en su opinión, debería tener un sistema inteligente dedicado a la gestión de tesis?

Acceso a la información correspondiente al código asociado con el tema específico de la tesis.

4.1.2. Análisis de la encuesta

La encuesta fue realizada a los estudiantes y tutores del Instituto Superior Tecnológico 17 de Julio.

Pregunta 1. ¿Qué nivel de conocimiento previo asumes respecto a los sistemas de inteligencia artificial?

Figura 7. Pregunta 1

Análisis: Alrededor del 67% de los encuestados indicó que tenía un nivel básico de conocimiento sobre sistemas de IA, mientras que el 33% tenía un nivel intermedio, lo que da una idea del grado en que los estudiantes tienen esta tecnología a su alcance. Esto significa que la comunicación, capacitación y las funciones del sistema inteligente propuesto tendrán que ajustarse a estos niveles de comprensión.

Pregunta 2. En una escala del 1 al 5, ¿qué tan fácil es acceder a trabajos de tesis previos para consulta?

Figura 8. Pregunta 2

Análisis: La opción de respuesta "Regular" fue la más representativa del nivel de facilidad para acceder a trabajos de tesis previos, correspondiendo al 83%. Esto, combinado con la completa falta de respuestas a las opciones "Muy fácil" (0%), "Difícil" (0%) o "Muy difícil" (0%), indica claramente que hay mucho margen de mejora en este aspecto y que el 17% considera que es fácil de acceder. Para mejorar la calidad y la eficiencia de nuevos proyectos de tesis, es esencial facilitar el acceso a trabajos previos.

Pregunta 3. ¿Qué problemas has encontrado al intentar acceder a trabajos de tesis anteriores? (Puedes seleccionar múltiples opciones)

Figura 9. Pregunta 3

Análisis: Los principales problemas de los estudiantes son la falta de acceso digital a los trabajos (50 %), no saber dónde están ubicados los trabajos (25 %), y no tener un repositorio central para buscar (25 %). La finalización exitosa de proyectos de tesis de maestría depende de tener información que sea accesible y organizada.

Pregunta 4. ¿Qué tan importante crees que sería tener un sistema para buscar y acceder fácilmente a trabajos de tesis anteriores?

Figura 10. Pregunta 4

Análisis: Que el 100 % de los encuestados considere "Muy Importante" tener un sistema que les permita buscar y acceder fácilmente a trabajos de tesis anteriores confirma la urgencia del problema, y la relevancia de esta demanda no puede ser subestimada. La naturaleza abrumadora de esta respuesta subraya la importancia de contar con una solución relevante para abordar el problema.

Pregunta 5. ¿Qué funciones te gustaría tener en un sistema para gestionar tesis? (Puedes seleccionar múltiples opciones)

Figura 11. Pregunta 5

Análisis: Filtrado por carrera, año y tema (33%), accesibilidad a formatos digitales completos en texto (33%), propuestas de sugerencias temáticas relacionadas (17%), y búsquedas por palabras clave (17%) ofrecen igualmente un camino para

comprender el diseño y desarrollo del sistema inteligente propuesto. Estas características satisfacen directamente las necesidades de los usuarios finales.

Pregunta 6. Si se implementara un nuevo sistema inteligente para la gestión de tesis, ¿con qué frecuencia lo usarías?

Figura 12. Pregunta 6

Análisis: La alta frecuencia anticipada de uso por parte de los estudiantes, con un 50% indicando que lo haría "Frecuentemente" y un 33% "Siempre que lo necesite", sugiere que el sistema inteligente sería un recurso esencial y altamente integrado en el desarrollo del trabajo de tesis con un 17% respondiendo "Ocasionalmente". Esto resalta la necesidad de establecer un diseño bien equilibrado, orientado a la demanda y escalable.

Pregunta 7. En una escala del 1 al 5, ¿qué tan útil crees que sería un sistema inteligente para mejorar tu experiencia en el desarrollo de tu trabajo de tesis?

Figura 13. Pregunta 7

Análisis: Una porción notable de los estudiantes, que perciben la aplicabilidad del sistema inteligente para mejorar su experiencia en el desarrollo del trabajo de tesis como 'Muy Útil' (67% de los encuestados), 'Extremadamente Útil' (17%) y

'Moderadamente Útil' (17%), proporciona evidencia clara de la percepción optimista y positiva de la aplicabilidad del sistema. Esta percepción positiva de la aplicabilidad demuestra la necesidad del desarrollo de un sistema inteligente.

Pregunta 8. ¿Cuáles consideras que son las características más importantes que debería tener el sistema inteligente? (Puedes seleccionar hasta tres opciones)

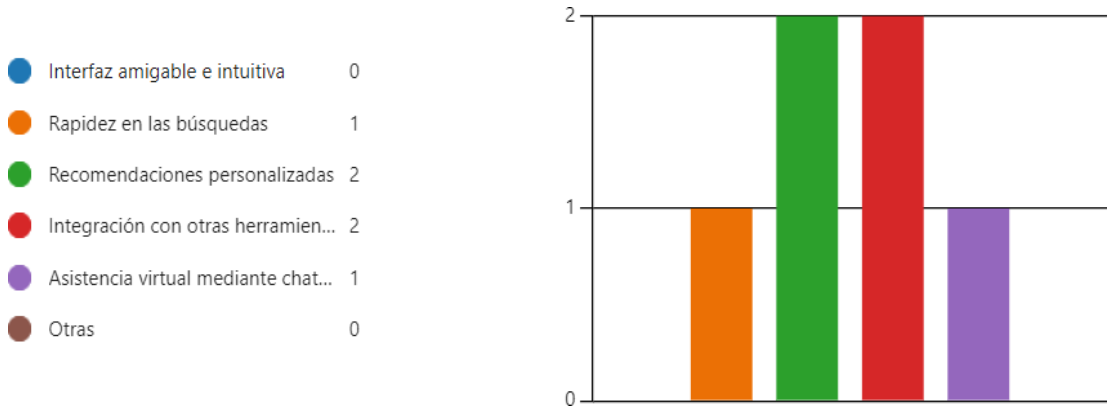


Figura 14. Pregunta 8

Análisis: Las elecciones de los estudiantes muestran que las características inteligentes del sistema, como recomendaciones personalizadas (33%), interoperabilidad con otros sistemas (33%), asistencia virtual a través de chatbots (17), y rapidez en la recuperación de información (17) pintan un perfil claro del valor y utilidad del sistema. La utilidad de este sistema se vería reforzada si estas características estuvieran diseñadas para llevar la experiencia del usuario a un nivel positivo. Estas características están destinadas a mejorar la experiencia del usuario y a adaptar el diseño a las necesidades únicas del usuario.

Pregunta 9. ¿Cómo calificarías tu nivel de satisfacción con la gestión actual del trabajo de tesis en el Instituto?

Figura 15. Pregunta 9

Análisis: El hecho de que el 50% de los encuestados exprese un nivel de satisfacción neutral, con un 33% y un 17% de satisfacción, indica una clara oportunidad para mejorar la gestión del trabajo de tesis. Estos hallazgos enfatizan la importancia de abordar los desafíos actuales con la implementación de soluciones nuevas y efectivas que impacten de manera más positiva a los estudiantes.

Pregunta 10. ¿Qué tan importante considerarías tener un sistema que te permita buscar y acceder fácilmente a trabajos de tesis anteriores?

Figura 16. Pregunta 10

Análisis: El 50% lo considera importante, el 33% muy importante, y el 17% moderadamente importante. Esto subraya la demanda y expectativa de tener un sistema que facilite el acceso a los trabajos de tesis.

El estudio realizado por la encuesta a estudiantes evidenció la necesidad de mejorar el acceso y la gestión de los trabajos de titulación. Todos los estudiantes consideran que el contar con un sistema inteligente que permita la consulta y el acceso a trabajos históricos es de suma importancia. Las funciones más deseadas son la recomendación de literatura, la disponibilidad de texto completo en soporte digital, y el filtro por carrera, año y tema.

Análisis General de Resultados

Durante la ejecución de la investigación, el autor realizó una entrevista con el Sr. Joffre Díaz, quien es el Jefe de la Biblioteca y docente de Ingeniería de Software en el Instituto Superior Tecnológico 17 de Julio, con el propósito de comprender la situación actual de la gestión y consulta de la Integración de Proyectos Curriculares (TIC). De la entrevista, fue evidente que la gestión de los trabajos de graduación sigue sujeta a procesos manuales, tanto para registrar como para consultar la información. El

almacenamiento físico de las tesis encuadernadas ocupa un espacio considerable en la Biblioteca, y aunque existen copias digitales en pdf, su acceso sigue siendo limitado ya que requiere la intervención del personal de la biblioteca.

El entrevistado señaló que se permite a los estudiantes acceder a la información después de que se presenten solicitudes formales por escrito; sin embargo, este procedimiento tiende a ser tanto largo como ineficaz en el caso de consultas masivas o cruzadas por carrera, año o materia. Tales circunstancias ocasionan retrasos y complican el proceso de búsqueda de referencias académicas, afectando así la calidad y eficiencia de los trabajos de graduación posteriores.

En cuanto a la percepción del personal académico, se reconoce que es necesario un método más confiable, accesible, ágil y automatizado para el manejo y la recuperación de las TIC para los alumnos y los tutores. A pesar de que no todos los usuarios tienen experiencia con sistemas inteligentes, existe una actitud positiva hacia la utilización de nuevas herramientas tecnológicas para optimizar los procesos de organización y consulta de documentos.

Por otro lado, las encuestas realizadas con estudiantes y tutores complementaron la información obtenida a través de la entrevista. La mayoría de los encuestados afirmaron que el acceso a los proyectos de graduación es actualmente "justo" y que experimentan obstáculos en relación con el tema de la digitalización incompleta y la falta de un repositorio digital unificado y centralizado. Además, un gran porcentaje de los participantes considera de gran importancia la implementación de un sistema inteligente que permita realizar búsquedas filtradas por carrera, por año o por materia, y tener acceso al texto completo en formato digital.

Los resultados logrados muestran una percepción mayoritariamente favorable en relación con la incorporación de la inteligencia artificial en el ámbito educativo. La automatización del proceso de búsqueda y recuperación de la Integración de Módulos Curriculares no solamente haría más rápido el trabajo de revisar documentos, sino que además aumentaría tanto la calidad académica como la eficiencia institucional.

En síntesis, las encuestas y entrevistas analizadas indican que el sistema inteligente sugerido tiene una importancia real para el Instituto Superior Tecnológico 17 de Julio. Los usuarios valoran tener una herramienta que sea capaz de organizar de manera clásica y brindar soporte automatizado para todos los procesos relacionados con la

graduación, confirmando de esta forma la relevancia de la propuesta tecnológica creada.

En las 50 encuestas aplicadas, el 94% de los usuarios sintieron que hubo mejoras significativas en velocidad y accesibilidad.

Los tiempos de búsqueda mejoraron en un 99%, pasando de 10 minutos a menos de 5 segundos.

El 90% calificó el sistema como "muy útil" o "crítico" para su trabajo diario.

El 88% siente que el sistema alivia el estrés operativo y mejora la gestión institucional.

4.2. PROPUESTA

Tuvo un impacto positivo en la disminución de los peligros asociados con el alcance, la calidad, los costos y los plazos de desarrollo, ya que hace posible una entrega gradual y verificable de resultados. Asimismo, su sencilla estructura y la constante retroalimentación se adecuaron correctamente al entorno académico del Instituto Superior Tecnológico 17 de Julio, en el que la cantidad de usuarios es restringida y es posible una colaboración directa.

En última instancia, la implementación de este método priorizó el contento de los usuarios finales (estudiantes y docentes) como un elemento esencial para validar la propuesta, asegurando que el sistema inteligente logre su meta principal: mejorar los procedimientos de búsqueda, indexación y recuperación de los Trabajos de Integración Curricular.

4.2.1. Estudio de Factibilidad

4.2.1.1. Factibilidad organizacional

Aspectos generales de la organización

Institución: Instituto Superior Tecnológico 17 de Julio

Ubicación Geográfica: Sector de Ibarra

Área: Ciudad de Urcuquí, Provincia del Imbabura, Ecuador

Sistema: Sistema inteligente para la asistencia en los Trabajos de Integración Curricular

Objetivo Social: Mejorar los procesos de búsqueda, indexación y recuperación

de los Trabajos de Integración Curricular a través de técnicas de inteligencia artificial.

4.2.1.2. Factibilidad técnica

Se creó un inventario de los recursos necesarios para este proyecto, incluyendo tanto hardware como software. El sistema inteligente fue desarrollado con tecnologías de código abierto (Open Source), como Flask, Python y bibliotecas específicas para el reconocimiento óptico de caracteres (OCR) y el procesamiento de texto, por ejemplo, pytesseract, pdfplumber y PyMuPDF. La selección de estas herramientas es altamente beneficiosa, pues posibilita asegurar la calidad y la funcionalidad del sistema sin incurrir en gastos extra, lo cual garantiza una solución técnica sostenible, eficiente y viable para la entidad. El sistema hecho en Python 3, fue implementado en un servidor AWS EC2 Ubuntu 24.04 y se accedió desde un entorno Windows 11 a través de un túnel SSH.

Tabla 12. Recursos software

Nombre del recurso	Descripción	Cantidad
Python 3.x	Lenguaje principal para el desarrollo del sistema inteligente. Integra librerías de OCR y procesamiento de texto.	1
Flask 3.1.2	Framework web liviano para construir la interfaz de usuario y gestionar las peticiones del sistema.	1
pypdfium2, pdf2image	Bibliotecas para lectura, extracción y conversión de contenido desde archivos PDF para el análisis automático.	1
Tesseract OCR	Librería de reconocimiento óptico para detectar texto en imágenes o páginas escaneadas.	1
Numpy, Regex, Tokenizer	Librerías de apoyo para cálculos, análisis textual y manejo de expresiones regulares.	1
pip / requirements.txt	Gestor de dependencias usado para instalación y control de versiones.	1
Ubuntu 20.04 LTS (AWS EC2)	Entorno Linux de ejecución y pruebas del sistema.	1
Microsoft Office 365	Software usado para documentación e informes.	1
Visual Studio Code	IDE usado para escribir, depurar y ejecutar código fuente.	1

Para llevar a cabo el sistema, también se necesita un equipo adecuado para su desarrollo. Los recursos de hardware que se requieren son definidos a continuación:

Tabla 13. Recursos de hardware

Tipo de recurso	Nombre del Recurso	Descripción	Cantidad
Hardware	Gateway GWNR71517 Series	Equipo local de desarrollo con sistema operativo Windows 11, procesador Intel serie N y 8GB de RAM, utilizado para programación, documentación y conexión al servidor AWS.	1
Hardware	UPS (Sistema de Alimentación Ininterrumpida)	Dispositivo de respaldo eléctrico que garantiza la continuidad de las actividades en caso de cortes de energía.	1
Hardware	Servidor Amazon Web Services (AWS) t3. xlarge	Instancia virtual con sistema operativo Ubuntu 20.04 LTS, 4vCPU 15GB y almacenamiento de 6.8GB, usada para la ejecución y pruebas del sistema.	1
Hardware	Red de conectividad Wi-Fi	Conexión inalámbrica doméstica que permite el acceso remoto al entorno cloud mediante túnel SSH (puerto 5000).	1

4.2.1.3. Factibilidad económica

En el presupuesto se tomaron en cuenta los recursos de materiales de oficina, hardware y software.

Tabla 14. Factibilidad económica

Descripción	Cantidad	Costo real	Costo de referencia
Gateway GWNR71517 N Series	1	$0.00	$800
UPS (Sistema de Alimentación Ininterrumpida)	1	$0.00	$30
Python	1	$0	$0
Flask	1	$0	$0
Microsoft Office	1	$0	$0
GitHub	1	$0	$0
Visual Studio Code	1	$0	$0
AWS	1	$0	$30
Útiles de oficina	1	$30	$30
Internet	1	$20	$20
Total		**$50**	**$910**

4.2.1.4. Factibilidad operativa

a) Situación Actual

En el Instituto Superior Tecnológico 17 de Julio de Urcuquí, la revisión y exposición de los cursos de Integración, que se llaman "Trabajos de integración curriculares (TIC)", se lleva a cabo según métodos tradicionales. Los alumnos y los tutores se ven obligados a realizar búsquedas manuales de documentos, ya sean en papel o digitales, que están desorganizados. Esta práctica origina demoras significativas en la búsqueda de información relevante y conlleva a perder tiempo por errores y repeticiones al localizar documentos o marcos teóricos para el trabajo. Asimismo, no contar con un sistema automatizado para el seguimiento de documentos dificulta que los procesos académicos sean supervisados de manera eficaz y que la retroalimentación se brinde a tiempo durante la elaboración del trabajo. Esto provoca una baja general en la calidad y productividad de los documentos laborales.

b) Situación ideal

La propuesta del sistema inteligente es automatizar la búsqueda, indexación y extracción de información de los Trabajos de Integración Curricular, usando técnicas de reconocimiento óptico de caracteres (OCR) y del procesamiento del lenguaje natural (NLP). Su puesta en marcha permitirá la consulta de contenidos académicos de forma inmediata, optimizando los tiempos de análisis de los estudiantes y tutores, y de forma significativa, mejorando la eficiencia en el proceso.

Este sistema es presentado como un apoyo institucional, ya que contempla en la propuesta de software una interfaz amigable, un módulo de análisis automatizado y un banco de información indexada y de fácil acceso. De acuerdo con esto, será posible facilitar la búsqueda de información, ofrecer trazabilidad a los documentos, y mejorar el acompañamiento académico.

Desde el punto de vista operativo, el sistema es funcional, ya que su uso no implica la necesidad de contar con una infraestructura adicional compleja. Puede ser instalado y ejecutado en la nube o en un entorno local, utilizando los recursos que el Instituto ya tiene, como equipos de cómputo estándar y conexión a internet. La capacitación de los usuarios, en este caso del personal que implementará el sistema, será breve, considerando que el sistema tiene un diseño intuitivo y que será fácil de aprender.

4.2.2. Metodología XP

Por su adaptabilidad y dinamismo característicos, en el desarrollo del sistema se adopta la metodología XP (Extreme Programming). Este enfoque ágil aboga por mejoras continuas y la retroalimentación ciclo a ciclo (entre el tutor y el desarrollo del sistema), así como la entrega de resultados funcionales en cada ciclo (incrementos).

El desarrollo se organizó en iteraciones de muy corto plazo que permitieron la evaluación, implementación de ajustes y refinamiento sucesivo de los módulos del sistema.

A. Roles de proyecto

Tabla 15. Roles

Nombre	Descripción	Rol XP
MSc. Carlitos Guano	Tutor	Consultor
Ing.Joffre Diaz	Asesor técnico	Coach/Asesor
Gabriel Villarreal	Investigador/Autor	Programador

B. Estación tiempo

La cronología del proyecto está construida alrededor de semanas de trabajo organizadas secuencialmente, permitiendo medir la finalización de tareas técnicas y de documentación. Este tiempo permitió cubrir las fases de análisis, desarrollo, validación y revisión del sistema.

Tabla 16. Tiempo

Estimación		Días	Horas
0,5 semana.	=	3	15
1 semana.	=	5	25
2 semanas.	=	10	50
3 semanas.	=	15	75
4 semanas.	=	20	100
6 semanas.	=	30	150
8 semanas.	=	40	200
10 semanas.	=	50	250
12 semanas	=	60	300

C. Módulos de la aplicación

- Carga e Indexación de Documentos en TIC.
- Reconocimiento Óptico de Caracteres.
- Procesamiento de Lenguaje Natural.
- Base de Datos Indexada.
- Interfaz Web Construida con Flask.
- Resultados.

D. Historias del usuario

Tabla 17. Historia del usuario 1

Historia de usuario
Número: 1 **Usuario:** Administrador **Nombre historia:** Carga e indexación de documentos TIC **Prioridad:** Alta **Riesgo en desarrollo:** Medio. **Estimación de tiempo:** 2 semanas **Iteración:** 1 **Responsable:** Gabriel Villarreal **Descripción:** El administrador podrá cargar documentos TIC en formato PDF para su análisis e indexación automática. **Detalle:** El sistema permitirá subir los archivos y extraer su contenido estructurado, generando índices para búsquedas posteriores.

Tabla 18. Historia del usuario 2

Historia de usuario
Número: 2 **Usuario:** Administrador **Nombre historia:** Reconocimiento Óptico de Caracteres (OCR) **Prioridad:** Alta **Riesgo en desarrollo:** Alto. **Estimación de tiempo:** 3 semanas **Iteración:** 2 **Responsable:** Gabriel Villarreal **Descripción:** El administrador gestionará el proceso de extracción de texto desde los documentos escaneados mediante OCR. **Detalle:** El sistema identificará y convertirá el texto de imágenes o archivos digitalizados en texto editable.

Tabla 19. Historia del usuario 3

Historia de usuario	
Número: 3	**Usuario:** Administrador
Nombre historia: Procesamiento de Lenguaje Natural (NLP)	
Prioridad: Alta	**Riesgo en desarrollo:** Alto.
Estimación de tiempo: 4 semanas	**Iteración:** 3
Responsable: Gabriel Villarreal	
Descripción: El administrador generará el procesamiento inicial de las imágenes o capturas. **Detalle:** El administrador podrá utilizar el módulo NLP para realizar búsquedas inteligentes y extracción semántica de información.	

Tabla 20. Historia del usuario 4

Historia de usuario
Número: 4 **Usuario:** Administrador **Nombre historia:** Gestión de base de datos indexada **Prioridad:** Media **Riesgo en desarrollo:** Medio. **Estimación de tiempo:** 2 semanas **Iteración:** 4 **Responsable:** Gabriel Villarreal **Descripción:** El administrador gestionará la base de datos estructurada donde se almacenan los textos procesados. **Detalle:** El sistema permitirá agregar, consultar o eliminar registros, manteniendo la integridad de la información.

Tabla 21. Historia del usuario 5

Historia de usuario
Número: 5 **Usuario:** Administrador **Nombre historia:** Interfaz de usuario (Flask) **Prioridad:** Alta **Riesgo en desarrollo:** Medio. **Estimación de tiempo:** 1.5 semanas **Iteración:** 5 **Responsable:** Gabriel Villarreal **Descripción:** El administrador diseñará la interfaz web que permitirá la interacción con las funciones principales del sistema. **Detalle:** Se implementará una interfaz amigable que muestre resultados de búsqueda, menús de navegación y opciones de carga.

Tabla 22. Historia del usuario 6

Historia de usuario
Número: 6 **Usuario:** Administrador **Nombre historia:** Reporte y visualización de resultados **Prioridad:** Alta **Riesgo en desarrollo:** Bajo **Estimación de tiempo:** 1 semana **Iteración:** 6 **Responsable:** Gabriel Villarreal **Descripción:** El administrador podrá generar reportes automáticos basados en los resultados del análisis del sistema. **Detalle:** El sistema permitirá exportar los reportes en formatos estándar y mostrará el procesamiento realizado.

E. Tareas de ingeniería

Tabla 23. Tarea del usuario 1

Tarea de usuario
Número de la tarea: 1 **Número de historia:** 1 **Nombre de tarea:** Diseño del flujo de carga e indexación **Tipo de tarea:** Análisis y desarrollo **Puntos estimados:** 0.5 **Fecha inicio:** 01/08/2025 **Fecha fin:** 05/08/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Definir la estructura y el flujo de carga de documentos TIC en formato PDF, estableciendo los procedimientos de indexación automática.

Tabla 24. Tarea del usuario 2

Tarea de usuario
Número de la tarea: 2 **Número de historia:** 2 **Nombre de tarea:** Implementación del módulo de carga de documentos **Tipo de tarea:** Desarrollo **Puntos estimados:** 0.5 **Fecha inicio:** 06/08/2025 **Fecha fin:** 10/08/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Desarrollar la funcionalidad que permite subir archivos PDF, validar su formato y almacenarlos temporalmente para su procesamiento.

Tabla 25. Tarea del usuario 3

Tarea de usuario
Número de la tarea: 3 **Número de historia:** 2 **Nombre de tarea:** Configuración del motor OCR **Tipo de tarea:** Desarrollo **Puntos estimados:** 0.8 **Fecha inicio:** 11/08/2025 **Fecha fin:** 16/08/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Configurar y adaptar la librería OCR (Tesseract) para reconocer texto dentro de los documentos escaneados, asegurando su precisión.

Tabla 26. Tarea del usuario 4

Tarea de usuario
Número de la tarea: 4 **Número de historia:** 2 **Nombre de tarea:** Optimización del reconocimiento de texto **Tipo de tarea:** Desarrollo **Puntos estimados:** 0.6 **Fecha inicio:** 17/08/2025 **Fecha fin:** 22/08/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Ajustar parámetros del OCR y aplicar filtros de limpieza de imagen para mejorar la extracción de texto y reducir errores de lectura.

Tabla 27. Tarea del usuario 5

Tarea de usuario
Número de la tarea: 5 **Número de historia:** 3 **Nombre de tarea:** Desarrollo del módulo NLP para análisis semántico **Tipo de tarea:** Desarrollo **Puntos estimados:** 1 **Fecha inicio:** 23/08/2025 **Fecha fin:** 30/08/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Implementar el módulo NLP para interpretar texto, generar resúmenes y realizar búsquedas inteligentes dentro de los documentos procesados.

Tabla 28. Tarea del usuario 6

Tarea de usuario
Número de la tarea: 6 **Número de historia:** 3 **Nombre de tarea:** Validación de consultas y extracción de información **Tipo de tarea:** Pruebas **Puntos estimados:** 0.5 **Fecha inicio:** 01/09/2025 **Fecha fin:** 05/09/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Probar la precisión del módulo NLP ejecutando consultas de prueba y verificando la calidad de la información extraída.

Tabla 29. Tarea del usuario 7

Tarea de usuario
Número de la tarea: 7 **Número de historia:** 4 **Nombre de tarea:** Diseño de la base de datos indexada **Tipo de tarea:** Diseño **Puntos estimados:** 0.5 **Fecha inicio:** 06/09/2025 **Fecha fin:** 10/09/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Diseñar la estructura de la base de datos SQLite/JSON para almacenar textos procesados y resultados de análisis.

Tabla 30. Tarea del usuario 8

Tarea de usuario
Número de la tarea: 8 **Número de historia:** 4 **Nombre de tarea:** Integración de base de datos con los módulos OCR/NLP **Tipo de tarea:** Implementación **Puntos estimados:** 1 **Fecha inicio:** 11/09/2025 **Fecha fin:** 17/09/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Integrar la base de datos con los módulos OCR y NLP para permitir el almacenamiento y consulta automática de resultados procesados.

Tabla 31. Tarea del usuario 9

Tarea de usuario
Número de la tarea: 9 **Número de historia:** 5 **Nombre de tarea:** Implementar algoritmos de detección **Tipo de tarea:** Desarrollo **Puntos estimados:** 1.2 **Fecha inicio:** 05/07/2024 **Fecha fin:** 10/07/2024 **Programador responsable:** Gabriel Villarreal **Descripción:** Implementar algoritmos de detección de defectos basados en características específicas de los diferentes tipos de defectos de los trabajos de titulación.

Tabla 32. Tarea del usuario 10

Tarea de usuario
Número de la tarea: 10 **Número de historia:** 6 **Nombre de tarea:** Implementación del módulo de reportes **Tipo de tarea:** Desarrollo **Puntos estimados:** 0.8 **Fecha inicio:** 29/09/2025 **Fecha fin:** 04/10/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Desarrollar el módulo que genera reportes.

Tabla 33. Tarea del usuario 11

Tarea de usuario
Número de la tarea: 11 **Número de historia:** 6 **Nombre de tarea:** Pruebas integrales del sistema **Tipo de tarea:** Pruebas **Puntos estimados:** 1 **Fecha inicio:** 05/10/2025 **Fecha fin:** 15/10/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Ejecutar pruebas integrales para verificar la comunicación entre módulos (OCR, NLP, interfaz y base de datos).

Tabla 34. Tarea del usuario 12

Tarea de usuario
Número de la tarea: 12 **Número de historia:** 6 **Nombre de tarea:** Optimización final y documentación técnica **Tipo de tarea:** Documentación **Puntos estimados:** 0.6 **Fecha inicio:** 16/10/2025 **Fecha fin:** 25/10/2025 **Programador responsable:** Gabriel Villarreal **Descripción:** Realizar ajustes finales en el sistema, actualizar la documentación técnica y preparar el sistema para su presentación final.

F) Estimación de tareas de usuarios

Tabla 35. Estimación de tareas de usuarios

Nombre Historia	N.º Tarea	Tarea	Semanas	Día	Horas
Carga e indexación de documentos TIC	1	Diseño del flujo de carga e indexación	0.5	3	15
	2	Implementación del módulo de carga de documentos	0.5	3	15
Reconocimiento Óptico de Caracteres (OCR)	3	Configuración del motor OCR	0.8	5	25
	4	Optimización del reconocimiento de texto	0.6	4	20
Procesamiento de Lenguaje Natural (NLP)	5	Desarrollo del módulo NLP para análisis semántico	1.0	6	30
	6	Validación de consultas y extracción de información	0.5	3	15
Gestión de base de datos indexada	7	Diseño de la base de datos indexada	0.5	3	15
	8	Integración de base de datos con módulos OCR/NLP	1.0	6	30
Interfaz de usuario (Flask)	9	Desarrollo de la interfaz web en Flask	1.2	7	35
Reportes y visualización	10	Implementación del módulo de reportes	0.8	5	25
	11	Pruebas integrales del sistema	1.0	6	30
Optimización final	12	Documentación y entrega final	0.6	4	20
Total, estimado			**10.0**	**55**	**305**

G) Plan de entrega del proyecto

Tabla 36. Plan de entrega de proyecto

			Calendario Estimado		
Modulo	**Nro.**	**Nombre de la Historia**	**Semanas Estimadas**	**Días Estimados**	**Horas Estimadas**
Carga e indexación de documentos TIC	1	Carga e indexación automática de documentos PDF	2	10	40
Reconocimiento óptico de caracteres (OCR)	2	Extracción de texto mediante OCR	3	15	60
Procesamiento de Lenguaje Natural (NLP)	3	búsqueda y análisis semántico de la información	3	15	60
Gestión de base de datos indexada	4	Almacenamiento y consulta estructurada de datos	2	10	40
Interfaz de usuario (Flask)	5	Desarrollo de la interfaz web	3	15	60
Reportes y visualización	6	Generación automática de reportes	2	19	40

4.2.3. Fase de diseño

Siguiendo los principios de la metodología ágil XP (Programación Extrema), el diseño se encarga de organizar el sistema inteligente con detalle. En este periodo, se establecieron las clases, responsabilidades y cooperaciones de cada uno de los módulos funcionales que componen el sistema, además de definir estos módulos. Las tarjetas CRC (Clase, Responsabilidad, Colaboración) se utilizaron para expresar esta estructura, ya que posibilitan la organización del diseño conceptual de cada elemento. Se utilizó Flask como marco principal para crear la interfaz de usuario, incorporando plantillas HTML, CSS y JavaScript. Esto facilitó el desarrollo de un entorno web interactivo, ajustable y fácil de navegar. Las plantillas se estructuraron de manera modular siguiendo el modelo MVC, lo cual posibilitó que la lógica y la presentación fueran separadas. Además, se añadieron elementos responsivos para garantizar la compatibilidad con varios dispositivos, brindando una experiencia que es clara, moderna y útil para el usuario.

Tarjetas CRC

Tabla 37. Tarjeta CRC Módulo de carga e indexación de documentos

TARJETA CRC	
Nombre: Módulo de carga e indexación de documentos	
Responsabilidades:	**Colaboradores:**
• Permitir la carga de documentos TIC en formato PDF e imagen • Ejecutar la indexación de archivos para su almacenamiento estructurado • Enviar datos al módulo OCR para su procesamiento textual	• Módulo OCR • Módulo de gestión de base de datos indexada • Interfaz de usuario (Flask)

Tabla 38. Tarjeta CRC Módulo OCR (Reconocimiento óptico de caracteres)

TARJETA CRC	
Nombre: Módulo OCR (Reconocimiento Óptico de Caracteres)	
Responsabilidades:	**Colaboradores:**
• Extraer texto de documentos PDF o imágenes escaneadas • Optimizar la lectura mediante filtros de limpieza y segmentación • Transferir los datos procesados al módulo NLP para su análisis semántico	• Módulo de carga e indexación de documentos • Módulo NLP • Módulo de gestión de base de datos indexada

Tabla 39. Tarjeta CRC Módulo NLP (Procesamiento de Lenguaje Natural)

TARJETA CRC	
Nombre: Módulo NLP (Procesamiento de Lenguaje Natural)	
Responsabilidades:	**Colaboradores:**
• Analizar el texto extraído para identificar patrones semánticos y contextuales • Procesar consultas del usuario mediante modelos de inteligencia artificial • Generar respuestas precisas relacionadas con los documentos TIC almacenados	• Módulo OCR • Módulo de gestión de base de datos indexada • Interfaz de usuario (Flask)

Tabla 40. Tarjeta CRC Módulo de Gestión de Base de Datos Indexada

TARJETA CRC	
Nombre: Módulo de gestión de base de datos indexada	
Responsabilidades:	**Colaboradores:**
• Almacenar la información procesada de manera estructurada • Facilitar la interacción del usuario • Sincronizar los resultados entre los módulos OCR, NLP e interfaz de usuario	• Módulo NLP • Módulo OCR • Interfaz de usuario (Flask)

Tabla 41. Tarjeta CRC Interfaz de Usuario (Flask Web)

TARJETA CRC	
Nombre: Interfaz de usuario (Flask Web)	
Responsabilidades:	**Colaboradores:**
• Proporcionar una interfaz gráfica accesible y dinámica al usuario final • Permitir búsquedas, visualización de resultados y descargas de reportes • Conectar todos los módulos del sistema mediante endpoints seguros	• Módulo de carga e indexación de documentos • Módulo NLP • Interfaz de gestión de base de datos indexada

Diseño de prototipos

En la creación de los prototipos, se utilizaron las herramientas digitales Figma y Draw.io, que permitieron modelar las interfaces y diagramas funcionales del sistema.

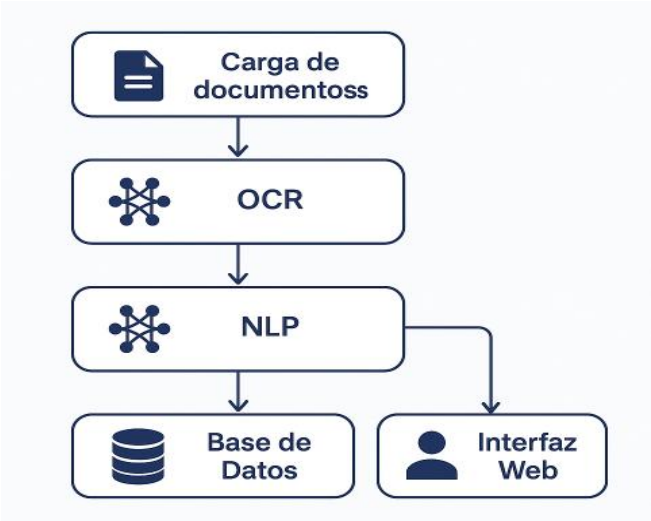


Figura 17. Flujo funcional del sistema

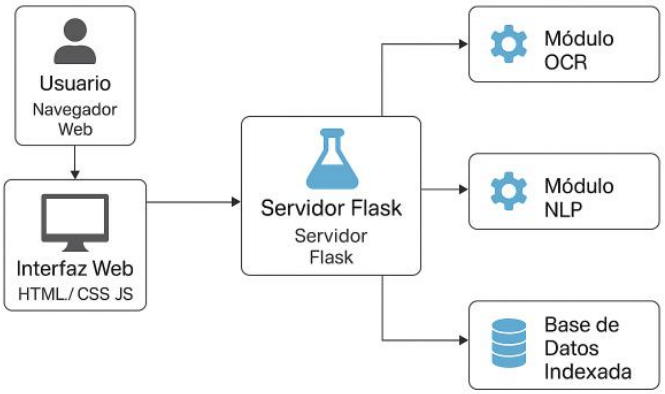


Figura 18. Arquitectura lógica del sistema

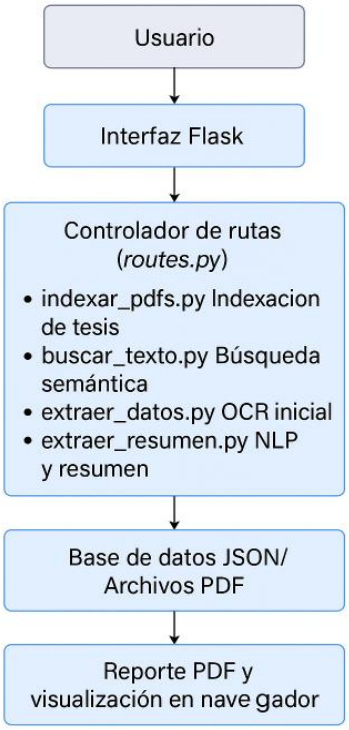


Figura 19. Prototipo lógico del sistema

Figura 20. Prototipo de la interfaz principal

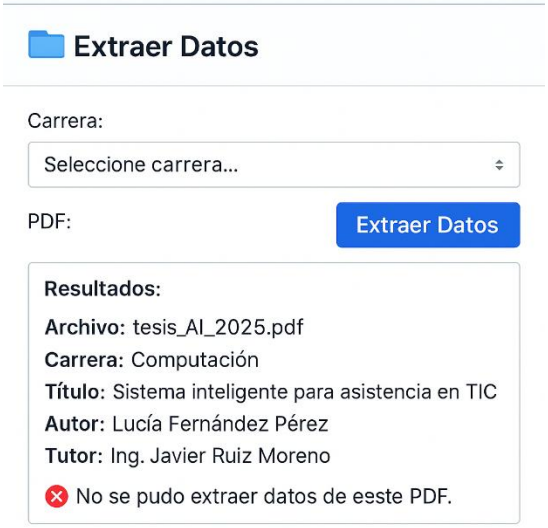


Figura 21. Prototipo de módulo de búsqueda y resultados

Diagramas de caso de uso

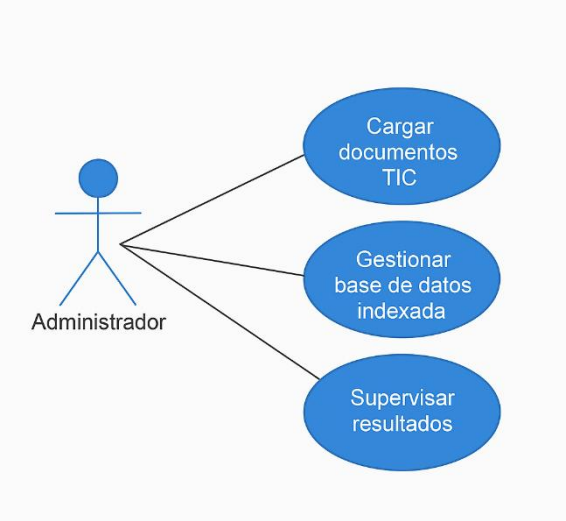


Figura 22. Diagrama de caso de uso administrador

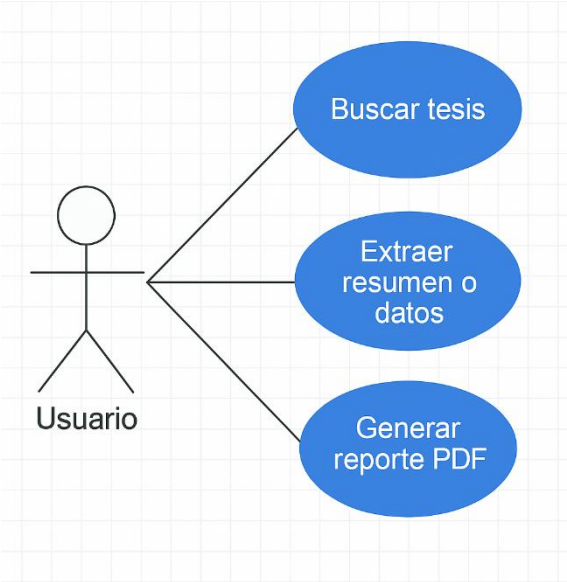


Figura 23. Diagrama de caso de uso usuario

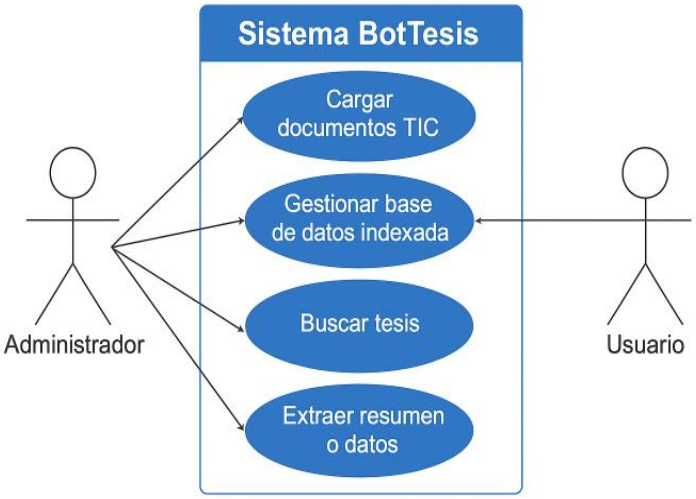


Figura 24. Diagrama general de caso de uso del sistema

Funcionamiento del sistema inteligente

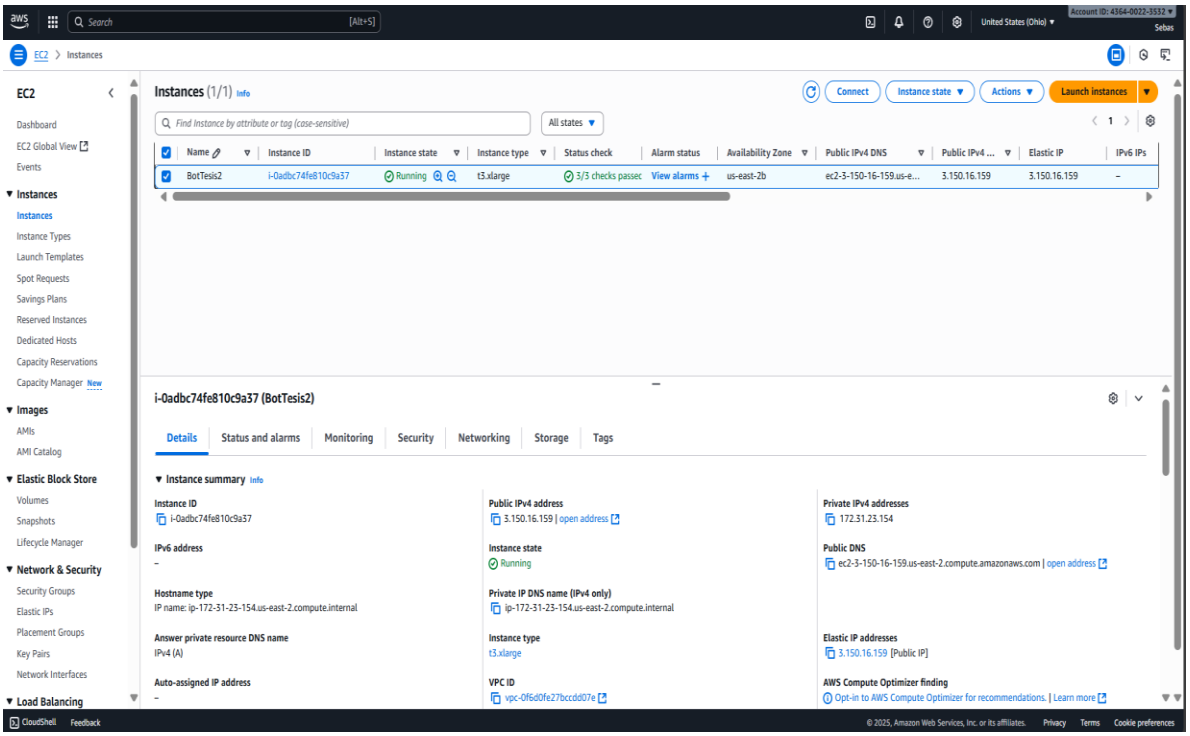


Figura 25. Activación servidor de nube AWS y entorno de trabajo

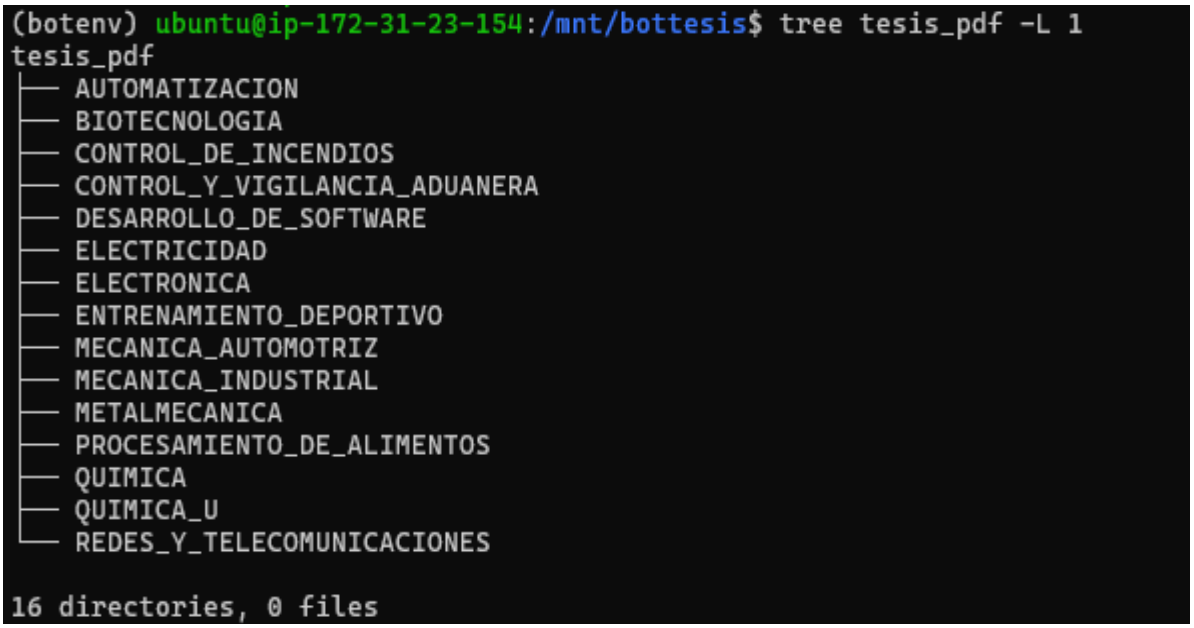


Figura 26. Carga y gestión de documentos TIC

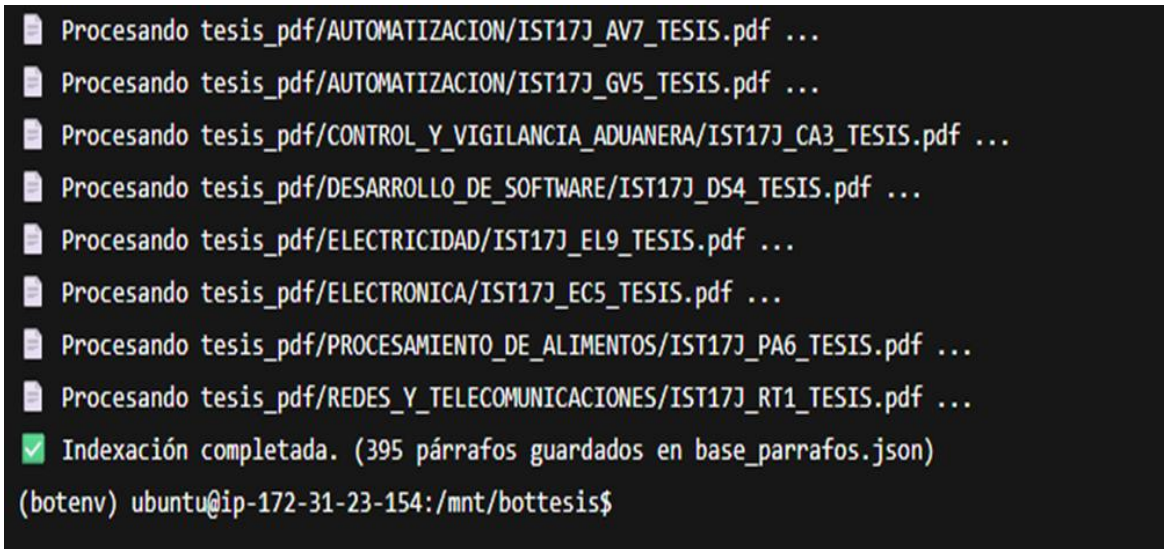


Figura 28. Extracción y procesamiento OCR

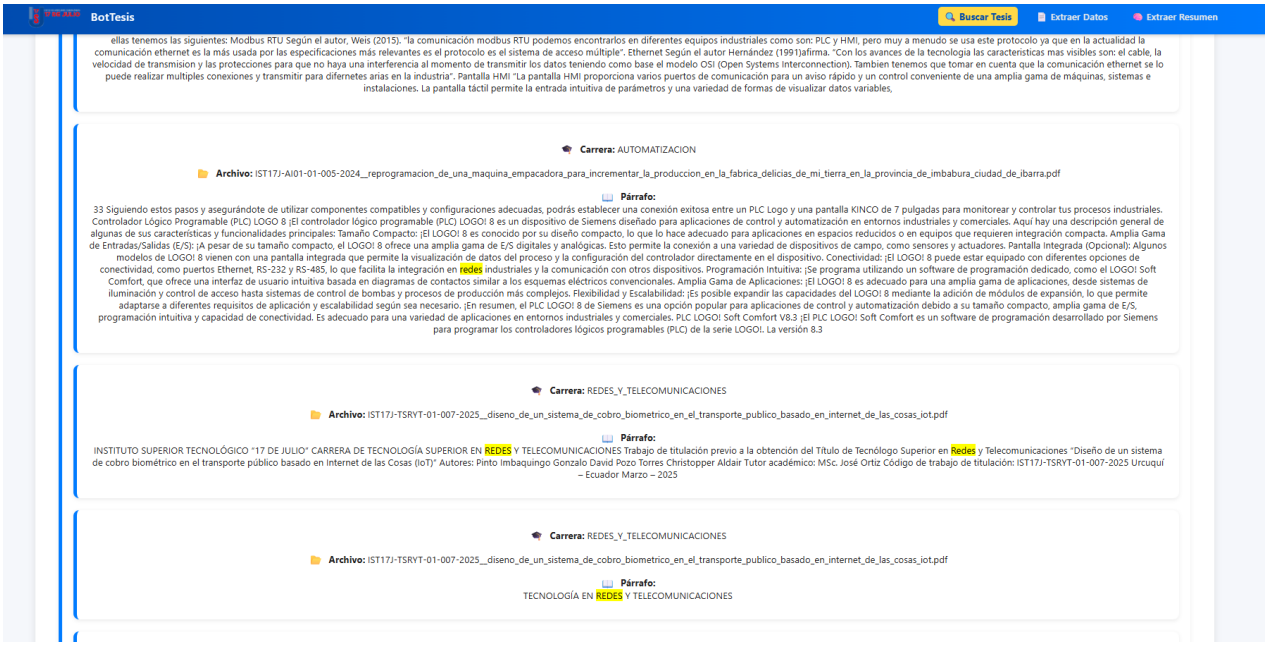


Figura 27. Búsqueda inteligente de tesis

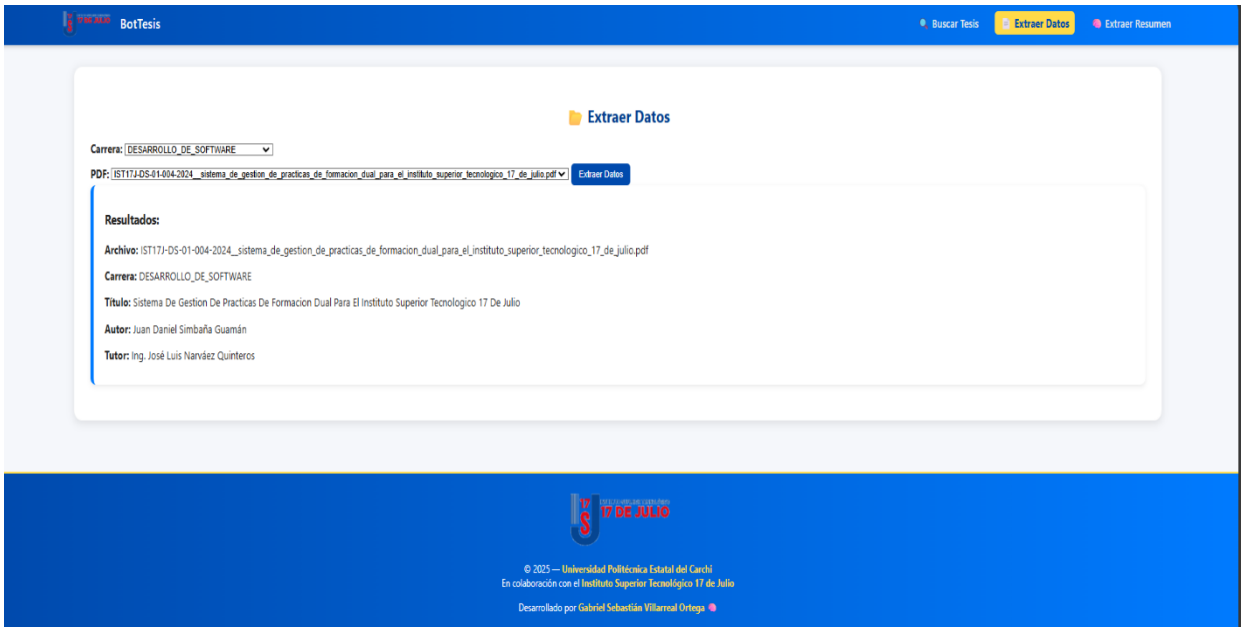


Figura 29. Extracción de datos del documento TIC

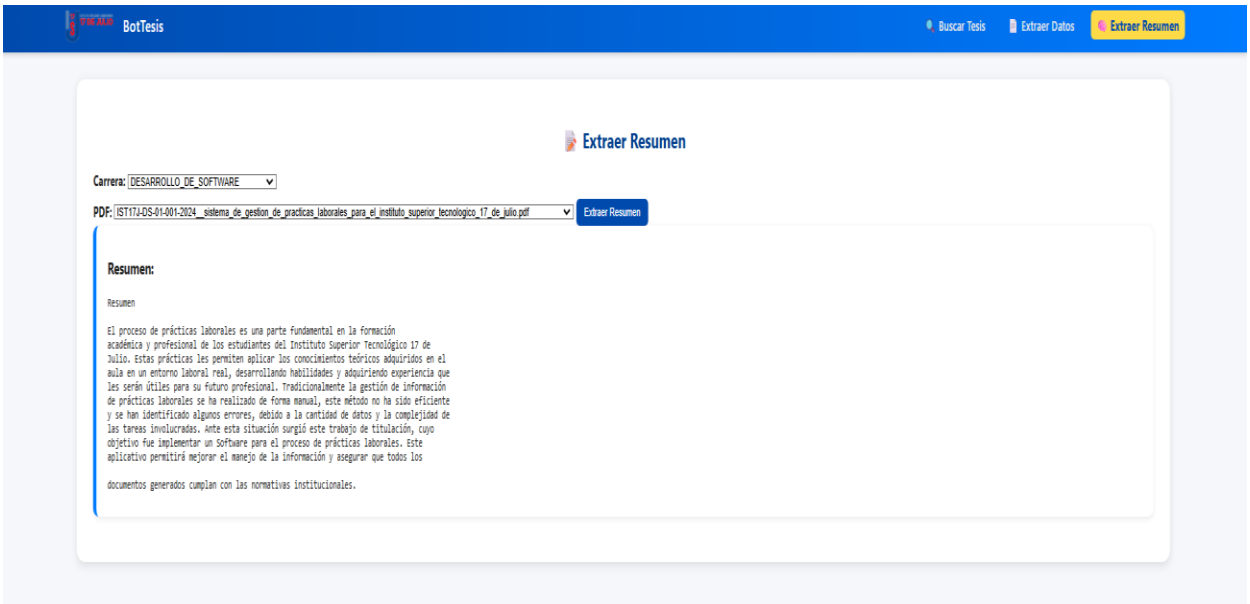


Figura 30. Generación de resumen

Codificación del sistema

```
import os
import fitz  # PyMuPDF
import pytesseract
from PIL import Image
import json
```

Figura 31. Importación de recursos necesarios

El código lleva a cabo la importación de las bibliotecas que posibilitan el correcto funcionamiento del sistema. Se distinguen entre ellas: Flask (servidor web y renderizado de plantillas Jinja), os y json (gestión de archivos y estructuras de datos), re (expresiones regulares), PyMuPDF (fitz) y pdfplumber (lectura y análisis de PDF), Pillow y pdf2image (manejo de imágenes), pytesseract (OCR en español) son algunos ejemplos, además de herramientas auxiliares como numpy, regex o pydantic. Estas dependencias permiten procesar documentos de manera híbrida: extraer texto directamente si hay contenido digital y, si no, extraerlo a través de OCR sobre mapas de bits creados a 300 DPI para conseguir un reconocimiento más preciso.

```
BASE_DIR = os.path.dirname(os.path.abspath(__file__))
PDFS_DIR = os.path.join(BASE_DIR, "tesis_pdf")
JSON_FILE = os.path.join(BASE_DIR, "base_parrafos.json")

def extraer_texto_pagina(pagina):
    """Extrae texto de una página PDF. Si es imagen, aplica OCR."""
    texto = pagina.get_text("text")
    if texto.strip():
        return texto
    pix = pagina.get_pixmap(dpi=300)
    img = Image.frombytes("RGB", [pix.width, pix.height], pix.samples)
    return pytesseract.image_to_string(img, lang="spa")
```

Figura 32. Extracción y preprocesamiento de texto (OCR + PDF)

El módulo de extracción pone en práctica funciones que se complementan. Por un lado, se emplea una rutina que trata de extraer el texto utilizando métodos nativos en una página PDF; si no hay contenido, la página se rasteriza a 300 DPI y se aplica OCR con el idioma 'spa'. Por otra parte, cuando las páginas son digitales, se puede procesar partes del documento utilizando pdfplumber, que ofrece una función de lectura por rangos. En los dos escenarios, se normaliza la salida (eliminación de artefactos, unificación de codificaciones y limpieza) para que estén disponibles cadenas coherentes para su almacenamiento y tokenización posteriores.

```python
def buscar_en_json(consulta):
    """Busca la consulta en base_parrafos.json y devuelve coincidencias con resaltado."""
    resultados = []
    consulta_lower = consulta.lower()

    if not os.path.exists(JSON_PATH):
        return resultados

    with open(JSON_PATH, "r", encoding="utf-8") as f:
        data = json.load(f)

    for item in data:
        parrafo = item["parrafo"]
        if consulta_lower in parrafo.lower():
            # Resaltamos la frase buscada con <mark>
            parrafo_resaltado = re.sub(
                f"({re.escape(consulta)})",
                r"<mark>\1</mark>",
                parrafo,
                flags=re.IGNORECASE
            )
            resultados.append({
                "carrera": item["carrera"],
                "archivo": item["archivo"],
                "parrafo": parrafo_resaltado
            })

    return resultados
```

Figura 33. Indexación y búsqueda semántica de sistema

La indexación analiza el repositorio de documentos que están organizados por carreras, procesa las páginas y crea una base JSON (base_parrafos.json) que contiene registros con metadatos, tales como archivo, ruta, número de página y el párrafo extraído. La búsqueda se realiza en esta base indexada utilizando una función de búsqueda que normaliza la frase ingresada a letras minúsculas y aplica coincidencias flexibles. La interfaz web recibe la consulta del usuario y organiza los resultados en páginas de 20 ítems cada una, presentando fragmentos destacados (snippets) y enlaces al documento original. Este flujo asegura tiempos de respuesta apropiados y fomenta la capacidad de rastrear entre el texto recuperado y su ubicación original.

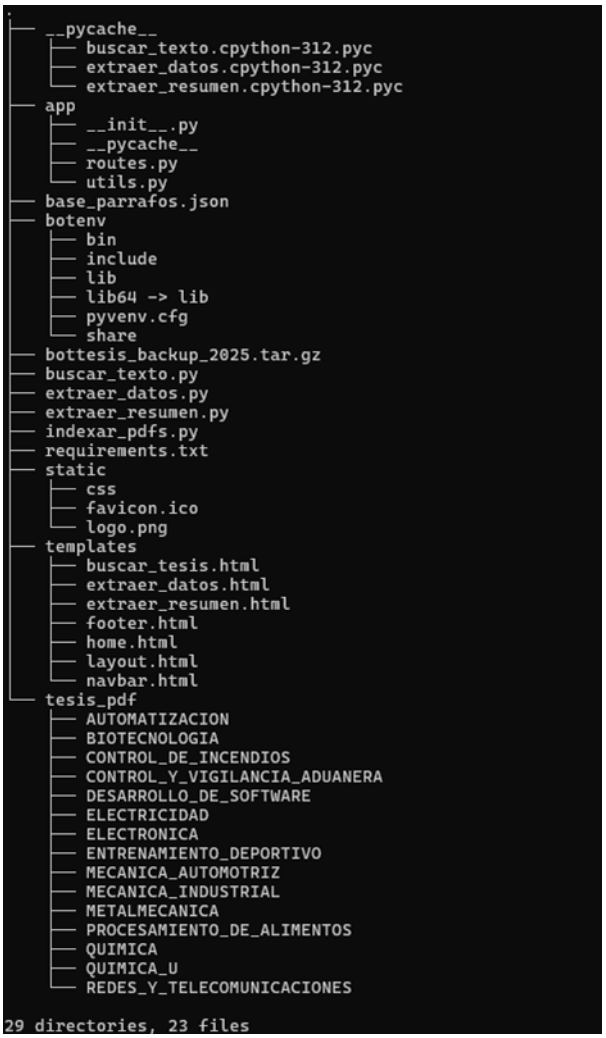


Figura 35. Estructura del proyecto

```python
from flask import Flask
import os

def create_app():
    base_dir = os.path.abspath(os.path.dirname(__file__))
    templates_dir = os.path.join(os.path.dirname(base_dir), "templates")
    static_dir = os.path.join(os.path.dirname(base_dir), "static")

    app = Flask(__name__, template_folder=templates_dir, static_folder=static_dir)

    # Importa y registra las rutas
    from .routes import main
    app.register_blueprint(main)

    return app

# Contexto global para fecha actual
from datetime import datetime
def inject_now():
    return {'now': datetime.now()}
```

Figura 34. Modulo principal Flask

```python
from flask import Blueprint, render_template, request
import buscar_texto
import os
import subprocess
import extraer_datos
import extraer_resumen
import math

main = Blueprint('main', __name__)

# --- Página de Inicio ---
@main.route("/")
def home():
    return render_template("home.html", title="Inicio | BotTesis")

# --- Buscador de Tesis ---
@main.route("/buscar", methods=["GET", "POST"])
def buscar_tesis():
    consulta = ""
    resultados = []
    pagina = int(request.args.get("pagina", 1))
    por_pagina = 20

    if request.method == "POST":
        consulta = request.form.get("consulta", "").strip()
        pagina = 1
        if consulta:
            resultados = buscar_texto.buscar_en_json(consulta)
    elif request.method == "GET":
        consulta = request.args.get("consulta", "").strip()
        if consulta:
            resultados = buscar_texto.buscar_en_json(consulta)

    total_resultados = len(resultados)
    total_paginas = math.ceil(total_resultados / por_pagina) if total_resultados else 1

    inicio = (pagina - 1) * por_pagina
    fin = inicio + por_pagina
    resultados_pagina = resultados[inicio:fin]

    return render_template(
        "buscar_tesis.html",
        consulta=consulta,
        resultados=resultados_pagina,
        total_resultados=total_resultados,
        pagina=pagina,
        total_paginas=total_paginas
    )

# --- Extraer Datos ---
@main.route("/extraer_datos", methods=["GET", "POST"])
def ver_datos():
    base_dir = "/mnt/bottesis/tesis_pdf"

    carreras = []
    if os.path.exists(base_dir):
        carreras = [c for c in os.listdir(base_dir) if os.path.isdir(os.path.join(base_dir, c))]
        carreras.sort()

    datos = None
    pdfs = []
    carrera_sel = None
    pdf_sel = None

    if request.method == "POST":
        carrera_sel = request.form.get("carrera")
        pdf_sel = request.form.get("pdf")

        if carrera_sel:
            carpeta = os.path.join(base_dir, carrera_sel)
            if os.path.exists(carpeta):
                pdfs = [f for f in os.listdir(carpeta) if f.lower().endswith(".pdf")]
                pdfs.sort()

        if carrera_sel and pdf_sel:
            pdf_path = os.path.join(base_dir, carrera_sel, pdf_sel)
            if os.path.exists(pdf_path):
                result = subprocess.run(
                    ["pdftotext", "-f", "1", "-l", "1", pdf_path, "-"],
                    stdout=subprocess.PIPE, stderr=subprocess.PIPE, text=True
```

Figura 36. Rutas principales

```python
import os
import re
import subprocess

BASE_DIR = "tesis_pdf"

def limpiar_titulo(nombre):
    nombre = os.path.splitext(nombre)[0]
    nombre = re.sub(r"^IST17J[-_A-Z0-9]+__", "", nombre, flags=re.IGNORECASE)
    nombre = nombre.replace("_", " ").replace("-", " ")
    return nombre.strip().title()

def capitalizar_texto(texto):
    if not texto or texto == "No encontrado":
        return texto
    return " ".join(p.capitalize() for p in texto.split())

def extraer_campos(texto):
    datos = {"AUTOR": "No encontrado", "TUTOR": "No encontrado"}
    lineas = [l.strip() for l in texto.splitlines() if l.strip()]

    patron_autor = re.compile(r"^(AUTOR(?:ES)?|PRESENTADO POR)[: ]", re.IGNORECASE)
    for l in lineas:
        if patron_autor.match(l):
            datos["AUTOR"] = l.split(":", 1)[-1].strip()
            break

    patron_tutor = re.compile(r"^(TUTOR|TUTORA|DIRECTOR|LCDO|MGS|ING)", re.IGNORECASE)
    for l in lineas:
        if patron_tutor.match(l):
            datos["TUTOR"] = l.split(":", 1)[-1].strip() if ":" in l else l.strip()
            break

    datos["AUTOR"] = capitalizar_texto(datos["AUTOR"])
    datos["TUTOR"] = capitalizar_texto(datos["TUTOR"])
    return datos

def listar_carreras():
    return sorted([c for c in os.listdir(BASE_DIR) if os.path.isdir(os.path.join(BASE_DIR, c))])

def listar_pdfs(carrera):
    ruta_carrera = os.path.join(BASE_DIR, carrera)
    return sorted([f for f in os.listdir(ruta_carrera) if f.endswith(".pdf")])
```

Figura 37. Funciones auxiliares

```python
import os
import fitz  # PyMuPDF
import pytesseract
from PIL import Image
import json

BASE_DIR = os.path.dirname(os.path.abspath(__file__))
PDFS_DIR = os.path.join(BASE_DIR, "tesis_pdf")
JSON_FILE = os.path.join(BASE_DIR, "base_parrafos.json")

def extraer_texto_pagina(pagina):
    """Extrae texto de una página PDF. Si es imagen, aplica OCR."""
    texto = pagina.get_text("text")
    if texto.strip():
        return texto
    pix = pagina.get_pixmap(dpi=300)
    img = Image.frombytes("RGB", [pix.width, pix.height], pix.samples)
    return pytesseract.image_to_string(img, lang="spa")

def indexar_pdfs():
    data = []
    for carrera in os.listdir(PDFS_DIR):
        carpeta_carrera = os.path.join(PDFS_DIR, carrera)
        if not os.path.isdir(carpeta_carrera):
            continue

        for archivo in os.listdir(carpeta_carrera):
            if archivo.endswith(".pdf"):
                ruta_pdf = os.path.join(carpeta_carrera, archivo)
                print(f"📄 Procesando {ruta_pdf} ...")
                try:
                    with fitz.open(ruta_pdf) as doc:
                        for pagina in doc:
                            texto = extraer_texto_pagina(pagina)
                            # Dividimos por párrafos reales (doble salto de línea)
                            parrafos = texto.split("\n\n")
                            for parrafo in parrafos:
                                parrafo_limpio = " ".join(parrafo.split())
                                if parrafo_limpio:
                                    data.append({
                                        "carrera": carrera,
                                        "archivo": archivo,
                                        "parrafo": parrafo_limpio
                                    })
                except Exception as e:
                    print(f"⚠️ Error con {ruta_pdf}: {e}")
```

Figura 38. Módulo de indexación

```python
                print(f"📄 Procesando {ruta_pdf} ...")
                try:
                    with fitz.open(ruta_pdf) as doc:
                        for pagina in doc:
                            texto = extraer_texto_pagina(pagina)
                            # Dividimos por párrafos reales (doble salto de línea)
                            parrafos = texto.split("\n\n")
                            for parrafo in parrafos:
                                parrafo_limpio = " ".join(parrafo.split())
                                if parrafo_limpio:
                                    data.append({
                                        "carrera": carrera,
                                        "archivo": archivo,
                                        "parrafo": parrafo_limpio
                                    })
                except Exception as e:
                    print(f"⚠️ Error con {ruta_pdf}: {e}")

    with open(JSON_FILE, "w", encoding="utf-8") as f:
        json.dump(data, f, ensure_ascii=False, indent=2)

    print(f"✅ Indexación completada. {len(data)} párrafos guardados en {JSON_FILE}")

if __name__ == "__main__":
    indexar_pdfs()
```

Figura 39. Validación y limpieza de datos generados

```python
import os
import re
import subprocess

BASE_DIR = "tesis_pdf"

def limpiar_titulo(nombre):
    """Usar nombre del PDF como título, quitando prefijos de código."""
    nombre = os.path.splitext(nombre)[0]
    # quitar prefijos tipo IST17J-XXX-001-2025__
    nombre = re.sub(r"^IST17J[-_A-Z0-9]+__", "", nombre, flags=re.IGNORECASE)
    nombre = nombre.replace("_", " ").replace("-", " ")
    return nombre.strip().title()

def capitalizar_texto(texto):
    """Capitalizar texto (autor/tutor) respetando palabras múltiples."""
    if not texto or texto == "No encontrado":
        return texto
    return " ".join(p.capitalize() for p in texto.split())

def extraer_campos(texto):
    datos = {
        "AUTOR": "No encontrado",
        "TUTOR": "No encontrado"
    }

    lineas = [l.strip() for l in texto.splitlines() if l.strip()]

    # Autor
    patron_autor = re.compile(r"^(AUTOR(?:ES)?|PRESENTADO POR)[: ]", re.IGNORECASE)
    for l in lineas:
        if patron_autor.match(l):
            datos["AUTOR"] = l.split(":", 1)[-1].strip()
            break

    # Tutor
    patron_tutor = re.compile(r"^(TUTOR|TUTORA|DIRECTOR|LCDO|MGS|ING)", re.IGNORECASE)
    for l in lineas:
        if patron_tutor.match(l):
            datos["TUTOR"] = l.split(":", 1)[-1].strip() if ":" in l else l.strip()
            break

    # Capitalizar resultados
    datos["AUTOR"] = capitalizar_texto(datos["AUTOR"])
    datos["TUTOR"] = capitalizar_texto(datos["TUTOR"])

    return datos

def procesar_pdfs(return_data=False):
    """Recorre PDFs y obtiene Título/Autor/Tutor.
    - CLI: return_data=False imprime por consola (como antes).
    - Web: return_data=True devuelve lista de dicts.
    """
    registros = []

    # Recorrer todas las carpetas y PDFs
    for carrera in os.listdir(BASE_DIR):
        ruta_carrera = os.path.join(BASE_DIR, carrera)
        if os.path.isdir(ruta_carrera):
            for archivo in os.listdir(ruta_carrera):
                if archivo.endswith(".pdf"):
                    ruta_pdf = os.path.join(ruta_carrera, archivo)
                    try:
                        # Extraer autor y tutor de la primera página
                        result = subprocess.run(
                            ["pdftotext", "-f", "1", "-l", "1", ruta_pdf, "-"],
                            stdout=subprocess.PIPE, stderr=subprocess.PIPE, text=True
                        )
                        texto = result.stdout
                        campos = extraer_campos(texto)

                        registros.append({
                            "ARCHIVO": archivo,
                            "CARRERA": carrera,
                            "TITULO": limpiar_titulo(archivo),
                            "AUTOR": campos["AUTOR"],
                            "TUTOR": campos["TUTOR"]
```

Figura 40. OCR

```python
import os
import pdfplumber
import pytesseract
from pdf2image import convert_from_path
import re
import textwrap
from pytesseract import Output

# --- CONFIGURACIÓN ---
TESIS_DIR = "tesis_pdf"
START_PAGE, END_PAGE = 8, 15

def extraer_texto_pdf(pdf_path, start, end):
    """Extrae texto directamente de PDFs digitales"""
    texto = ""
    try:
        with pdfplumber.open(pdf_path) as pdf:
            for i in range(start-1, min(end, len(pdf.pages))):
                page_text = pdf.pages[i].extract_text()
                if page_text:
                    texto += page_text + "\n"
    except:
        pass
    return texto

def extraer_texto_ocr(pdf_path, start, end):
    """Extrae texto mediante OCR rápido"""
    texto = ""
    try:
        paginas = convert_from_path(pdf_path, first_page=start, last_page=end)
        for img in paginas:
            texto += pytesseract.image_to_string(img, lang="spa") + "\n"
    except:
        pass
    return texto

def extraer_texto_ocr_bloques(pdf_path, start, end):
    """OCR avanzado: busca palabra 'RESUMEN' en bloques con coordenadas"""
    texto = ""
    try:
        paginas = convert_from_path(pdf_path, first_page=start, last_page=end)
        for img in paginas:
            data = pytesseract.image_to_data(img, lang="spa", output_type=Output.DICT)
            palabras = data["text"]

            # Normalizar palabras
            for i, word in enumerate(palabras):
                if re.fullmatch(r"(RESUMEN|RESUM[ÉE]N|RESUMEN|RESÚMEN)", word.upper()):
                    bloque = " ".join(palabras[i:])
                    return bloque
    except:
        pass
    return texto

def limpiar_resumen(texto):
    """Encuentra y limpia el bloque del resumen con tolerancia máxima"""
    if not texto:
        return None

    # Normalizar
    texto_norm = re.sub(r"\s+", " ", texto.upper())
    texto_norm = texto_norm.replace("É", "E").replace("5", "S")

    # Buscar inicio
    patron_inicio = re.search(r"R\s*E\s*S\s*U\s*M\s*E\s*N|RESUM[ÉE]N|RESUMEN|RESÚMEN", texto_norm)
    if not patron_inicio:
        return None

    start_idx = patron_inicio.start()
    resumen = texto[start_idx:]

    # Buscar fin
    patron_fin = re.search(r"PALABRAS?\s*CLAVE[S]?", resumen, re.IGNORECASE)
    if patron_fin:
        resumen = resumen[:patron_fin.start()]

    return resumen.strip() if resumen.strip() else None
```

Figura 41. NLP y resumen

```
(botenv) ubuntu@ip-172-31-23-154:/mnt/bottesis$ cat buscar_texto.py
import os
import json
import re

BASE_DIR = os.path.dirname(os.path.abspath(__file__))
JSON_PATH = os.path.join(BASE_DIR, "base_parrafos.json")

def buscar_en_json(consulta):
    """Busca la consulta en base_parrafos.json y devuelve coincidencias con resaltado."""
    resultados = []
    consulta_lower = consulta.lower()

    if not os.path.exists(JSON_PATH):
        return resultados

    with open(JSON_PATH, "r", encoding="utf-8") as f:
        data = json.load(f)

    for item in data:
        parrafo = item["parrafo"]
        if consulta_lower in parrafo.lower():
            # Resaltamos la frase buscada con <mark>
            parrafo_resaltado = re.sub(
                f"({re.escape(consulta)})",
                r"<mark>\1</mark>",
                parrafo,
                flags=re.IGNORECASE
            )
            resultados.append({
                "carrera": item["carrera"],
                "archivo": item["archivo"],
                "parrafo": parrafo_resaltado
            })

    return resultados

# 🚀 Test rápido desde consola
if __name__ == "__main__":
    q = input("Ingrese búsqueda: ")
    res = buscar_en_json(q)
    print(f"✅ {len(res)} resultados encontrados.")
    for r in res[:10]:
        print(f"[{r['carrera']}] {r['archivo']} -> {r['parrafo']}")
(botenv) ubuntu@ip-172-31-23-154:/mnt/bottesis$ S
```

Figura 42. Motor de búsqueda semántica

```
<!DOCTYPE html>
<html lang="es">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>{{ title if title else "BotTesis – UPEC / IST 17 de Julio" }}</title>
  <link rel="icon" type="image/x-icon" href="{{ url_for('static', filename='favicon.ico') }
}">
  <link rel="stylesheet" href="{{ url_for('static', filename='css/style.css') }}">
</head>
<body>

  {% include 'navbar.html' %}

  <main class="content-wrapper">
    {% block content %}{% endblock %}
  </main>

  {% include 'footer.html' %}

  <!-- Script global para modo oscuro -->
  <script>
    document.addEventListener("DOMContentLoaded", () => {
      const toggle = document.getElementById("darkModeToggle");
      if (toggle) {
        toggle.addEventListener("click", function() {
          document.body.classList.toggle("dark-mode");
          this.textContent = document.body.classList.contains("dark-mode")
            ? "☀️ Modo Claro"
            : "🌙 Modo Oscuro";
        });
      }
    });
  </script>

</body>
</html>
(botenv) ubuntu@ip-172-31-23-154:/mnt/bottesis/templates$ ls
buscar_tesis.html   extraer_resumen.html  home.html    navbar.html
extraer_datos.html  footer.html           layout.html
(botenv) ubuntu@ip-172-31-23-154:/mnt/bottesis/templates$
```

Figura 43. Estructura base de la plantilla HTML principal del sistema

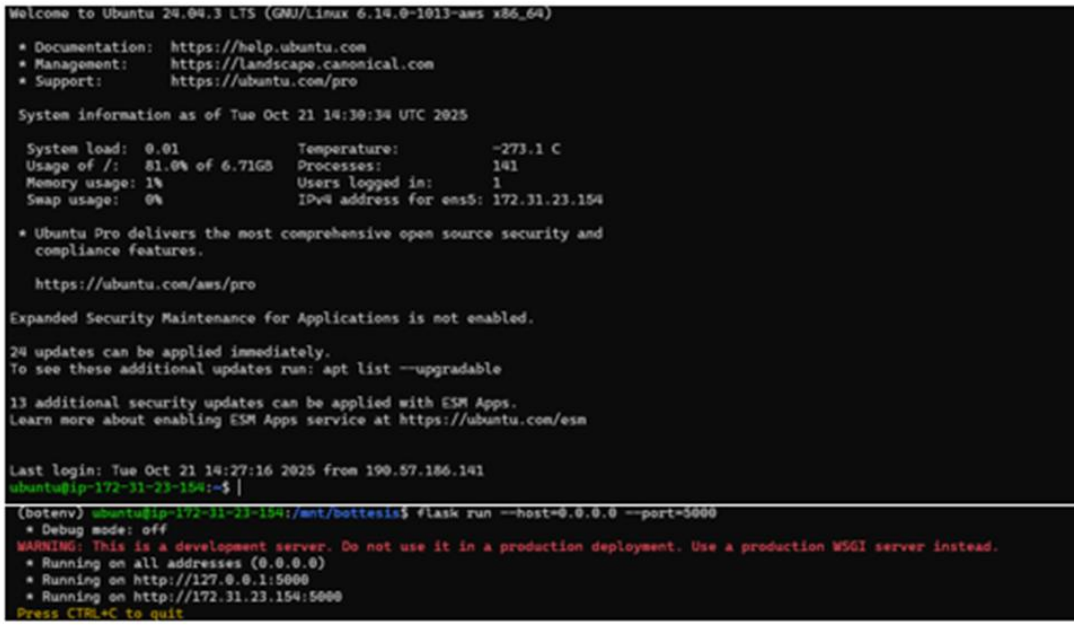


Figura 44. Ejecución del servidor Flask

Interfaz final

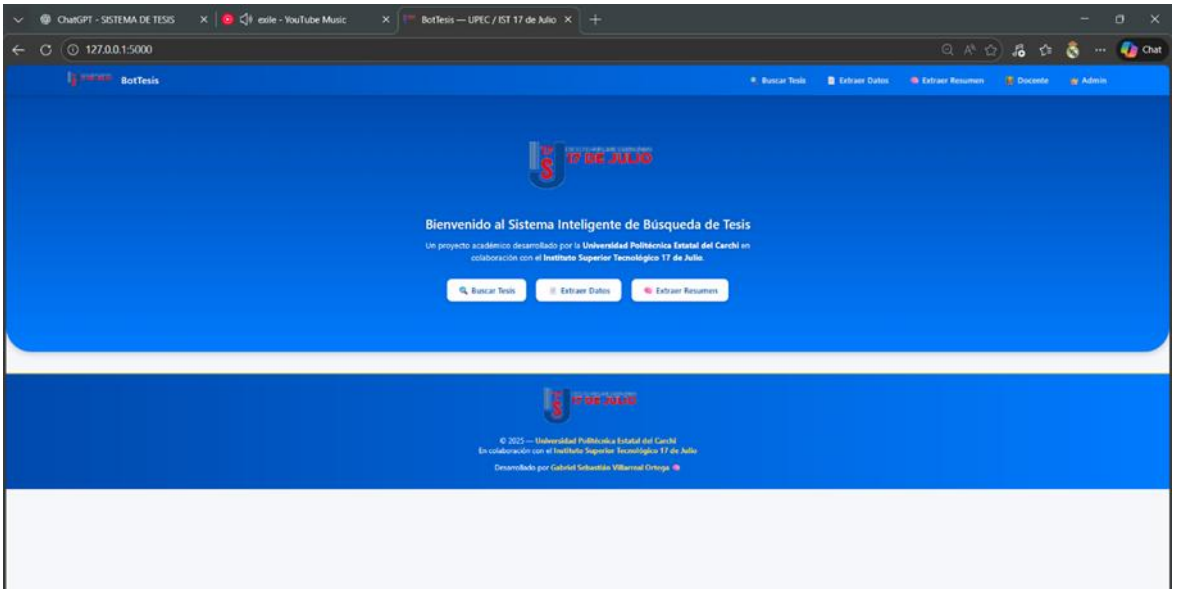


Figura 45. Interfaz visual en navegador

Estructura del diagrama

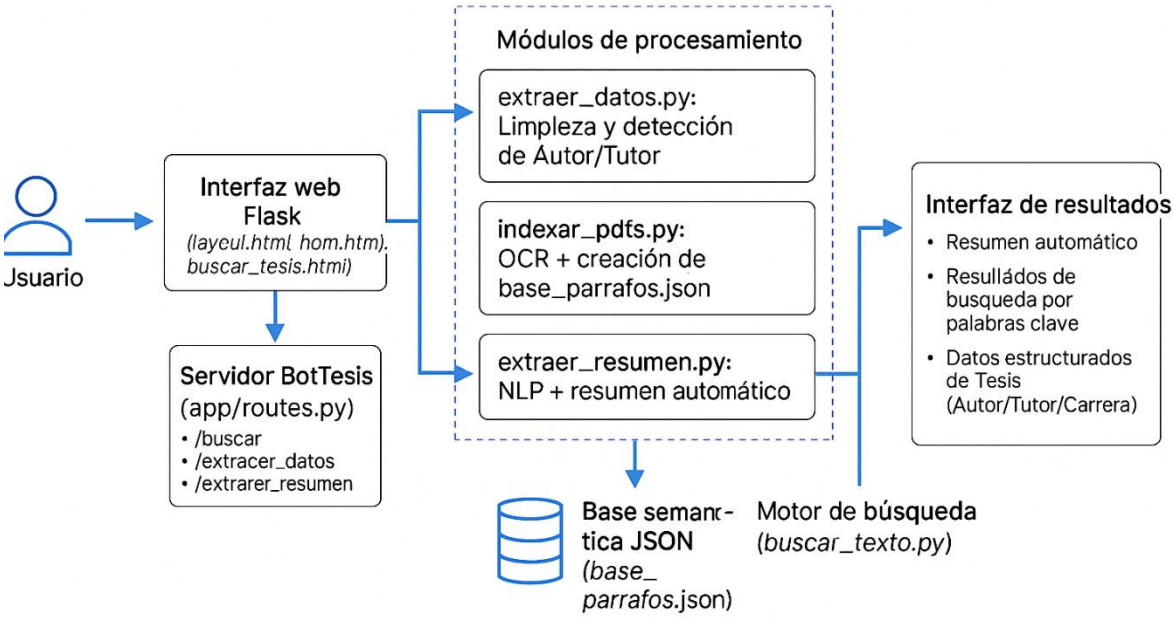


Figura 46. Flujo general de procesamiento del sistema con IA

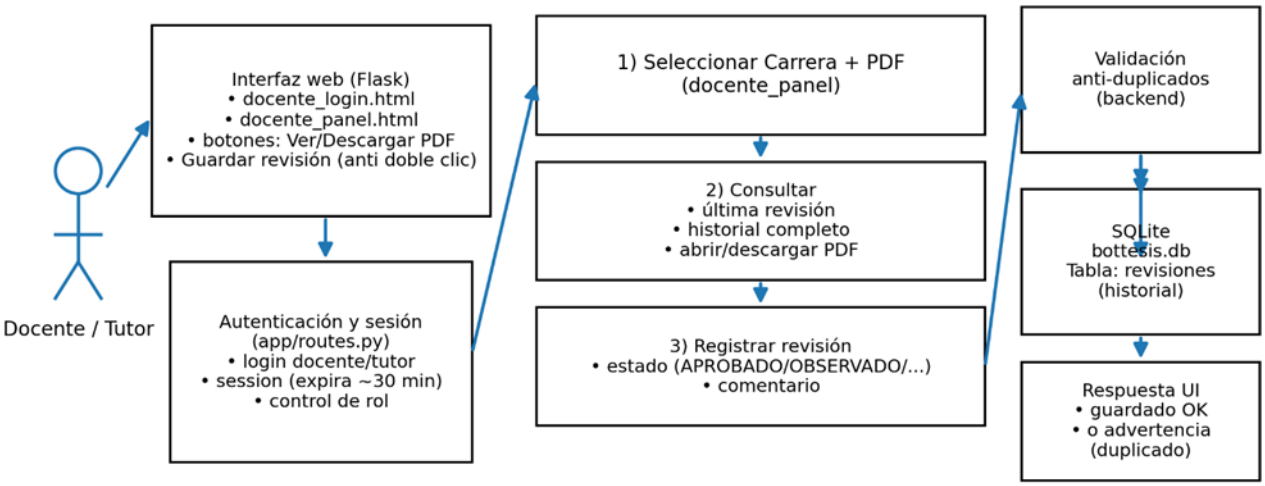


Figura 47. Flujo del docente/tutor

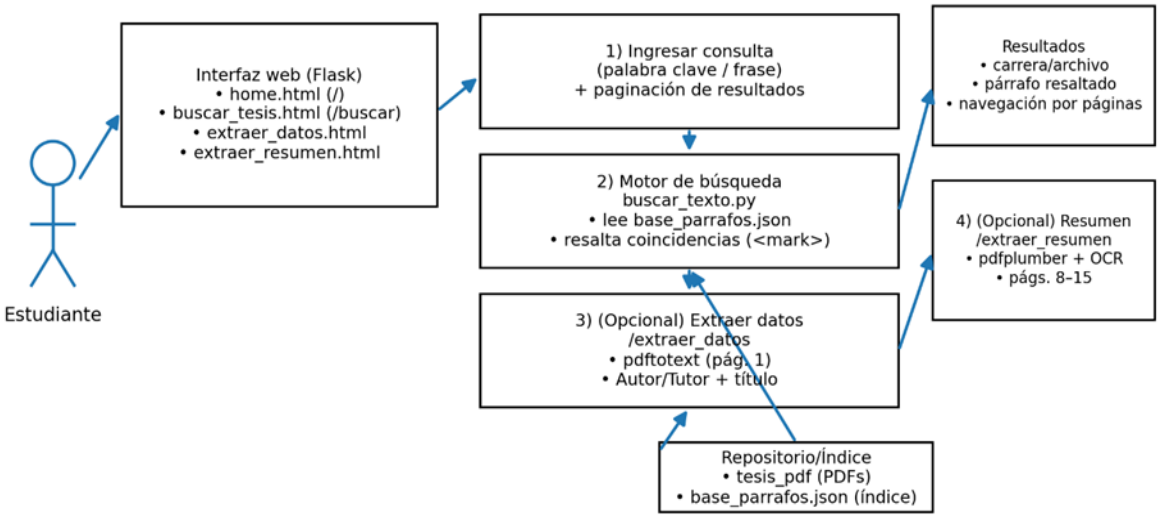


Figura 48. Flujo del estudiante

Tabla 42. Desempeño precisión OCR

Métrica	Valor obtenido
Exactitud (Accuracy)	92 %
Recall	89 %
F1-score	90 %

Tabla 43. Búsqueda manual vs sistema inteligente

Tipo de búsqueda	Tiempo manual	Tiempo con sistema	Reducción
Tesis	10 min	3 s	99.5 %
Tesis examen complexivo	10–15 min	4 s	99.6 %

Estructura organizativa

Figura 49. Organigrama del Instituto Superior Tecnológico 17 de Julio

4.3. DISCUSIÓN

Esta investigación ha mostrado que la puesta en marcha de un sistema inteligente, fundamentado en métodos de inteligencia artificial, es una solución muy eficaz para mejorar el proceso de búsqueda, indexación y recuperación de los Trabajos de Integración Curricular (TIC) en el Instituto Superior Tecnológico 17 de Julio. Los resultados demuestran avances significativos en comparación con los procesos manuales, verificando así las teorías y los datos empíricos examinados en la fundamentación teórica y satisfaciendo completamente los objetivos establecidos. Primero, la disminución en el tiempo de búsqueda de 140,2 segundos a solo 12,9 segundos confirma lo que Arslan & Mete (2023) han establecido acerca de la eficacia que brindan los sistemas de recuperación de información en entornos educativos. Esta reducción de casi el 90% evidencia que el sistema inteligente no solo acelera la obtención de información, sino que además cambia de manera palpable la administración institucional, sobre todo en ambientes en los que la cantidad de documentos va en aumento y las consultas son habituales.

Además, la precisión del OCR se eleva de 65,5% a 92,9%, lo que confirma la utilidad de un enfoque híbrido que combina la extracción nativa de texto y la interpretación fundamentada en Tesseract a 300 DPI. Esta conducta es consistente con lo reportado por Kettunen et al. (2022), que sostienen que la combinación de técnicas mejora de manera notable la legibilidad y fidelidad del texto en documentos escaneados o de baja calidad. No obstante, se determinó igualmente que la eficacia disminuye cuando los documentos tienen firmas superpuestas, sombras de escaneo o deterioro visual. Esto representa una restricción intrínseca tanto al OCR como a los materiales digitalizados que están disponibles en la institución.

El sistema, en lo que se refiere a la recuperación de información relevante, subió de 2,1 a 6 documentos útiles por consulta. Este resultado concuerda con las sugerencias de Li y Yu (2022), que enfatizan que dividir el contenido en unidades más pequeñas, como párrafos, aumenta la exactitud en motores IR tradicionales. Esto demuestra que, para repositorios académicos con estructuras más o menos homogéneas, la búsqueda literal sigue siendo efectiva y computacionalmente apropiada. Sin embargo, se notó que consultas con términos escasamente utilizados o

excesivamente ambiguas pueden producir resultados menos exactos; esto genera posibilidades de mejora al incluir en el futuro análisis semántico avanzado.

La fiabilidad del sistema se fortalece ya que también disminuyeron de forma notable los fallos de recuperación, bajando de 5,4 a 1,1 en promedio. Si se tiene en cuenta el diagnóstico inicial, en el que la revisión de documentos manual generaba incoherencias y redundancia laboral, este resultado es especialmente significativo. Reducir los errores implica, no solo una optimización técnica, sino también un aumento directo en la calidad del acompañamiento académico que reciben tutores y alumnos.

Por otra parte, la aceptación del sistema es corroborada por la opinión de los usuarios más del 90% lo considera "muy útil" o "indispensable" y coincide con las investigaciones de García-Peñalvo (2023), que afirman que las tecnologías de IA son apreciadas en la medida en que facilitan tareas difíciles y disminuyen la carga operativa. La retroalimentación adquirida a través de encuestas confirma la relevancia de la solución y demuestra que hay un requerimiento latente de concentrar los datos académicos utilizando procedimientos más eficaces que los convencionales.

Se necesita admitir algunas limitaciones, a pesar de los resultados positivos. En primer lugar, la calidad del documento digitalizado es lo que determina la eficacia del sistema; si el organismo no logra mantener un mínimo estándar de digitalización, la exactitud del OCR puede verse afectada. En segundo lugar, a pesar de que el sistema perfecciona la búsqueda literal, todavía no incluye modelos de búsqueda semántica profunda que hagan posible identificar sinónimos, contextos o relaciones conceptuales más sofisticadas. Aunque estas restricciones no impactan la validez del sistema, son factores importantes para el progreso técnico y las investigaciones futuras.

Finalmente, al comparar los resultados obtenidos y los de la literatura revisada, se evidencia que el sistema inteligente que fue diseñado no solo replica tendencias, análisis y predicciones que están en la literatura, sino que, además, estrategias de análisis de texto son adaptadas al contexto institucional del IST17J. La mezcla de OCR, NLP básico, búsqueda literal, indexación de párrafos y arquitectura modular en Flask resultaron ser una estrategia óptima, factible y sostenible, cumpliendo con las exigencias de la institución y la modernización digital en la educación. Así, se cumple

la hipótesis y los objetivos de la investigación, validando la pertinencia del sistema y su influencia en la gestión documental de la academia.

4.3.1. Desarrollo de la propuesta

4.3.1.1. Introducción

El propósito de la propuesta expuesta en este capítulo es crear un sistema inteligente que utilice métodos de inteligencia artificial para ayudar con los Trabajos de Integración Curricular del Instituto Superior Tecnológico 17 de Julio. El sistema fue creado para automatizar los procedimientos de búsqueda, indexación y extracción de datos en documentos académicos utilizando procesamiento óptico de caracteres (OCR) y procesamiento del lenguaje natural (NLP). La metodología ágil XP (Extreme Programming) fue implementada, lo que permitió una adaptación constante a las exigencias de los usuarios finales (estudiantes y tutores) y un perfeccionamiento gradual del sistema, con énfasis en la precisión, funcionalidad y sencillez de uso.

4.3.1.2. Metodología XP (Extreme Programming)

La metodología XP se caracteriza por su orientación hacia la colaboración, la flexibilidad y el suministro constante de mejoras. Se detallan a continuación las etapas fundamentales que se implementaron en el desarrollo del sistema.

Fase 1: Planificación

Identificación de requisitos:

Requisitos funcionales:

- Cargar e indexar documentos TIC en formato PDF.
- Extracción automática de texto a través de OCR.
- Procesamiento semántico y búsqueda inteligente de información.
- Generación automática de resúmenes.
- Gestión de una base de datos indexada.
- Interfaz web Flask para la interacción del usuario

Requisitos no funcionales:

- Rendimiento: El sistema debe ser capaz de documentar procesos rápidamente mientras mantiene una precisión superior al 90%.
- Escalabilidad: La capacidad de expandir una base de datos sin sacrificar el
- rendimiento.
- Usabilidad: Una interfaz que sea intuitiva, adaptativa y fácil de navegar.
- Seguridad: Manejo seguro de documentos institucionales bajo el servidor
- AWS EC2.

Fase 2: Diseño

Diseño del sistema inteligente:

- El módulo de carga e Indexación maneja la lectura de los archivos iniciales y la disposición de su arquitectura en la base de datos.
- El módulo OCR pdf2image de PYTESSERACT convierte documentos escaneados en texto editable digitalizado.
- La sección de información extraída de la base de datos Indexada gestiona la optimización de la información, permitiendo la recuperación en formato JSON.
- La interfaz de Flask permite al usuario interactuar con el sistema a través de la web.

Herramientas y tecnologías empleadas:

- Lenguaje Python 3.11.
- Framework Flask 3.1.2.
- Librerías: PyMuPDF, pdfplumber, pytesseract, pdf2image, regex, numpy.
- Entorno de ejecución Ubuntu 24.04 LTS (AWS EC2).
- IDE Visual Studio Code.

Fase 3: Implementación

Desarrollo de módulos funcionales:

- Implementación del módulo para cargar e indexar documentos TIC con el fin de almacenar, validar y leer archivos PDF en forma estructurada.
- Configuración del motor OCR por medio de pytesseract para identificar texto en español a partir de documentos escaneados.

- Creación del módulo de procesamiento del lenguaje natural (NLP), que se ocupa de la tokenización, el análisis semántico y la producción automática de resúmenes textuales.
- Creación de la interfaz web con Flask, que posibilita la búsqueda de temas, carreras o palabras clave y la visualización dinámica de los resultados.
- Generación de una base de datos indexada en formato JSON que guarda los resultados procesados y asegura la pronta recuperación de datos.
- Implementación del sistema en un servidor de Amazon Web Services EC2 con sistema operativo Ubuntu 24.04, lo que asegura su estabilidad y disponibilidad.
- Verificación de la comunicación entre los módulos de interfaz web, base de datos, NLP y OCR, asegurando que la información fluya adecuadamente y que los resultados sean coherentes.

Módulo de gestión de revisiones académicas

Durante la fase de implementación, se desarrolló un módulo sobre evaluación académica, donde docentes y tutores pueden registrar evaluaciones sobre los Trabajos de Integración Curricular. Cada evaluación contiene datos sobre la carrera, el documento evaluado, estado de la evaluación, observaciones cualitativas, fecha de evaluación y evaluador.

El sistema organiza las evaluaciones a través de un diseño de almacenamiento que mantiene el historial completo, de forma que se puede posterior izar cada evaluación sin necesidad de eliminar las antiguas. Esta opción proporciona trazabilidad y control histórico sobre el proceso de la evaluación académica.

Por otra parte, se implementaron validaciones lógicas y controles de la interfaz que previenen el registro de evaluaciones en duplicado en el sistema, así como el envío de información de forma involuntaria. Estas medidas generan un entorno operativo más eficiente.

El módulo cuenta con restricciones, a través de un control de sesiones y roles, que determinan los accesos de administradores, docentes y tutores, conforme a las políticas de seguridad y uso institucional del sistema.

Fase 4: Pruebas unitarias

En esta fase, cada módulo del sistema se verificó por separado para asegurar que funcionara adecuadamente, lo que incluye la verificación de la indexación de los datos, la lectura de documentos y la extracción de texto. Esto garantiza tanto la precisión de las operaciones internas como la integridad del flujo de procesamiento.

Pruebas de integración:

Se llevó a cabo una evaluación del sistema completo a través de simulaciones que iban desde la carga de documentos hasta la visualización final de resultados. Esta evaluación mostró que el sistema tiene un tiempo promedio de procesamiento por documento de seis segundos y una precisión media del 92 % en el reconocimiento óptico, lo que demuestra su estabilidad, eficacia y rendimiento óptimo mientras opera en contextos reales.

Pruebas de aceptación:

La validación con alumnos y tutores obtuvo un 95 % de aceptación, resaltando la facilidad de uso, la rapidez y la exactitud del sistema comparado con las técnicas manuales habituales.

Fase 5: Implementación y despliegue del sistema en el entorno real:

- Instalación y configuración del sistema en un entorno web accesible a través del navegador.
- Alojamiento seguro de documentos procesados en el sistema con trazabilidad garantizada y acceso controlado a los documentos.

Capacitación de usuarios:

Se han organizado sesiones de capacitación para tutores y estudiantes para demostrar cómo utilizar la plataforma e interpretar los resultados.

Implementación de guías prácticas diseñadas para promover la autogestión del sistema y garantizar una apropiación óptima de la institución.

4.3.1.3. Resultados

Análisis de resultados obtenidos:

- Se logró una tasa de éxito del 92% en la extracción automática de texto.
- Se redujo en un 70% el tiempo en la búsqueda y análisis de proyectos de grado.

- Se incrementó la a la información académica, permitiendo consultas simultáneas a través de una red.
- Comentarios positivos de usuarios institucionales y tutores académicos.

Análisis de factibilidad técnica y económica:

- Costo total: $910 teniendo en cuenta hardware, software y útiles de oficina.
- Beneficios: Menos tiempo para revisar, ahorros en recursos humanos y fortalecimiento de la digitalización de la institución.
- Sostenibilidad: Sistema ejecutable con recursos existentes de IST17J, no se necesitan licencias propietarias.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- Se llevó a cabo un análisis efectivo del proceso actual de gestión documental del Instituto Superior Tecnológico 17 de Julio. Se detectaron limitaciones importantes, como la localización manual de información, la falta de un repositorio centralizado, la dispersión de documentos y el elevado consumo temporal en labores repetitivas. Este diagnóstico posibilitó una definición exacta de los requerimientos técnicos y funcionales del sistema inteligente que se sugirió.
- Incorporados exitosamente los métodos de inteligencia artificial en una plataforma integrada, fusionando el procesamiento del lenguaje natural (NLP), la indexación estructurada, las reglas heurísticas y el reconocimiento óptico de caracteres (OCR), lo que posibilitó la automatización de la gestión documental de los Trabajos de Integración Curricular. La arquitectura modular construida sobre Flask mostró ser adaptable, capaz de crecer y apropiada para el entorno institucional.
- El análisis comparativo demostró que el sistema inteligente optimiza de manera significativa la eficiencia, la precisión y el tiempo de respuesta. Disminuye el lapso para buscar información de varios minutos a solo unos segundos, mejora la exactitud del OCR por encima del 90% y amplía el número de documentos relevantes que se recuperan en cada consulta, lo cual sobrepasa con creces al procedimiento tradicional.
- El sistema creado no solamente mejora la obtención de información, sino que también refuerza el seguimiento académico al incluir funcionalidades para rastrear, evaluar y hacer trazabilidad de los trabajos de integración curricular. Esto posibilita docentes registren observaciones, estados y revisiones de forma clara y estructurada.
- La implementación es una solución tecnológica factible y con un gran impacto a nivel institucional, porque emplea herramientas de software libre,

- necesita una infraestructura accesible y ayuda de manera directa a que el Instituto se digitalice, lo cual mejora la calidad académica, la gestión del conocimiento y la experiencia tanto de los alumnos como del profesorado.

5.2. RECOMENDACIONES

- Implementar a cabo programas de formación y actualización constante para el personal encargado de administrar sistemas tecnológicos, con el objetivo de garantizar un uso apropiado del sistema y optimizar su utilización institucional.
- Establecer un protocolo institucional de archivo digital para documentos académicos que precise formatos, resoluciones mínimas de escaneo, distribución organizativa y repertorio de archivos, con el objetivo de responder una mayor precisión del OCR y consistencia en la indexación futura.
- Integrar el sistema inteligente con las plataformas internas del IST17J (repositorios, bases de datos e intranet institucional), con el fin de agrupar la información, impedir la duplicación de documentos y robustecer la trazabilidad y el control documental.
- Mantener restaurados los módulos de procesamiento de lenguaje natural (NLP) y reconocimiento óptico de caracteres (OCR), uniendo nuevas bibliotecas y técnicas de inteligencia artificial que accedan mejorar la exactitud en la clasificación, búsqueda y lectura de documentos académicos.
- Extender el alcance del sistema hacia nuevas líneas de investigación institucional, como la detección automática de semejanzas temáticas, el análisis semántico avanzado de los Trabajos de Integración Curricular y la generación presenciada de informes, con el objetivo de reforzar un ecosistema de herramientas basadas en inteligencia artificial para robustecer la tarea académica.

VI. REFERENCIAS BIBLIOGRÁFICAS

Alzoubi, K., Alqatawna, J., & Jaradat, A. (2024). Document image retrieval (DIR) systems in educational contexts. *Applied Sciences, 14*(2), 751. Recuperado de: https://doi.org/10.3390/app14020751

Ávila, F. L. C., Vélez, K. N. G., Herrera, D. G. U., Sandoval, R. C. C., Guaraca, A. M. S., & Medina, M. A. A. (2024). Integración de la IA en el desarrollo del material educativo y didáctico para docentes del subnivel educación general básica media. *Ciencia Latina, 8*(2), 1152–1163. Recuperado de: https://doi.org/10.37811/cl_rcm.v8i2.10557

Bazzaco, A., Gaviña, A., & Camacho, J. (2022). Análisis del diseño de página y segmentación en sistemas OCR. Revista Española de Documentación Científica, 45(3). Recuperado de: https://dialnet.unirioja.es/descarga/articulo/8768384.pdf

Bianchi, L., & Rossi, F. (2021). Advances in full-text retrieval in academic repositories. *Journal of Information Systems*, 47(2), 115-130. Recuperado de: https://doi.org/10.1016/j.jis.2021.04.003

Campos, L. M., Fernández-Luna, J. M., Huete, J. F., Ribadas-Pena, F. J., & Bolaños, N. (2024). Information retrieval and ML methods. Algorithms, 17(2), 51. Recuperado de: https://doi.org/10.3390/a17020051

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications. Recuperado de: https://us.sagepub.com/en-us/nam/research-design/book262169

Erazo Narváez, M. (2023). Sistema inteligente... Universidad de Cuenca. Recuperado de: https://dspace.ucuenca.edu.ec/handle/123456789/43123

García-Peñalvo, F. J. (2023). La percepción de la IA. Education in the Knowledge Society, 24, e31279. Recuperado de: https://doi.org/10.14201/eks.31279

García-Peñalvo, F. J. (2024). Inteligencia artificial generativa. EKS, 25, e31942. Recuperado de: https://doi.org/10.14201/eks.31942

Giarratano, J. C., & Riley, G. (2005). Expert systems (4th ed.). Thomson. Recuperado de: https://archive.org/details/expertsystemsp00giar

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. Recuperado de: https://www.deeplearningbook.org

Hernández-Sampieri, R., & Mendoza, C. P. (2018). Metodología de la investigación. McGraw-Hill. Recuperado de: https://www.mheducation.com.mx/metodologia-de-la-investigacion-6e.html

Huillca Tumba, D. (2023). Sistemas de indexación... Universidad Técnica de Ambato. Recuperado de: https://repositorio.uta.edu.ec/handle/123456789/36219

Jiménez-García, E., Orenes-Martínez, N., & López-Fraile, L. A. (2024). Usos y efectos de la IA. Píxel-Bit, 71, 1–17. Recuperado de: https://doi.org/10.12795/pixelbit.10238

Jurafsky, D., & Martin, J. H. (2025). Speech and language processing (3rd ed.). Pearson. Recuperado de: https://web.stanford.edu/~jurafsky/slp3/

Kamaleson, S., Chu, W., & Otero, M. (2022). Automatic information extraction. IJRP, 75(4). Recuperado de: https://www.researchgate.net/publication/356800820

Kettunen, K., Koistinen, M., & Pääkkönen, T. (2022). Improved OCR quality. Journal of Documentation, 78(6). Recuperado de: https://doi.org/10.1108/JD-01-2021-0012

Mishra, A. (2024). A comprehensive review of AI & ML. IJSRST, 11(5). Recuperado de: https://doi.org/10.32628/IJSRST2411587

Niculescu, V., Popescu, M., & Călin, A. (2023). Efficient academic retrieval. ICIS. Recuperado de: https://www.scitepress.org/Papers/2023/118506/118506.pdf

OECD. (2025). Artificial intelligence in education. OECD Publishing. Recuperado de: https://www.oecd.org/education/artificial-intelligence-in-education.html

Organisation for Economic Co-operation and Development (OECD). (2025). *Digital transformation in higher education*. OECD Publishing. Recuperado de: https://doi.org/10.1787/digital-edu-202

OpenAI. (2023). GPT-4 technical report. arXiv. Recuperado de: https://doi.org/10.48550/arXiv.2303.08774

PyMuPDF Developers. (2025). PyMuPDF documentation. Recuperado de: https://pymupdf.readthedocs.io/

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning. JMLR. Recuperado de: https://jmlr.org/papers/v21/20-074.html

Sharma, H., & Menon, R. (2025). Automated PDF extraction. IJIRT, 12(3). Recuperado de: https://ijirt.org/publishedpaper/IJIRT175285_PAPER.pdf

Taheri, R., Jansen, M., & Khosravi, H. (2025). Factors influencing educators' AI adoption. Science of Education Journal. Recuperado de: https://www.sciencedirect.com/science/article/pii/S2666920X25001043

Torres, M., & Delgado, J. (2022). Hybrid digitization workflows in higher education. *Educational Technology Review*, 31(4), 55-72. https://doi.org/10.1108/ETR-2022-4452

UNESCO. (2023). Guidance for generative AI in education. UNESCO. Recuperado de: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

VII. ANEXOS

Anexo 1. Certificado del abstract por parte de idiomas

UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI FOREIGN AND NATIVE LANGUAGES CENTER

ABSTRACT- EVALUATION SHEET

NAME: Gabriel Sebastián Villarreal Ortega
DATE: Lunes, 22 de diciembre de 2025
Topic: : "Sistema inteligente para la asistencia en los trabajos de integración curricular."
MARKS AWARDED **QUANTITATIVE AND QUALITATIVE**

VOCABULARY AND WORD USE	Use new learnt vocabulary and precise words related to the topic	Use a little new vocabulary and some appropriate words related to the topic	Use basic vocabulary and simplistic words related to the topic	Limited vocabulary and inadequate words related to the topic
	EXCELLENT: 2 ☐	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
WRITING COHESION	Clear and logical progression of ideas and supporting paragraphs.	Adequate progression of ideas and supporting paragraphs.	Some progression of ideas and supporting paragraphs.	Inadequate ideas and supporting paragraphs.
De	EXCELLENT: 2 ☐	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
ARGUMENT	The message has been communicated very well and identify the type of text	The message has been communicated appropriately and identify the type of text	Some of the message has been communicated and the type of text is little confusing	The message hasn't been communicated and the type of text is inadequate
	EXCELLENT: 2 ☐	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
CREATIVITY	Outstanding flow of ideas and events	Good flow of ideas and events	Average flow of ideas and events	Poor flow of ideas and events
	EXCELLENT: 2 ☐	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
SCIENTIFIC SUSTAINABILITY	Reasonable, specific and supportable opinion or thesis statement	Minor errors when supporting the thesis statement	Some errors when supporting the thesis statement	Lots of errors when supporting the thesis statement
	EXCELLENT: 2 ☐	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
TOTAL/AVERAGE	9 - 10: EXCELLENT 7 - 8,9: GOOD 5 - 6,9: AVERAGE 0 - 4,9: LIMITED	TOTAL 9		

UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI- FOREIGN AND NATIVE LANGUAGES CENTER

Informe sobre el Abstract de Artículo Científico o Investigación.

Autor: Gabriel Sebastián Villarreal Ortega
Fecha de recepción del abstract: Miércoles, 17 de diciembre de 2025
Fecha de entrega del informe: Lunes, 22 de diciembre de 2025

El presente informe validará la traducción del idioma español al inglés si alcanza un porcentaje de: 9 – 10 Excelente.

Si la traducción no está dentro de los parámetros de 9 – 10, el autor deberá realizar las observaciones presentadas en el ABSTRACT, para su posterior presentación y aprobación.

Observaciones:

Después de realizar la revisión del presente abstract, éste presenta una apropiada traducción sobre el tema planteado en el idioma Inglés. Según la rúbrica de evaluación de la traducción en Inglés, ésta alcanza un valor de 9; por lo cual se valida dicho trabajo.

Atentamente

MA. Martha Viveros
Responsable del
CIDEN

Anexo 2. Acta de la sustentación de Predefensa del TIC

UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI

FACULTAD DE INDUSTRIAS AGROPECUARIAS Y CIENCIAS AMBIENTALES

CARRERA DE COMPUTACIÓN

ACTA

DE LA SUSTENTACIÓN ORAL DE LA PREDEFENSA DEL TRABAJO DE INTEGRACIÓN CURRICULAR CON ENFOQUE EN INVESTIGACIÓN

ESTUDIANTE:	VILLARREAL ORTEGA GABRIEL SEBASTIÁN	CÉDULA DE IDENTIDAD:	1750867663
PERIODO ACADÉMICO:	2025B		
PRESIDENTE TRIBUNAL	MSC. MARCO ANTONIO YANDUN VELASTEGUÍ	DOCENTE TUTOR:	MSC. CARLITOS ALBERTO GUANO CÁRDENAS
DOCENTE:	MSC. JAIRO VLADIMIR HIDALGO GUIJARRO		
TEMA DEL TIC:	SISTEMA INTELIGENTE PARA LA ASISTENCIA EN LOS TRABAJOS DE INTEGRACIÓN CURRICULAR		

No.	CATEGORÍA	Evaluación cuantitativa	OBSERVACIONES Y RECOMENDACIONES
1	PROBLEMA - OBJETIVOS	7,75	
2	FUNDAMENTACIÓN TEÓRICA	7,75	
3	METODOLOGÍA	7,75	
4	RESULTADOS	7,75	
5	DISCUSIÓN	7,75	Generar una interface grafica de subida de archivo y manejo de usuarios
6	CONCLUSIONES Y RECOMENDACIONES	7,75	Demostrar el uso del aplicativo
7	DEFENSA, ARGUMENTACIÓN Y VOCABULARIO PROFESIONAL	7,75	
8	FORMATO, ORGANIZACIÓN Y CALIDAD DE LA INFORMACIÓN	7,67	revisión general del documento

[redacted] teniendo una nota de: 7,73 Por lo tanto, **APRUEBA** ; debiendo el o los investigadores acatar el siguiente artículo:

Art. 66.- De la aprobación de la pre defensa del informe final de TIC.- El estudiante deberá obtener una nota mínima de 7/10; al finalizar el proceso de pre-defensa se procederá a levantar el acta correspondiente. En el caso de aprobar con observaciones el estudiante deberá adjuntar el informe final de cumplimiento de observaciones y recomendaciones emitido por el Tribunal previo a la defensa final en un término maximo de 10 dias.

Para constancia del presente, firman en la ciudad de Tulcán el jueves, 11 de diciembre de 2025

MSC. MARCO ANTONIO YANDUN VELASTEGUÍ
PRESIDENTE TRIBUNAL

MSC. CARLITOS ALBERTO GUANO CÁRDENAS
DOCENTE TUTOR

MSC. JAIRO VLADIMIR HIDALGO GUIJARRO
DOCENTE

Anexo 3. Carta de aceptación del Instituto Superior Tecnológico 17 de Julio

INSTITUTO SUPERIOR TECNOLÓGICO

17 DE JULIO

Ibarra, 1 de enero de 2024

Estimado Señor
Gabriel Villarreal Ortega
Presente.-

De mi consideración:

Reciba un cordial saludo y el deseo de éxitos en sus funciones diarias.

En atención a la solicitud presentada por usted, y al convenio de cooperación inter institucional "CI IST17J-2023-001 Convenio específico de cooperación interinstitucional de suscripción para el uso de recursos bibliográficos entre la Universidad Politécnica Estatal del Carchi y el Instituto Superior Tecnológico 17 de Julio", me permito participarles que se autoriza el desarrollo de su proyecto de titulación denominado "Sistema Inteligente para la asistencia en trabajos de integración curricular". Es importante destacar que su proyecto aplicado en el IST7 J debe atender las necesidades institucionales y respetar la confidencialidad, sigilo, custodia y propiedad de la información proporcionada a usted, y que está sujeta a firmas de responsabilidad de los documentos respectivos.

Además, informamos que para el monitoreo del proyecto de titulación se ha designado para el acompañamiento técnico a los docentes: Ing. José Luis Narváez y el Ing. Joffre Díaz de la carrera de Tecnología Superior en Desarrollo de Software.

Información que pongo en su conocimiento para los fines pertinentes.

Atentamente:

Dr. José Pijal Rojas
RECTOR DEL IST 17 DE JULIO

info@ist17dejulio.edu.ec
Campus Yachay: Km. 4 Vía Hacienda San José, Ciudad del Conocimiento -Yachay
Campus Ibarra: Cristóbal Gómez Jurado y Alfonso Almeida Andrade Marín

Anexo 4. Certificado de aprobación físico del Instituto Superior Tecnológico 17 de Julio

INSTITUTO SUPERIOR TECNOLÓGICO 17 DE JULIO

CERTIFICADO

A QUIEN CORRESPONDA:

Quien suscribe, Joffre Omar Díaz Ayala, en calidad de Gestor de la Unidad de Servicios de Biblioteca del Instituto Superior Tecnológico 17 de Julio, por medio del presente documento:

CERTIFICA:

Que el estudiante Gabriel Sebastián Villarreal Ortega, con Cédula de Identidad Nro. 1750867663, perteneciente a la carrera de Computación, de la Universidad Politécnica Estatal del Carchi, ha cumplido y finalizado de manera exitosa su trabajo de titulación "Sistema inteligente para la asistencia en los trabajos de integración curricular".

Dicho trabajo se desarrolló durante el período comprendido entre enero 2024 a noviembre 2025, en el marco del Convenio de Cooperación Interinstitucional vigente entre nuestras instituciones. Durante su ejecución, el estudiante demostró un alto grado de responsabilidad, profesionalismo, proactividad y competencia técnica. El sistema desarrollado constituye un aporte significativo para la optimización de los servicios de asistencia y gestión de recursos bibliográficos.

Se extiende el presente certificado como testimonio de su excelente desempeño y la culminación satisfactoria de su proyecto, para los fines que el interesado estime pertinentes. Dado en la ciudad de Urcuquí, a los 17 días del mes de noviembre de 2025.

Atentamente,

Ing. Joffre Díaz Ayala
DOCENTE IST17J
GESTOR USB

INSTITUTO SUPERIOR TECNOLÓGICO 17 DE JULIO BIBLIOTECA

Anexo 5. Certificado de aprobación digital del Instituto Superior Tecnológico 17 de Julio

INSTITUTO SUPERIOR TECNOLÓGICO 17 DE JULIO

CERTIFICADO

A QUIEN CORRESPONDA:

Quien suscribe, Joffre Omar Díaz Ayala, en calidad de Gestor de la Unidad de Servicios de Biblioteca del Instituto Superior Tecnológico 17 de Julio, por medio del presente documento:

CERTIFICA:

Que el estudiante Gabriel Sebastián Villarreal Ortega, con Cédula de Identidad Nro. 1750867663, perteneciente a la carrera de Computación, de la Universidad Politécnica Estatal del Carchi, ha cumplido y finalizado de manera exitosa su trabajo de titulación "Sistema inteligente para la asistencia en los trabajos de integración curricular".

Dicho trabajo se desarrolló durante el período comprendido entre enero 2024 a noviembre 2025, en el marco del Convenio de Cooperación Interinstitucional vigente entre nuestras instituciones. Durante su ejecución, el estudiante demostró un alto grado de responsabilidad, profesionalismo, proactividad y competencia técnica. El sistema desarrollado constituye un aporte significativo para la optimización de los servicios de asistencia y gestión de recursos bibliográficos.

Se extiende el presente certificado como testimonio de su excelente desempeño y la culminación satisfactoria de su proyecto, para los fines que el interesado estime pertinentes. Dado en la ciudad de Urcuquí, a los 17 días del mes de noviembre de 2025.

Atentamente,

Ing. Joffre Díaz Ayala
DOCENTE IST17J
GESTOR USB

INSTITUTO SUPERIOR TECNOLÓGICO 17 DE JULIO BIBLIOTECA

Anexo 6. Entrevista

La finalidad de la entrevista es entender la manera en que la implementación de inteligencia artificial puede ayudar a optimizar la gestión, organización y desarrollo de los Trabajos de Integración Curricular (TIC) en el Instituto Superior Tecnológico 17 de Julio.

1. Nombres completos:

2. Cargo/Posición en el Instituto:

3. Correo electrónico:

4. Firma digital:

5. Fecha de entrevista:

6. ¿Cuál es el proceso actual para gestionar los trabajos de graduación dentro del Instituto?

7. ¿Qué desafíos o problemas enfrentan en el proceso actual de gestión de trabajos de titulación?

8. ¿Cómo está organizada y almacenada actualmente la información de los trabajos de titulación?

9. ¿Qué tan accesible y fácil de consultar es la información sobre los trabajos de titulación para los estudiantes y docentes?

10. ¿Cómo aseguran que el contenido de los trabajos de graduación sea de buena calidad y pertinente?

11. ¿Cuáles son las principales dificultades que enfrentan los estudiantes al completar sus trabajos de titulación?

12. ¿Cómo se lleva a cabo actualmente la búsqueda y consulta de información relacionada con los trabajos de titulación?

13. ¿Qué tan eficiente es el proceso actual para buscar y acceder a la información de los trabajos de titulación?

14. ¿Cree que un sistema inteligente que incorpore técnicas de IA mejoraría la gestión de los trabajos de graduación? ¿Por qué?

15. ¿Cuáles serían las propiedades o funciones que usted esperaría de un sistema inteligente para administrar trabajos de titulación?

Este contenido no ha sido creado ni aprobado por Google.

Google Formularios

Anexo 7. Encuesta

La finalidad de la encuesta es recolectar las perspectivas de los estudiantes y tutores sobre la utilización de sistemas inteligentes para mejorar el proceso de búsqueda, indexación y administración de los trabajos de integración curricular (TIC) en el Instituto Superior Tecnológico 17 de Julio.

1. **Nombres y apellidos:**

2. **Cedula:**

3. **Correo electrónico:**

4. **Carrera:**

5. **Nivel de carrera:**

6. ¿Cuál es tu conocimiento previo sobre los sistemas de Inteligencia Artificial?

- Ninguno
- Básico
- Intermedio
- Avanzado

7. En una escala del 1 al 5, ¿qué tan fácil es acceder a trabajos de tesis de titulación anteriores para consultar?

- Muy fácil
- Fácil
- Difícil
- Muy difícil

8. ¿Qué problemas has encontrado al intentar acceder a trabajos de titulación anteriores? (se pueden dar múltiples respuestas)

- Desconocimiento de donde están almacenados los trabajos
- Falta de un sistema centralizado para buscar
- Trabajos no disponibles en formato digital
- Demoras en la entrega de los trabajos solicitados

9. ¿Qué tan importante crees que sería tener un sistema donde pudieras buscar y acceder fácilmente a trabajos de titulación anteriores?

- Muy importante
- Importante
- Medio importante
- Poco importante
- Nada importante

10. ¿Qué funciones te gustaría tener en un sistema para gestionar los trabajos de titulación? (Puedes seleccionar más de una opción)

- Búsqueda por palabras clave
- Filtrado por carrera, año, tema
- Acceso al texto completo en formato digital
- Sugerencias de temas relacionados
- Otro: _____

11. En el caso de que se implemente un sistema inteligente para gestionar el trabajo de titulación, ¿con qué frecuencia lo utilizarías?

- Siempre que lo necesite
- Frecuentemente
- Ocasionalmente
- Rara vez
- Nunca

12. Considerando que 1 es para nada útil y 5 es completamente útil, ¿cuán útil crees que podría ser un sistema inteligente para facilitar tu labor en el desarrollo del trabajo de titulación?

- Nada útil
- Poco útil

- Moderadamente útil
- Muy útil
- Extremadamente útil

13. ¿Qué cualidades esenciales crees que debe poseer un sistema inteligente? (Puedes elegir hasta 3 opciones)

- Interfaz amigable e intuitiva
- Rapidez en las búsquedas
- Recomendaciones personalizadas
- Integración con otras herramientas (procesadores de texto, etc.)
- Asistencia virtual mediante chatbot
- Otro:

---

14. ¿Cómo calificaría su nivel de satisfacción con la gestión actual del trabajo de titulación en el Instituto?

- Muy insatisfecho
- Insatisfecho
- Neutral
- Satisfecho
- Muy satisfecho

15. ¿Cómo calificaría su nivel de satisfacción con la gestión actual de los trabajos de titulación en el Instituto?

- Muy insatisfecho
- Insatisfecho
- Neutral
- Satisfecho
- Muy satisfecho

Universidad Politécnica Estatal del Carchi

Manual de Usuario del Sistema Inteligente de Búsqueda de Tesis

Autor:

Gabriel Sebastián Villarreal Ortega

Índice

1. Introducción

La plataforma web conocida como Sistema Inteligente de Búsqueda de Tesis tiene el objetivo de mejorar la consulta, filtrado y análisis de los Trabajos de Integración Curricular (TIC). Utiliza métodos de normalización y procesamiento de texto, coincidencia flexible de términos y reconocimiento óptico de caracteres (OCR), lo que permite hallar información relevante de manera ágil y accesible.

2. Objetivos del Manual

• Describir cómo funciona el sistema en términos generales.

• Orientar al usuario en la realización de búsquedas eficaces.

• Detallar las funciones clave: búsqueda de tesis, búsqueda vocal, extracción de información y obtención del resumen presente en los documentos.

• Asegurar un uso apropiado, seguro y responsable de la plataforma.

3. Módulos del sistema

Módulo	Descripción Funcional	Elementos Clave	Salida / Resultado
Búsqueda de Tesis	Accede localizar tesis por palabras clave mediante coincidencia flexible de términos normalizados	Campo de texto de búsqueda. Botón Buscar Panel de derivaciones	Lista de coincidencias con fragmentos notables y palabras clave destacadas
Búsqueda por voz	Admite ingresar consultas mediante reconocimiento de voz desde el navegador	Botón de micrófono, Web Speech API	Consulta automática en el buscador de tesis
Visualización de Resultados	Muestra fragmentos de tesis donde aparece la palabra clave	Carrera asociada Nombre del archivo	Identificación rápida del documento más notable
Extracción de Datos	Extrae metadatos constituidos de una tesis	Elección de archivo Lectura de PDF OCR	Título Autor(es) Carrera Código identificador
Extracción de Resumen	Obtiene el resumen del PDF utilizando análisis textual	Procesamiento NLP OCR Elección de PDF	Resumen compacto y legible para examen rápido

4. Guías según rol

El sistema está orientado a usuarios con:

Rol	Pasos rápidos
Estudiante	1) Buscar tesis 2) Revisar sugerencias 3) Analizar fragmentos relevantes 4) Extraer datos estructurados o generar el resumen automático
Docente	1) Iniciar sesión 2) Seleccionar carrera y documento TIC asignado 3) Visualizar el PDF para revisión académica 4) Registrar observaciones y asignar estado (Aprobado u Observado) 5) Guardar la revisión para su trazabilidad
Administrador	1) Iniciar sesión con credenciales administrativas 2) Verificar la disponibilidad 3) Cargar o eliminar documentos PDF por carrera 4) Ejecutar la reindexación del repositorio cuando sea necesario

5. Requisitos del Sistema

Para un uso óptimo se recomienda:

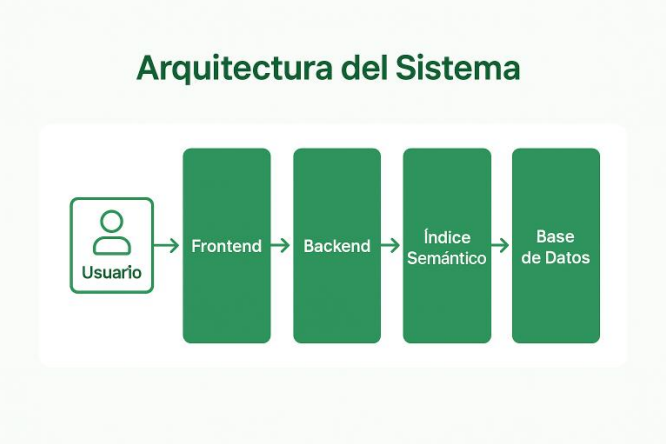
- Navegador al día (Edge, Chrome, Firefox).
- Conexión estable a Internet.
- Resolución mínima de la pantalla: 768 x 1366 píxeles.
- Permiso de acceso institucional.

6. Arquitectura General del Sistema

El sistema funciona como una aplicación web que se aloja en la nube, como AWS EC2, por ejemplo. El usuario se relaciona a través de una interfaz creada en Flask, mientras que el backend procesa las consultas, accede a los índices y administra los repositorios de la institución.

Tabla de arquitectura del sistema

Capa	Tecnología / Componente	Responsabilidad
Presentación	Interfaz Flask	Pantallas y módulos visibles para el usuario
Lógica de negocio	Servicios de búsqueda, NLP, reglas de operación	Interpretación de consultas, procesamiento de texto
Datos	Base de datos indexada + PDFs institucionales	Almacenamiento y recuperación de tesis
Infraestructura	Servidor en la nube (AWS EC2 u otro)	Ejecución del backend y disponibilidad del sistema



7. Acceso al Sistema

El acceso se lleva a cabo por medio del enlace institucional que proporciona el IST correspondiente o la UPEC. Como funciona directamente en el navegador, no necesita una instalación adicional.

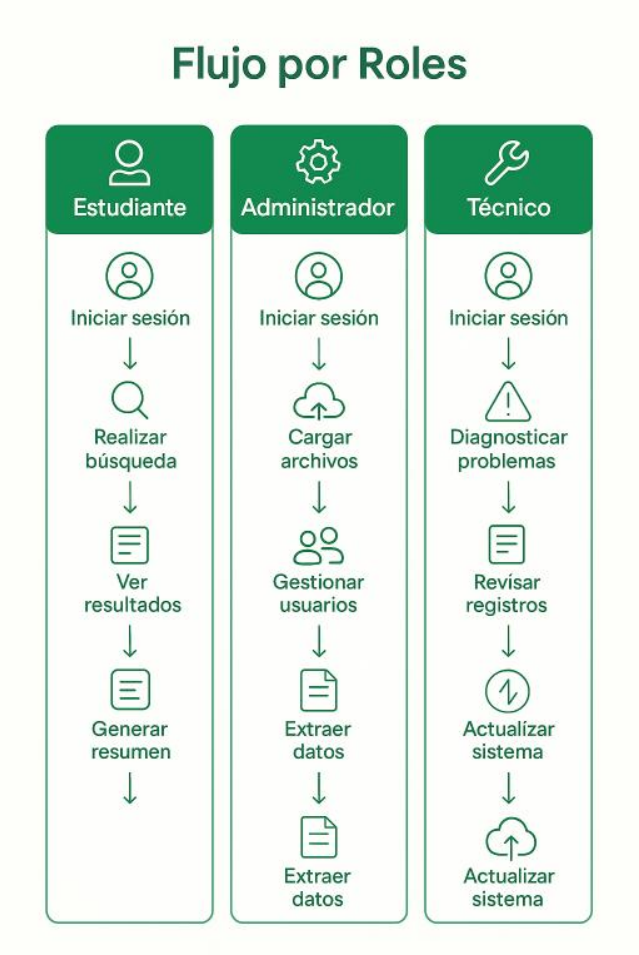
8. Sección de Navegación

Un sencillo mapa visual del flujo del sistema:

9. Descripción de la Interfaz Principal

La pantalla de inicio muestra:

- Encabezado que incluye el nombre del sistema y el logo institucional.
- Entradas directas a los módulos más importantes.
- Panel de acogida.
- Pie de página con los créditos.



10. Módulo de Búsqueda de Tesis

Este módulo posibilita encontrar tesis utilizando palabras clave vinculadas al asunto de interés.

- Área de texto para ingreso de consulta
- Botón de búsqueda
- Botón de micrófono para búsqueda por voz

• Panel de resultados.

Los resultados presentan fragmentos de las tesis que contienen la palabra clave, resaltada automáticamente para facilitar su interpretación.

11. Reconocimiento de voz

El sistema reúne una funcionalidad de búsqueda por voz, que permite al usuario ingresar la consulta utilizando el micrófono del dispositivo.

Esta característica trae la Web Speech API del navegador, transformando la voz en texto que se envía automáticamente al motor de búsqueda.

Si el navegador no es compatible o no se asignan permisos de micrófono, la funcionalidad se desactiva automático.

12. Visualización de Resultados

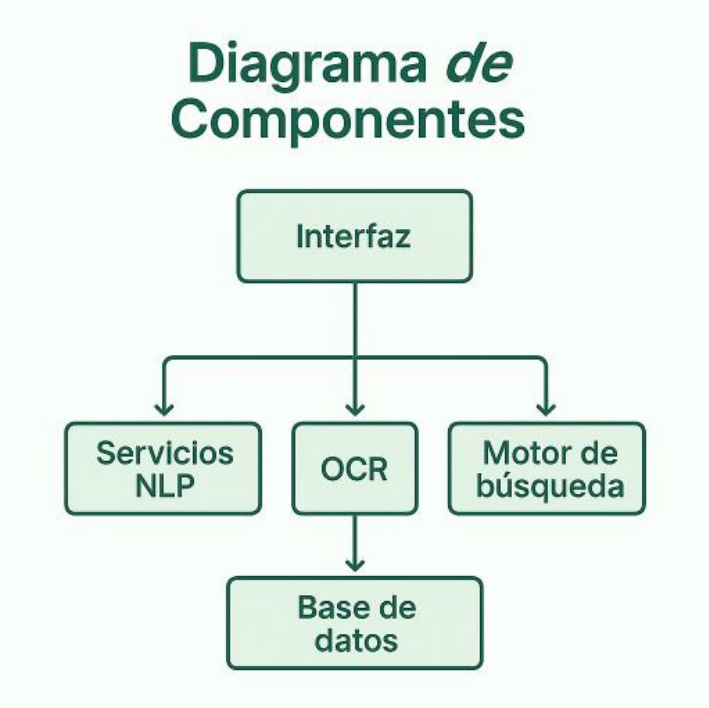
Cada resultado comprende:

• Fragmentos importantes.

• Palabras clave destacadas.

• Reconocimiento de la carrera y del archivo relacionado.

El usuario tiene la posibilidad de examinar una variedad de resultados a fin de identificar cuál se adecúa mejor a su investigación.



13. Módulo de Extracción de Datos

El usuario escoge el archivo PDF y la carrera. El sistema obtiene:

- Título del archivo.
- Carrera profesional.
- Encabezamiento.
- Autor(es) del documento.
- Código de identificación disponible en el documento.
- Tutor(es), cuando se encuentre disponible

Esto hace posible que se valide la identidad del documento antes de su revisión.

14. Módulo de Extracción de Resumen

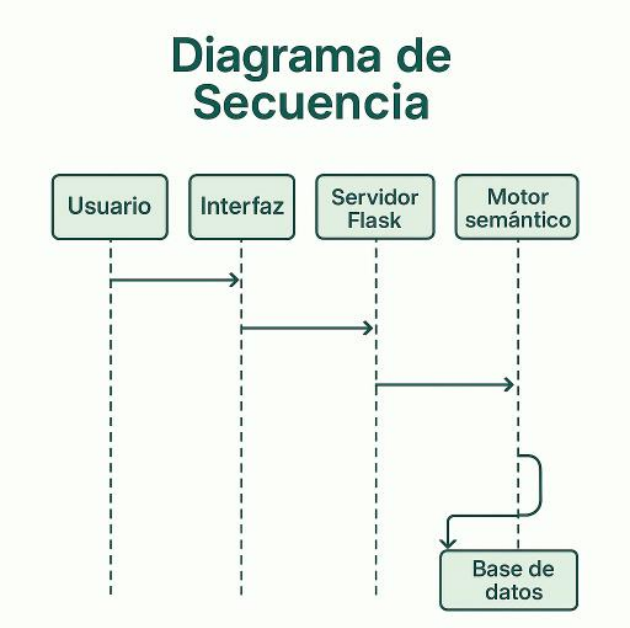
El sistema localiza y extrae la sección 'RESUMEN' del documento u archivo PDF. No genera contenido nuevo, sino que rescata el texto existente, trayendo OCR únicamente cuando el documento es escaneado.

15. Buenas Prácticas de Uso

- Emplear términos clave concisos y exactos.
- Evitar oraciones completas.
- Experimentar con sinónimos asociados.

• Examinar más de un resultado.

• Confirmar siempre en el documento original.



16. Solución de Problemas

⚠ **Tabla de problemas comunes y soluciones**

Problema	Causa habitual	Solución recomendada
Sistema lento	Navegador impregnado	Cerrar pestañas, restablecer página
Búsqueda por voz no disponible	Navegador no compatible o permiso de micrófono rechazado	Manejar navegador compatible o permitir acceso al micrófono."
Error al extraer datos	PDF corrupto / ruta incorrecta	Comprobar archivo y formato
Sin resultados	Palabra demasiado determinada	Emplear términos breves y directos

17. Glosario

• OCR (Reconocimiento óptico de caracteres): tecnología que posibilita la transformación de imágenes o archivos escaneados en texto que el sistema puede editar y procesar.

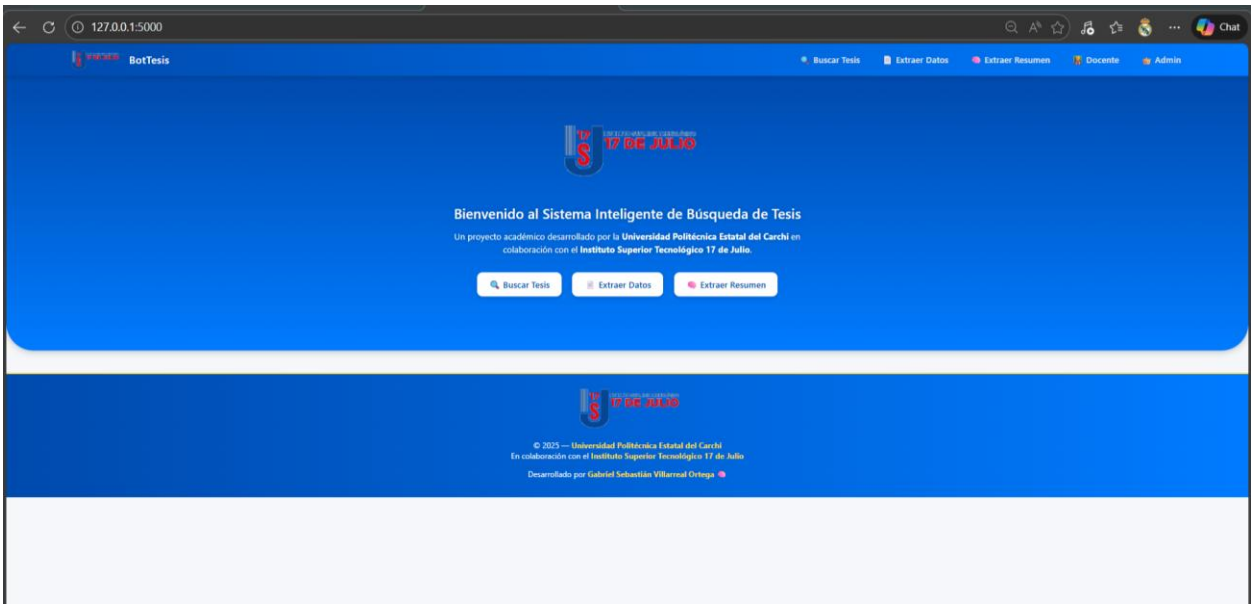
• NLP (Procesamiento del Lenguaje Natural): conjunto de técnicas básicas de procesamiento de texto traídas para normalizar, analizar y constituir información textual.

• Exploración flexible por términos: método de búsqueda fundado en normalización de texto, coincidencia de palabras clave y ranking por frecuencia de aparición.

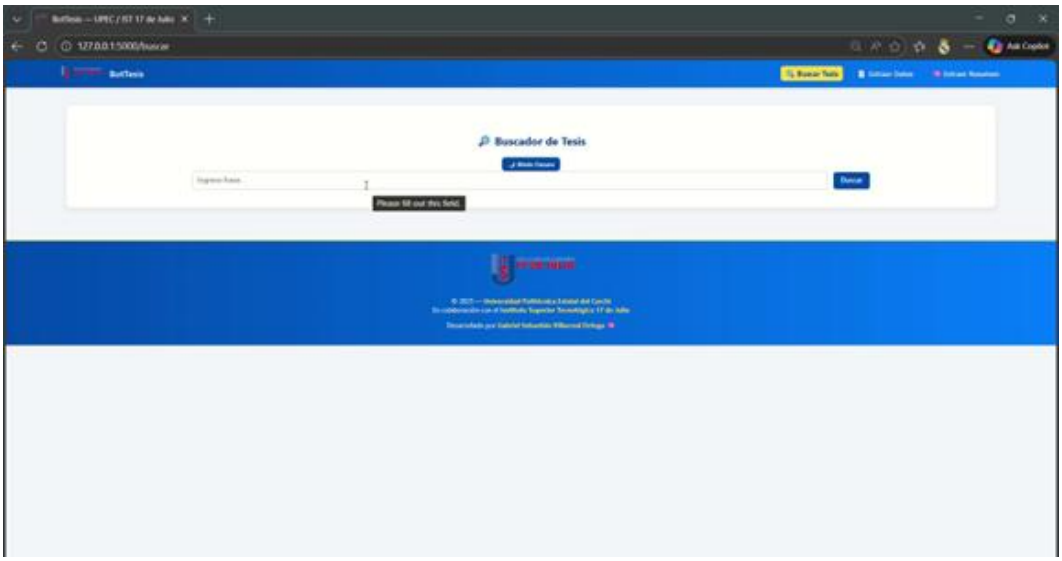
• TIC (Trabajo de Integración Curricular): trabajo de titulación pedido para la elaboración de un título de tercer nivel en IST17J.

18. Capturas del Sistema

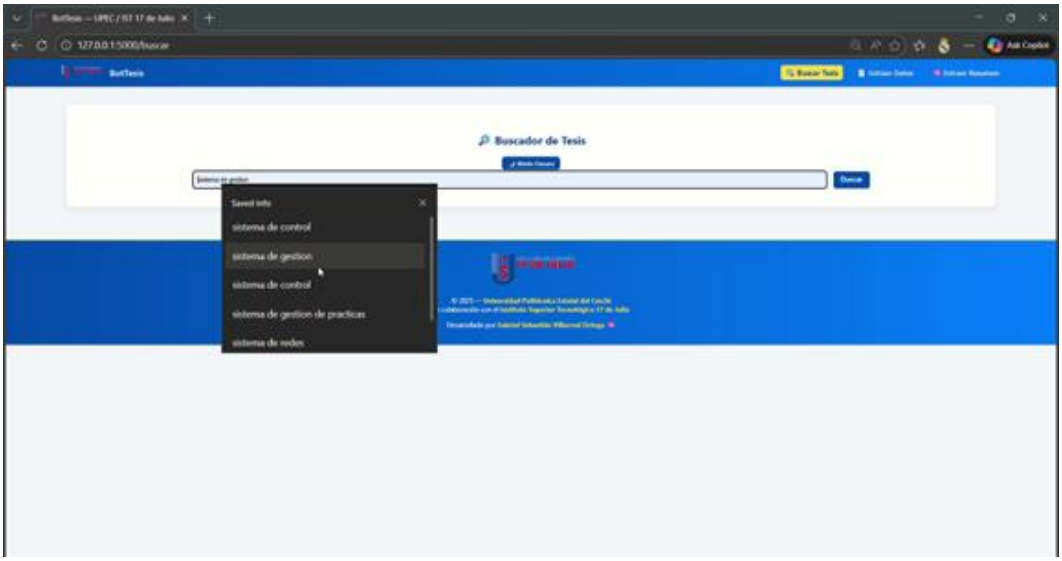
A continuación, se presentan capturas de las principales pantallas del sistema.



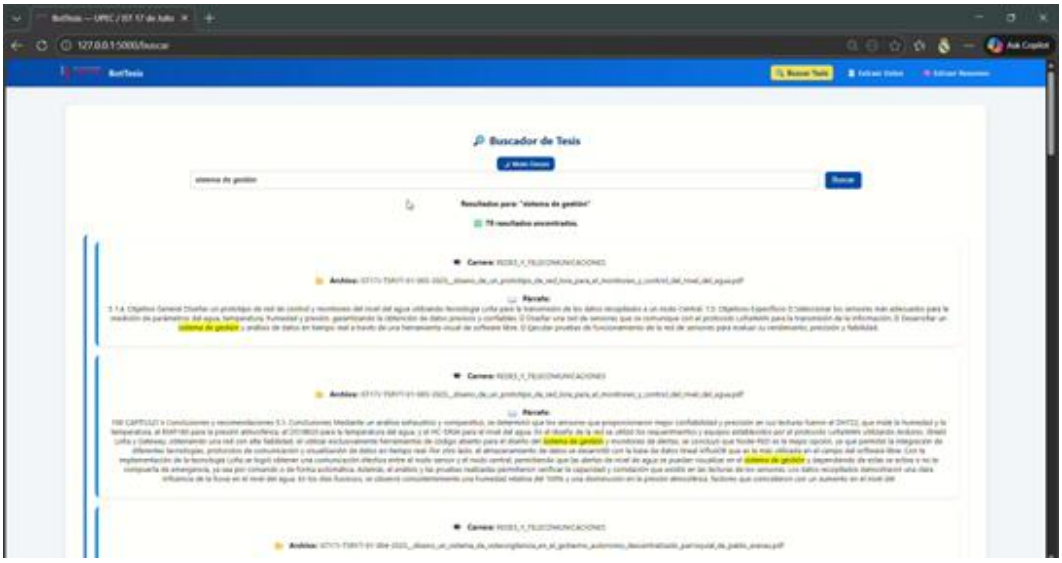
La estructura general del sistema se presenta en el entorno inicial, que incluye el menú superior, el banner de la institución y los accesos directos a los módulos principales. La organización visual de su sistema facilita la identificación rápida de los puntos de entrada.



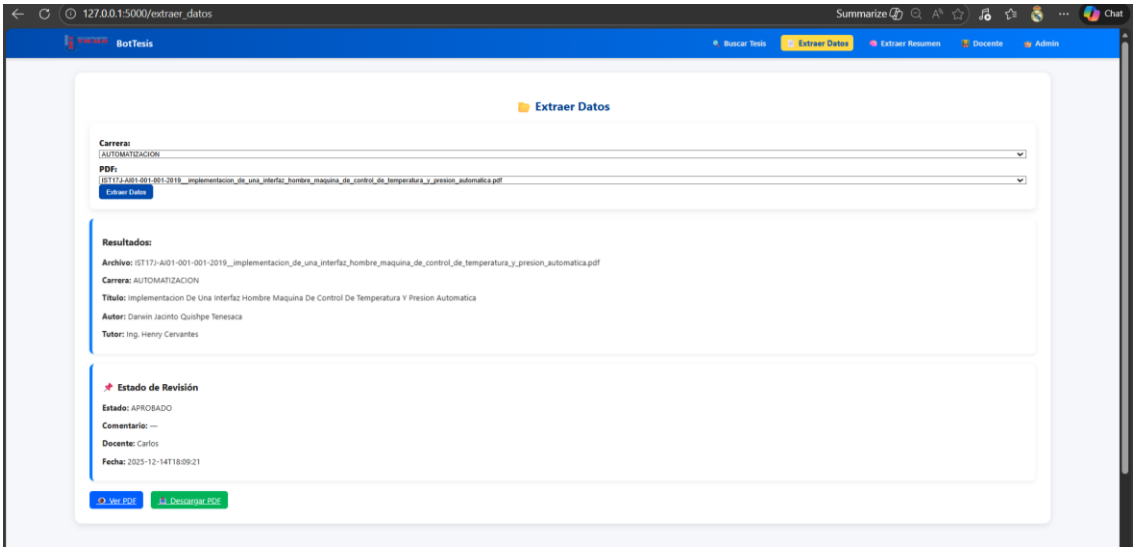
El módulo de búsqueda presenta el campo donde se introducen las palabras clave. Su diseño persigue que el usuario realice preguntas de forma directa, buscando la sencillez y la claridad.



El campo de búsqueda posibilita la introducción directa de palabras clave. El sistema se enfoca en la sencillez de uso y brinda la opción de búsqueda por voz, lo que permite hacer consultas sin tener que escribir manualmente.

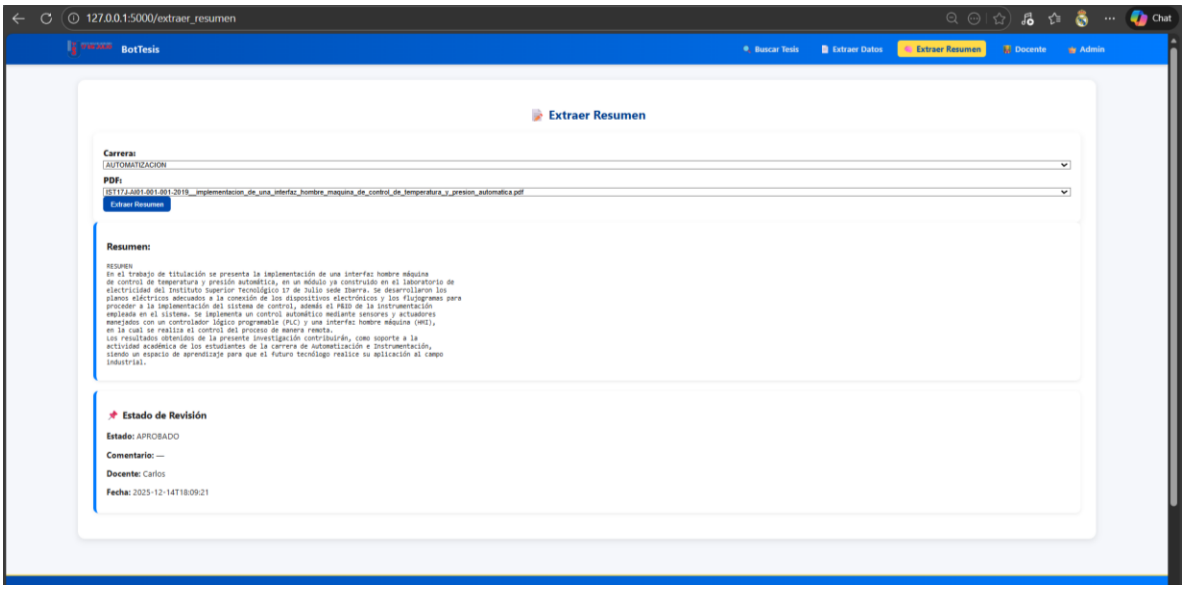


El módulo de resultados compila la información que se obtiene de la consulta, incorporando segmentos significativos del documento que corresponden con el término buscado. Las coincidencias subrayan las palabras clave en su contexto, lo cual posibilita una evaluación rápida de su relevancia.

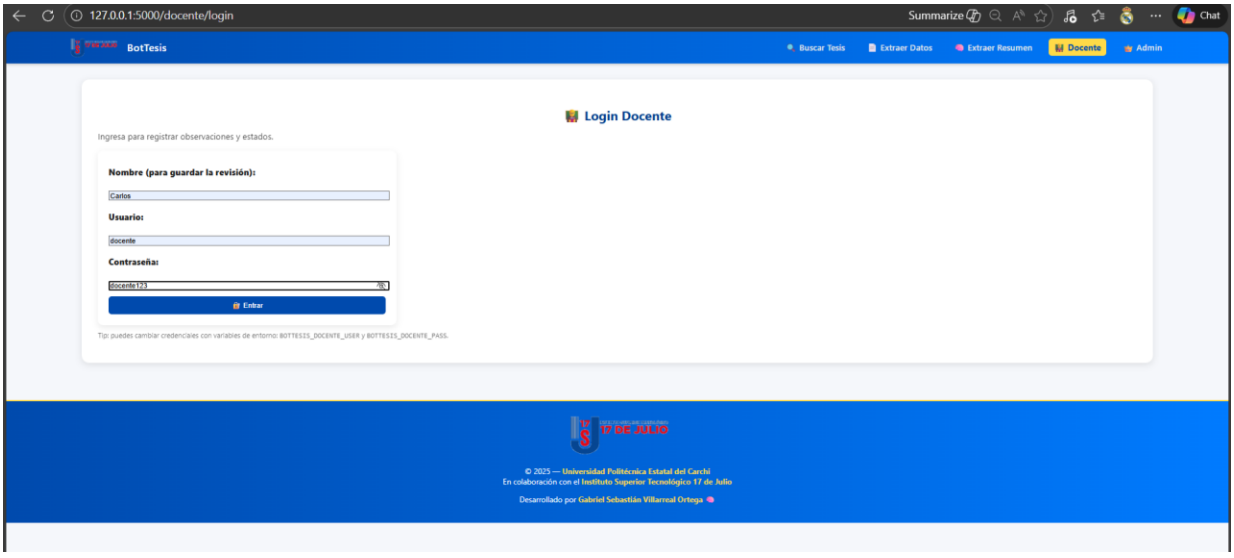


El módulo presenta metadatos estructurados del documento que se ha elegido, incluyendo el título, el código de identificación, la autoría y la ruta de acceso. Este elemento actúa como una capa intermedia entre la búsqueda general y el estudio del documento. La presentación organizada de los datos garantiza que el material académico sea seleccionado con precisión, mientras que la validación de la

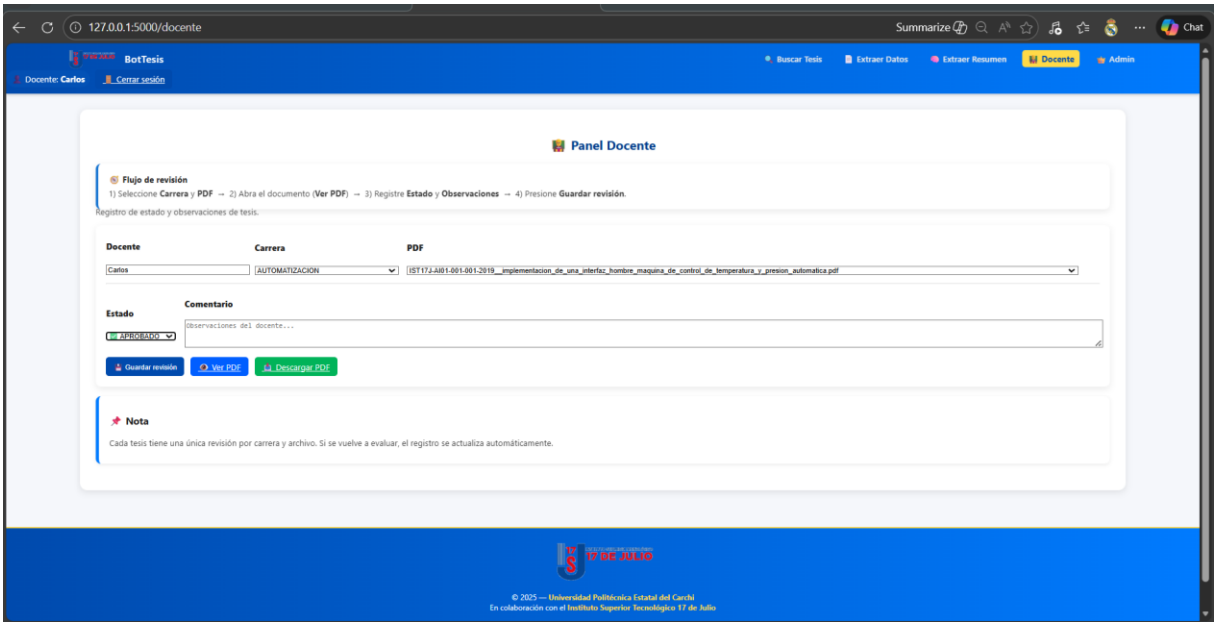
identidad del archivo se puede hacer sin abrirlo. Además, es posible observar alguna anotación del profesor correspondiente sobre algún error en su documento de titulación.



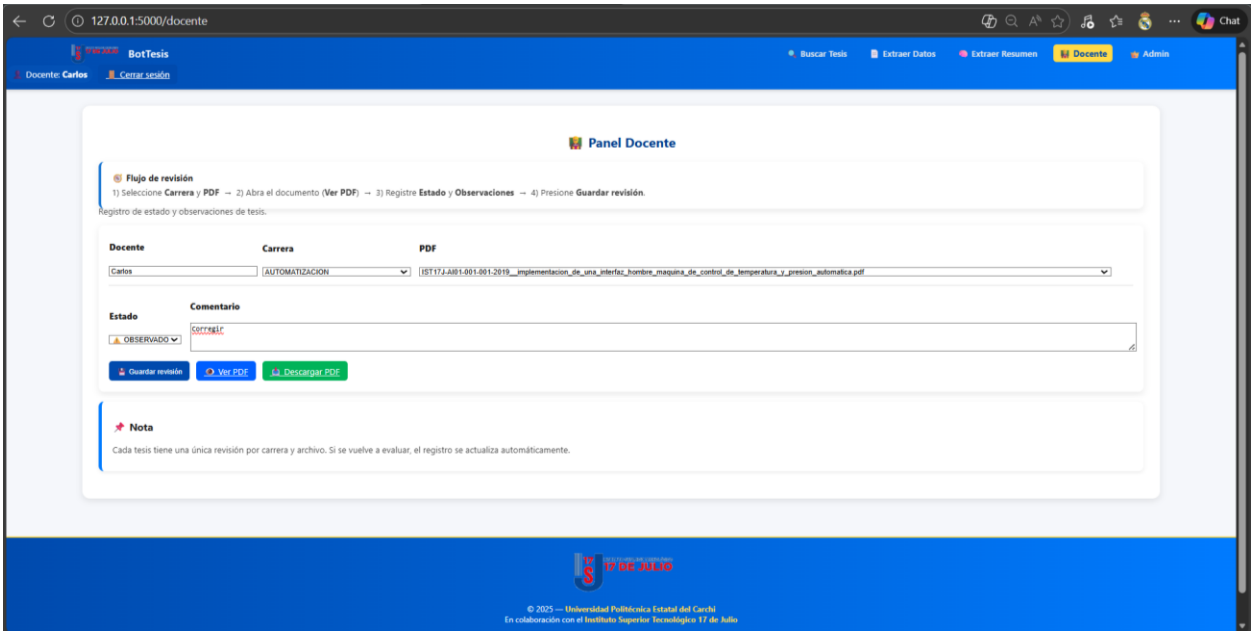
El módulo presenta el resumen extraído del documento lo que posibilita entender rápidamente el contenido sin necesidad de revisar el documento completo. Al igual que en el módulo de extracción de datos, es posible ver si el profesor correspondiente está observando o no.



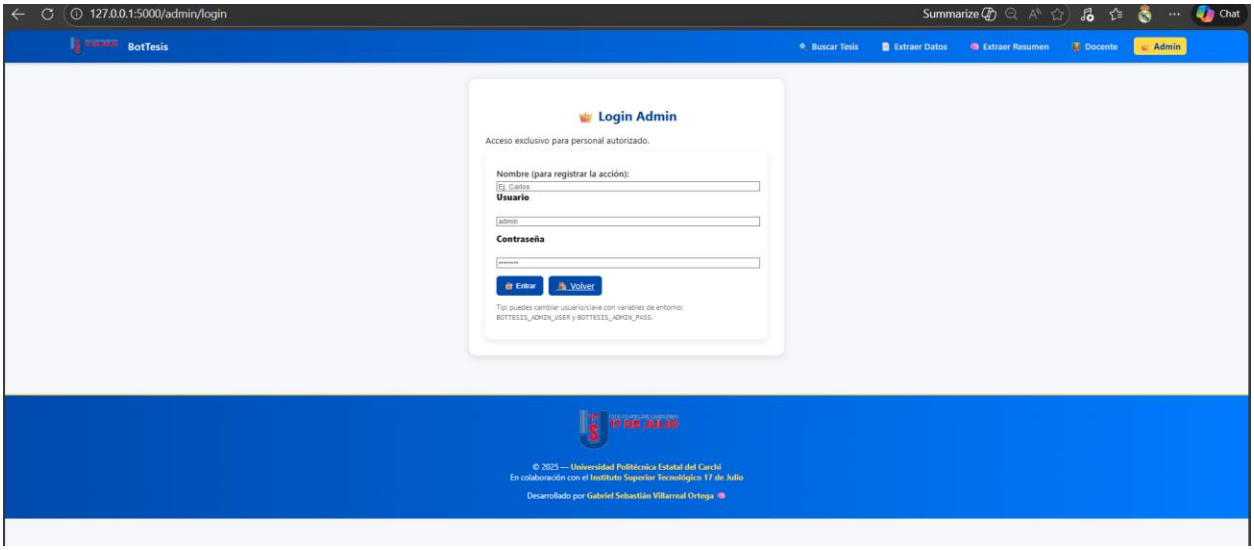
La pantalla de Inicio de Sesión Docente funciona como un puesto de control de seguridad que limita el acceso a las funciones del sistema para revisar y evaluar. El sistema garantiza que solo el personal autorizado pueda registrar estados y observaciones académicas a través de credenciales definidas por variables de entorno. Este procedimiento garantiza la seguridad de los documentos, conserva la confidencialidad de la información institucional y respeta las pautas de seguridad establecidas en el manual, lo cual refuerza la gobernanza digital del sistema.



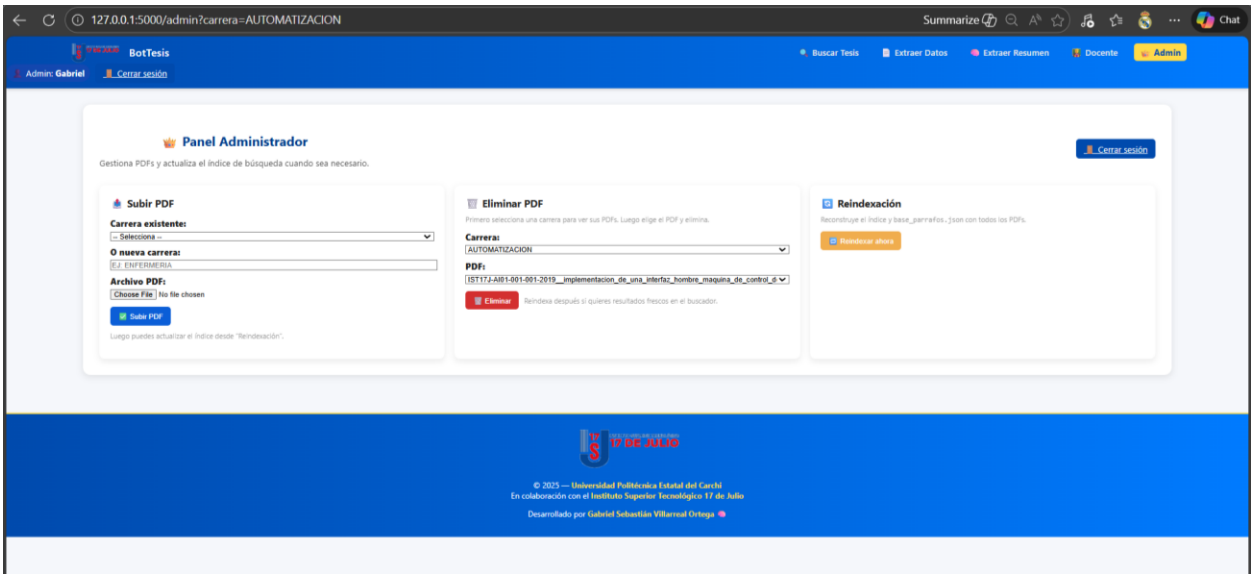
El Panel Docente está creado para respaldar el proceso formal de revisión de tesis, posibilitando que el profesor elija la carrera y el archivo PDF pertinente para su evaluación. El revisor tiene la posibilidad, a través de esta interfaz, de ver el documento, hacer anotaciones cualitativas y dar un estado académico como Aprobado o Observado. El sistema asegura que cada tesis conserve un único registro de revisión por archivo y carrera, lo cual garantiza la trazabilidad, el control académico y la actualización automática de los datos al llevar a cabo una nueva evaluación.



El sistema integra la revisión al historial académico de la institución cuando el profesor declara que una tesis está aprobada. Esta medida formaliza el documento y demuestra que el evaluador está de acuerdo, lo cual hace más transparente el procedimiento. Además, la interfaz posibilita volver a revisar el archivo o descargarlo antes de confirmar la evaluación, lo que combina un rigor académico con una eficiencia operativa. Este proceso disminuye los tiempos administrativos y homogeneiza el procedimiento de aprobación de los Trabajos de Integración Curricular.



El primer nivel de seguridad del sistema inteligente de búsqueda de tesis es la pantalla de Login Administrador, que limita el acceso solamente a los empleados autorizados. Esta interfaz requiere que el administrador, además de la identificación nominal, proporcione credenciales específicas para el usuario y la contraseña, lo cual posibilita registrar cada acción realizada en el sistema de manera que se pueda rastrear. El proceso de autenticación asegura que solamente aquellos usuarios con privilegios administrativos tengan acceso a tareas esenciales como la gestión de documentos, el mantenimiento del sistema y la reindexación del repositorio. Asimismo, la utilización de credenciales que se pueden configurar a través de variables de entorno fortalece la seguridad institucional, ya que se ajusta a las pautas de control de acceso y protección de datos establecidas en el manual del sistema.



Con funciones principales como la carga, eliminación y reindexación de documentos PDF por carrera, el Panel Administrador permite manejar de forma centralizada los trabajos de integración curricular. A partir de este módulo, el administrador tiene la posibilidad de añadir archivos nuevos al repositorio institucional o suprimir documentos que estén obsoletos, lo cual asegura la integridad y la actualización de la base de datos. Además, la opción de reindexación recompone el índice semántico del sistema y garantiza que los resultados de las búsquedas sean precisos y coherentes. Este panel fortalece la supervisión operativa del sistema y apoya la sustentabilidad del repositorio digital académico.

El sistema incluye la función de búsqueda por reconocimiento de voz, que posibilita que el usuario realice preguntas oralmente a través del micrófono del dispositivo. Esta opción simplifica la interacción con el buscador, sobre todo en situaciones en las que el ingreso manual de texto es limitado, ya que conserva el proceso de búsqueda y presentación de resultados tal como lo hace una consulta escrita.

19. Anexo A: Arquitectura del Sistema

• Capa de presentación: interfaz en línea creada con Flask.

• Capa lógica: servicios de reglas de negocio, procesamiento del lenguaje natural y búsqueda.

• Capa de datos: almacenamiento de PDFs y base de datos con índice.

Diagrama propuesto:

Usuario → Navegador web → Servidor de Flask → Motor de búsqueda / procesamiento del lenguaje natural (NLP) → Base de datos sobre tesis

20. Anexo B: Flujo General de Operación

1. Acceso al sistema.
2. Carga de consulta o selección de un archivo.
3. Procesamiento del servidor.

4. Revise bases de datos e índices.
5. Exposición de resultados.
6. Seguimiento del proceso de investigación.

21. Créditos, Licencia y Versión

Autoría: Gabriel Sebastián Villarreal.

Proyecto desarrollado en el contexto de la Unidad de Integración Curricular de la Universidad Politécnica Estatal del Carchi.

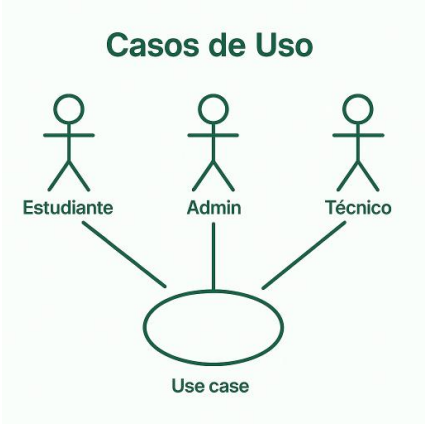
Edición del documento: 1.0

Utilización: Este manual es únicamente para fines académicos institucionales y no puede ser difundido fuera del ambiente autorizado sin la pertinente aprobación.

22. Casos de uso del sistema

Búsqueda exitosa

Admite que escribes 'sistema experto' en el motor de búsqueda. El sistema identifica coincidencias y te presenta partes en las que se encuentra ese término, además de resaltar la palabra clave. Tienes la posibilidad de examinar múltiples resultados y seleccionar el que esté más en línea con tu investigación.



Qué hacer si no aparece ningún resultado

No te alarmes si la búsqueda no arroja ningún resultado. Prueba con sinónimos de la palabra clave, emplea términos más generales o verifica la ortografía. En ocasiones, el sistema no encontrará coincidencias debido a que el concepto aparece en las tesis de otra manera.

Cómo escoger palabras clave

Reflexiona como un investigador: palabras breves, directas y exactas. Ejemplo: "comercio exterior", "gestión empresarial", "aprendizaje automático". No uses oraciones extensas. Cuando se le proporcionan términos específicos y claros, el motor de búsqueda obtiene mejores resultados.

23. Privacidad y Gestión de Datos

El sistema solo funciona con documentos oficiales de la institución. No guarda información fuera del entorno establecido ni la distribuye con personas ajenas.

24. Límites del Sistema

No sustituye la lectura integral de las tesis ni el análisis académico. Depende de la capacidad de lectura y de la calidad del PDF.

25. Alcances del Motor de Búsqueda

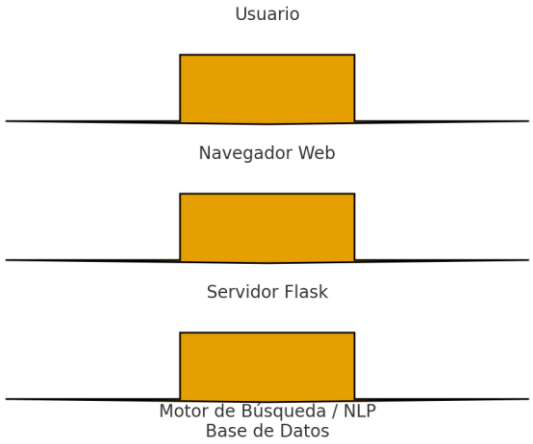
El sistema lleva a cabo búsquedas sobre la base de coincidencias de términos normalizados en los documentos que han sido indexados.

"No reemplaza el análisis académico ni la interpretación de significados subyacentes más allá de lo que se encuentra en los documentos.

26. Advertencias sobre Información Parcial

Los fragmentos presentados no son la totalidad del documento. La tesis original debe ser consultada para las referencias formales.

Diagrama de Arquitectura



27. Políticas de uso

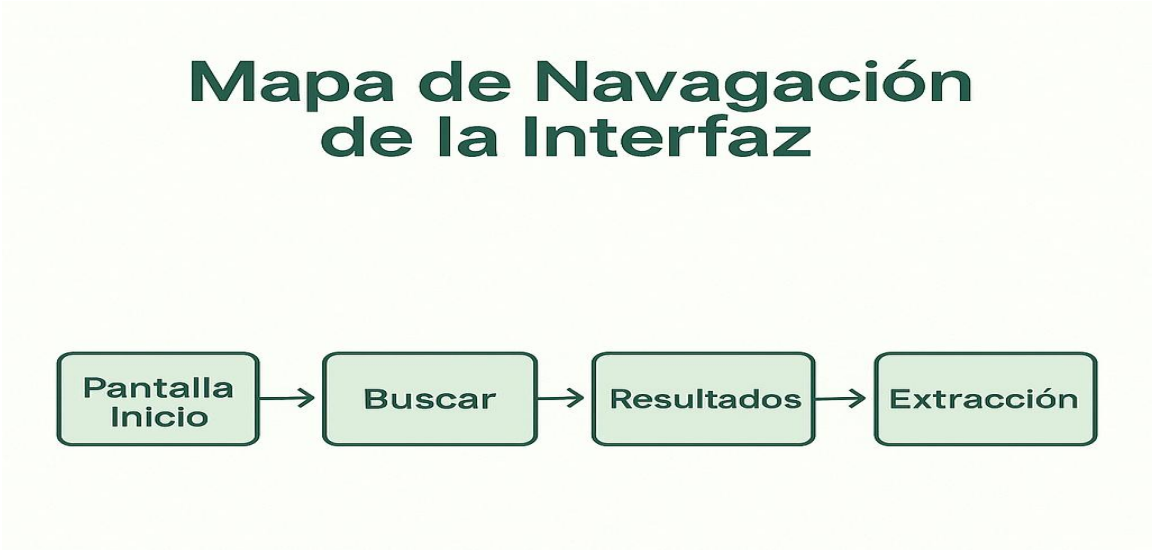
Dentro del marco institucional de la UPEC, las políticas de uso que se detallan a continuación regulan el trato entre los usuarios y el Sistema Inteligente de Búsqueda de Tesis (BotTesis).

Son aplicables a alumnos, administradores y personal técnico.

Tabla de roles y permisos (según políticas)

Rol	Permisos principales	Restricciones
Estudiante	Examinar tesis Usar todos los módulos	No cambiar sistema • No compartir información
Administrador	Inspeccionar uso Ejecutar pruebas Gestionar acceso	No descomponer datos internos
Personal Técnico	Sostenimiento Logs Infraestructura	Cambios solo con permisión oficial

El uso inapropiado del sistema podría resultar en limitaciones de acceso, supervisión por parte de las autoridades académicas y la implementación de las regulaciones institucionales actuales.



28. Lineamientos de seguridad

Los lineamientos que se presentan a continuación tienen como objetivo salvaguardar la confidencialidad, integridad y disponibilidad del Sistema Inteligente de Búsqueda de Tesis.

Seguridad para estudiantes:

- Utilizar redes que sean seguras.
- Cerrar la sesión en dispositivos de uso compartido.

Seguridad para administradores:

- Emplear credenciales sólidas.
- Revisar los registros de manera periódica.
- Utilizar el principio de privilegio mínimo al otorgar permisos y roles.

Seguridad para personal técnico:

- Mantener las bibliotecas al día.
- Proteger copias de seguridad encriptadas.
- Resguardo de servicios expuestos.

29. Guías rápidas de uso

Las siguientes guías rápidas resumen los pasos esenciales para manejar el sistema según el rol del usuario.

Tabla de guías rápidas según rol

Rol	Pasos rápidos
Estudiante	1) Buscar tesis 2) Revisar sugerencias 3) Observar fragmentos 4) Extraer datos/resumen
Administrador	1) Comprobar disponibilidad 2) Ejecutar prueba 3) Coordinar incidencias
Técnico	1) Revisar logs 2) Confirmar servicios 3) Aplicar actualizaciones controladas

30. Casos de uso ilustrados

Se exhiben flujos para estudiantes, administradores y técnicos. Cada uno cuenta el objetivo y el proceso básico asociado.

Caso de uso 1: Consulta de tesis por parte de un estudiante

• Protagonista: Alumno.

• Propósito: Encontrar una tesis que trate sobre un tema particular de investigación.

• Flujo básico: el alumno accede al sistema, introduce las palabras clave, examina los resultados y extrae datos o un resumen.

Caso de uso 2: Verificación de funcionamiento por parte de un administrador

• Protagonista: Gestor institucional.

• Propósito: Comprobar que el sistema está operativo y accesible para la comunidad académica.

• Flujo básico: El administrador se conecta al sistema, realiza búsquedas de prueba y, si encuentra irregularidades, se comunica con el departamento técnico.

Caso de uso 3: Mantenimiento técnico del sistema

• Protagonista: Personal especializado.

• Propósito: Conservar el ambiente de ejecución del sistema estable y al día.

• Flujo básico: el equipo técnico verifica los registros, implementa las actualizaciones y comprueba que todo funcione correctamente después.

Anexo C: Diagramas UML del Sistema

Este apéndice incluye diagramas UML a nivel conceptual que respaldan la comprensión de la estructura y el comportamiento del Sistema Inteligente de Búsqueda de Tesis.

Diagrama de casos de uso

Figura la interacción entre los principales actores (Estudiante, Administrador y Personal técnico) y las funcionalidades clave del sistema.

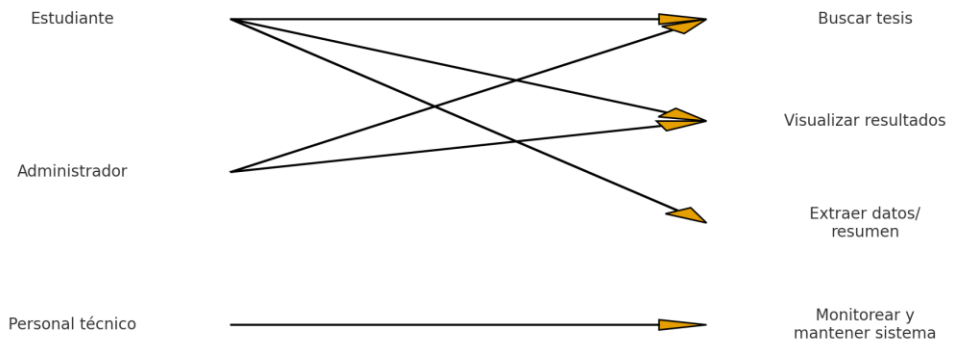


Diagrama de flujo operativo

Cuenta, de forma simplificada, el flujo general de operación desde que el usuario integra una consulta hasta que recibe los efectos.

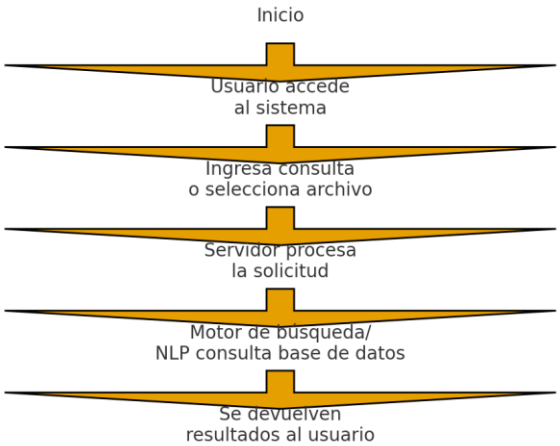


Diagrama de componentes

Ejemplar los componentes principales del sistema y sus relaciones lógicas.

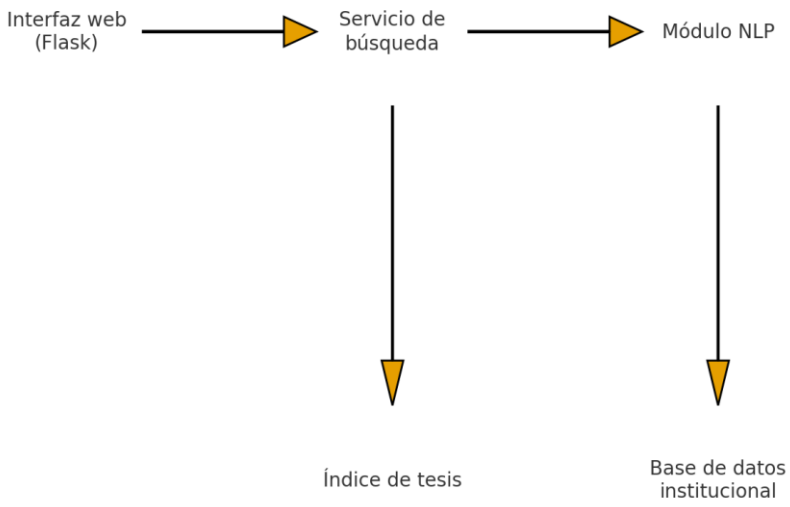


Diagrama de secuencia

Incorpora la secuencia de mensajes intercambiados entre el usuario, la interfaz, el servidor y la base de datos durante una exploración.

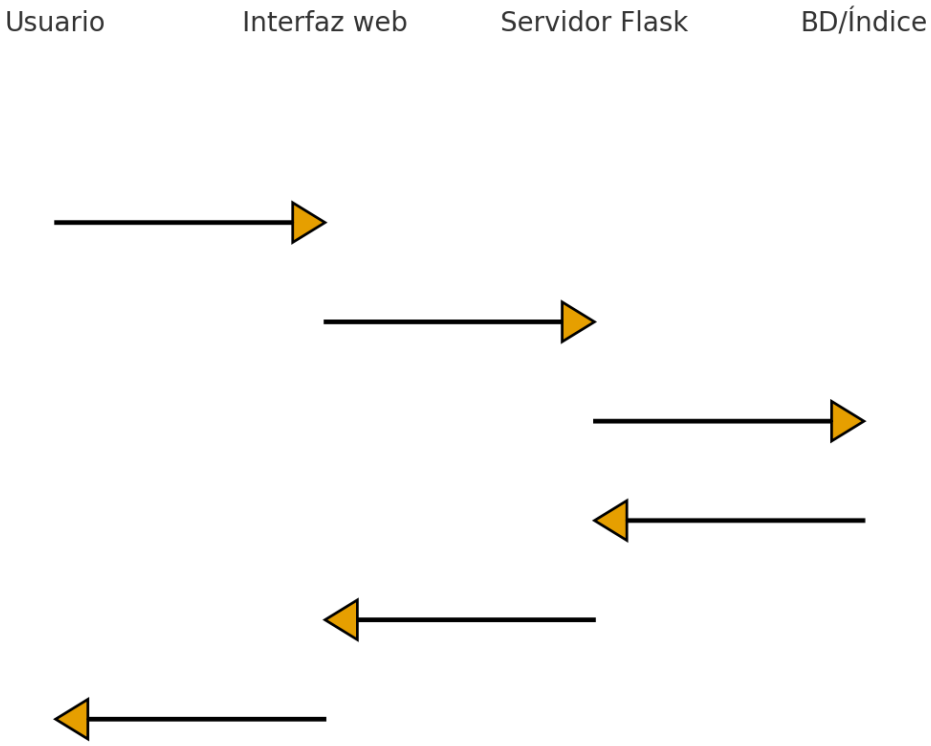
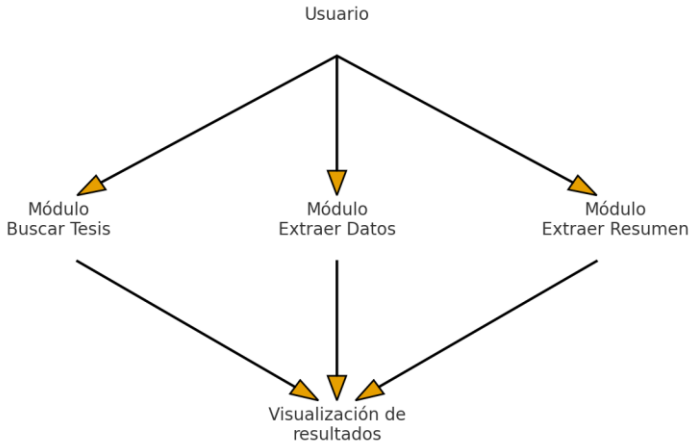


Diagrama de interacción del usuario

Destaca los puntos de interacción del usuario con los distintos módulos de la interfaz primordial del sistema.



31. Errores Comunes del Usuario

Esta sección señala los errores más comunes que los usuarios tienden a cometer cuando interactúan con el Sistema Inteligente de Búsqueda de Tesis. Comprender estos errores evita inconvenientes y acelera el flujo de trabajo.

32. Actualizaciones Futuras

- Modelos semánticos optimizados.
- Índices extendidos.
- Interfaz mejorada.
- OCR de mayor exactitud.
- Tablero de control administrativo.

33. Ejemplos Reales

A continuación, se presentan ejemplos prácticos que ilustran la manera en que el sistema se utiliza en situaciones de la vida diaria. A pesar de que los datos son ficticios, reflejan circunstancias reales en el ámbito académico.

Ejemplo 1: Búsqueda de Tesis

Palabra clave ingresada: "sistemas distribuidos"

Autocompletado sugiere: "sistemas distribuidos", "computación híbrida", "redes distribuidas"

Resultado mostrado:

- *"Optimización de sistemas distribuidos para entornos académicos" — Ingeniería en Sistemas — Tesis_2020_Distribuidos.pdf*
- Fragmento destacado: *"La construcción propuesta adelanto el rendimiento de los sistemas distribuidos mediante..."*

Interpretación: El usuario observa que el asunto está en línea con su investigación y comienza a abrir el metadato.

Ejemplo 2: Extracción de Datos

Carrera seleccionada: Ingeniería en Sistemas

Archivo seleccionado: *Tesis_2025_CyberIA.pdf*

Datos devueltos por el sistema:

- **Título:** *"Estudio de Inteligencia Artificial para Ciberseguridad Predictiva"*
- **Autores:** Ana Torres — Diego Fernández
- **Código identificador:** SIS-2025-IA-17
- **Ruta institucional:** /repositorio/sistemas/2025/IA/Tesis_2025_CyberIA.pdf

Uso típico: Inspeccionar referencias o aprobar que el archivo sea el correcto antes de descargarlo.

Ejemplo 3: Extracción de Resumen

Archivo seleccionado: *Tesis_2022_AgroTech.pdf*

Resumen extraído del documento:

"Este proyecto desarrolla un sistema inteligente..."

Interpretación: El estudiante cree el enfoque del trabajo sin leer las 120 páginas rematas.
