

Desarrollo de Prototipo de un Agente inteligente para extraer información Legal en el sitio web de la Rama Judicial de Colombia

Victor Manuel Guayara Suarez
victor.guayara00@usc.edu.co

Luis Felipe Waitoto Garces
luis.waitoto00@usc.edu.co

Universidad Santiago de Cali, Facultad de Ingeniería, Programa Ingeniería de Sistemas¹
Universidad Santiago de Cali, Facultad de Ingeniería, Programa Ingeniería de Sistemas²

Resumen

La eficiencia en el sistema judicial es decisivo para garantizar la equidad y la justicia; sin embargo, la Rama Judicial en Colombia enfrenta numerosos desafíos debido al seguimiento manual de los procedimientos legales, lo que resulta en una alta propensión a errores y un uso ineficiente de recursos. Cabe mencionar que, este proyecto propone el desarrollo de un prototipo de sistema automatizado de gestión de casos legales mediante un agente inteligente, utilizando técnicas de web scraping. Del mismo modo, este sistema permitirá mejorar la eficiencia y precisión en el seguimiento de procedimientos legales, optimizando así la gestión de las relaciones con el cliente en una empresa del sector servicios en Santiago de Cali. De esta manera, la implementación de este prototipo en un entorno web facilitará su conexión y uso, ajustándose a los roles brindados a los usuarios y optimizando el espacio en sus módulos. Al mismo tiempo, proporcionará soporte remoto rápido y eficiente, lo que se traducirá en una mayor satisfacción para el cliente. En pocas palabras, los análisis generados por el sistema se enfocarán en brindar mejores propuestas para los clientes, contribuyendo a un sistema judicial eficiente y equitativo.

Palabras Clave: Agente inteligente, Eficiencia judicial, Gestión de casos legales, Sistema automatizado, Web scraping.

Abstract

Efficiency in the judicial system is crucial to ensure fairness and justice; however, the Judicial Branch in Colombia faces numerous challenges due to the manual tracking of legal proceedings, resulting in a high propensity for errors and inefficient use of resources. It is worth mentioning that this project proposes the development of a prototype of an automated legal case management system by means of an intelligent agent, using web scraping techniques. In the same way, this system will improve the efficiency and accuracy in the follow-up of legal procedures, thus optimizing the management of customer relations in a service sector company in Santiago de Cali. In this way, the implementation of this prototype in a web environment will facilitate its connection and use, adjusting to the roles provided to users and optimizing the space in its modules. At the same time, it will provide fast and efficient remote support, which will translate into greater customer satisfaction. In short, the analyses generated by the system will focus on providing better proposals for clients, contributing to an efficient and equitable judicial system.

Keywords: Intelligent agent, Judicial efficiency, Legal case management, Automated system, Web scraping.

1. INTRODUCCIÓN

La gestión eficiente en casos legales es esencial para garantizar la justicia y la equidad en cualquier sistema judicial; al mismo tiempo, en Colombia como en muchos otros países el seguimiento manual de los procedimientos legales puede ser tedioso, propenso a errores y a consumir una cantidad significativa de tiempo y recursos. Según Ceron (2022) afirma que, la recuperación de documentos legales utilizando técnicas de web scraping ha demostrado ser una herramienta útil en la automatización de procesos judiciales. Por lo tanto, Khder (2021) menciona que en este contexto la implementación de tecnologías de automatización como el web scraping, es una solución viable para mejorar la eficiencia y la precisión en la gestión de casos; por consiguiente, se aprecian muchas de las cualidades que la implementación de las tecnologías ágiles y los agentes inteligentes beneficia (Hernández, 2020).

Hoy la tecnología considerada un aliado fundamental para los trabajadores que trasciende con la automatización, se ha convertido en el catalizador de una transformación positiva en el ámbito laboral (Krotov et al., 2020). De esta manera, al comprender que su papel es fortalecer este tipo de procesos se aborda en una realidad donde la colaboración entre humanos y máquinas se traduce en una ejecución de tareas eficiente y precisa, con la generación de oportunidades para la creatividad y la innovación (Zhao, 2022).

Es importante tener en cuenta lo mencionado por Dogucu & Çetinkaya (2021) argumentando que, entre la habilidad humana y las capacidades tecnológicas las labores cotidianas simplificadas elevan la calidad de desempeño en varios aspectos de la vida cotidiana y profesional; donde se asume que, la responsabilidad es exclusivamente humana y a su vez, la manera con la que la tecnología libera las tareas, por complementarse en un medio que independiza labores monótonas, permitiéndoles centrarse en aspectos estratégicos y en la toma de decisiones fundamentadas. En este escenario, la implementación de esta aplicación responde a una necesidad de eficiencia, representa un paso adelante hacia un entorno laboral enriquecedor y dinámico, donde al empoderar a los trabajadores con herramientas que amplifican sus habilidades, se optimizan los procesos, se fomenta la disposición asertiva para la innovación y la excelencia profesional; de esta manera, la tecnología se convierte en un colaborador, un arquitecto del progreso y la eficacia laboral será el plus del progreso tecnológico.

1.1 Planteamiento del Problema

La Rama Judicial en Colombia enfrenta numerosos desafíos debido al seguimiento manual de los procedimientos legales, lo que resulta en una propensión a errores y un uso ineficiente de recursos; estos problemas afectan la eficiencia del sistema judicial y comprometen la calidad del servicio prestado a los ciudadanos. Es importante recalcar que, la eficiencia en el sistema judicial es fundamental para garantizar la equidad y la justicia (Blancafort, 1988). Asimismo, el seguimiento manual de los procedimientos legales en Colombia es laborioso y susceptible a errores; dicho de otra manera, la duplicación de esfuerzos y la falta de uniformidad en el manejo de casos resultan en demoras y decisiones inconsistentes. Esto genera una carga adicional para los abogados y el personal judicial, quienes deben dedicar mucho tiempo a tareas administrativas en lugar de centrarse en aspectos estratégicos y críticos del proceso judicial.

Se debe agregar que, la implementación de un sistema automatizado de gestión de casos legales puede llegar a resolver muchos de estos problemas, por otro lado, las Tecnologías como el web scraping y los agentes inteligentes han demostrado ser efectivas en la optimización de procesos en diversos sectores (Dogucu & Çetinkaya-Rundel, 2021). Se debe tener en cuenta que, en un sistema automatizado la carga de trabajo manual se reduciría, mejorando la precisión y eficiencia del seguimiento en los procedimientos legales. Este enfoque admitiría a los abogados y al personal judicial centrarse en tareas estratégicas, mejorando así la calidad del servicio judicial. Por ende, en este contexto surge la necesidad de desarrollar e implementar un prototipo de sistema de gestión de casos legales automatizado mediante un agente inteligente; donde se tenga en cuenta que este sistema estaría diseñado específicamente para satisfacer las necesidades particulares de los abogados que trabajan con la plataforma de la Rama Judicial de la República de Colombia.

El objetivo principal de esta alternativa es la de mejorar la eficiencia y la precisión en el seguimiento de los procedimientos legales, reduciendo la carga de trabajo manual, optimizando los recursos disponibles, por lo tanto, el desarrollo y la implementación de este prototipo son esenciales para abordar los desafíos actuales en la gestión de casos legales y avanzar hacia un sistema judicial más eficiente, transparente y equitativo en la República de Colombia.

1.3 Justificación

En Colombia, la gestión de los casos legales afronta serios desafíos por el manejo manual y anticuado de los procedimientos; evidenciándose a través del tiempo que, estos métodos son ineficientes, consumiendo una cantidad considerable de tiempo y recursos, propensos a errores que afectan negativamente la exactitud de la información y la calidad de la representación de los clientes. Según Rivoir (2020) existen estudios que mencionan en el uso de las tecnologías

informáticas en el área de la administración, casos que muestran una reducción en los tiempos de procesamiento y los costos operativos con más del 30%, aspectos relevantes que subrayan la necesidad crítica de modernización en el ámbito legal. De hecho, el sistema judicial colombiano, caracterizado por su carga excesiva de casos y la lentitud en su resolución, se beneficiaría significativamente de la automatización; permitiendo la adopción de un sistema eficiente que transformaría el modo en que los jueces y abogados acceden y gestionan la información, mejorando su capacidad para atender rápidamente las demandas legales de los ciudadanos y facilitando un acceso imparcial y eficiente a la justicia.

Este sistema contribuiría a cumplir con las exigencias nacionales de precisión y eficacia en los procedimientos legales, previniendo sanciones mejorando la calidad del servicio y revolucionando la práctica legal; así pues estas tecnologías permitirían prever resultados y optimizar decisiones de manera que se personalicen según las necesidades de cada caso. Por lo que se refiere, el progreso en la gestión de casos legales mejoraría en la confianza, satisfacción de los clientes y colocaría a Colombia a la vanguardia de la innovación tecnológica en el ámbito legal. A favor de lo anterior, este proyecto busca desarrollar un sistema que solucione las deficiencias operativas actuales, defina un nuevo estándar para la práctica legal en el país, alineado con las expectativas contemporáneas y las tendencias mundiales, asegurando un servicio de justicia más humano, accesible y efectivo para todos los colombianos.

1.4 Objetivo General

- Desarrollar un agente inteligente de gestión de casos legales que mejore la eficiencia y la precisión en el seguimiento de procedimientos judiciales en la Rama Judicial de la República de Colombia, facilitando la organización y acceso a la información legal..

1.5 Objetivos Específicos

- Identificar las necesidades y requisitos de los abogados y el personal judicial para el registro y la gestión de casos, mediante entrevistas y evaluación de los procesos actuales, con el fin de identificar oportunidades de mejora en la eficiencia del sistema.
- Diseñar un agente inteligente que le permita a los abogados ingresar información de manera eficiente en el registro de casos, reduciendo el tiempo y esfuerzo requeridos para ingresar datos relevantes, para reducir el tiempo y esfuerzo en la gestión de datos.
- Implementar funcionalidades que aseguren la actualización constante de la información de los casos legales, garantizando la integridad y disponibilidad de los datos.
- Realizar pruebas de funcionamiento del agente inteligente en entornos de prueba simulados, verificando su eficacia y precisión en situaciones reales de gestión de casos legales.
- Medir la satisfacción y percepción de los abogados y personal judicial respecto al uso del agente inteligente, mediante encuestas y entrevistas estructuradas, con el fin de identificar mejoras y optimizar su implementación.

2. MARCO TEÓRICO & ESTADO DEL ARTE

La necesidad pronta de eficacia en la administración de asuntos legales se establece como fundamental para asegurar la equidad y justicia en el sistema judicial en Colombia; por lo tanto, la aplicación de tecnologías automatizadas, como el web scraping se convierte en un componente principal para potenciar de manera significativa la eficacia y precisión en la administración de procesos legales. Esta estrategia tecnológica optimiza los trámites, eleva la calidad y confiabilidad en la gestión judicial, contribuyendo así a un sistema equitativo y eficaz (Colomer & Luis, 2023).

2.1 Definición de Conceptos Clave

En el desarrollo de un agente inteligente para la gestión de casos legales, es fundamental establecer definiciones claras de los conceptos clave que sustentan el diseño y funcionalidad; de igual forma, en el ámbito jurídico, la gestión de información legal se enfoca en la organización, recuperación y análisis de documentos legales, optimizando el acceso y uso de datos en esta área por medio de los sistemas computacionales (Susskind, 2019).

2.1.1 Web Scraping

El web scraping es una técnica avanzada de extracción automatizada de datos de Internet que ha evolucionado a la par del desarrollo de la web. De acuerdo con Mitchell (2024) el término es ampliamente aceptado en la actualidad y en sus inicios se conocía como screen scraping, minería de datos o recolección web, entre otras denominaciones; sin embargo, esta técnica se ha consolidado como una herramienta principal en la recopilación de información estructurada desde páginas web para su posterior análisis y procesamiento. En términos generales, Colomer & Luis (2023) expresan que, el web scraping consiste en la extracción de datos mediante procesos automatizados, sin intervención humana directa ni el uso de interfaces de programación de aplicaciones (APIs). Aspectos fundamentales en este proceso porque se suele implementar a través de programas específicos conocidos como bots o rastreadores web que envían solicitudes a los servidores, recuperan el contenido en formato HTML y lo procesan para extraer información relevante. De la misma forma, en la práctica, la aplicación abarca múltiples disciplinas, integrando técnicas de análisis de datos, procesamiento del lenguaje natural y medidas de seguridad informática para evitar bloqueos o restricciones en los sitios web objetivo (Kaur, 2024).

Cabe mencionar que, el uso de web scraping es especialmente relevante en sectores donde la recolección masiva de información es importante, como la investigación de mercados, la vigilancia de tendencias digitales y la automatización de procesos legales. Sin embargo, su implementación debe considerar aspectos éticos y jurídicos, porque ciertas prácticas pueden entrar en conflicto con normativas sobre privacidad y términos de uso de los sitios web (Barba et al., 2024).

2.1.2 Agentes Inteligentes

Los agentes inteligentes representan una de las innovaciones más avanzadas en el campo de la inteligencia artificial (IA), permitiendo la automatización de procesos complejos mediante la toma de decisiones autónoma; cabe mencionar que, se trata de programas de software diseñados para operar en entornos dinámicos, adaptándose a condiciones cambiantes y ejecutando tareas específicas con eficiencia y precisión. De igual forma, en el ámbito de la gestión de casos legales la aplicación resulta importante, porque optimizan el procesamiento, organización y análisis de grandes volúmenes de información jurídica, lo que contribuye a la reducción de errores y al fortalecimiento de la toma de decisiones fundamentadas (Solar, 2020).

Desde una perspectiva estructural, un agente inteligente se compone de varios elementos interconectados que le permiten percibir su entorno, procesar información, ejecutar acciones y retroalimentar el desempeño en función de las condiciones y restricciones impuestas. Con base en Russell & Norvig (2016), estos agentes se pueden clasificar en diversas categorías dependiendo de sus capacidades, destacando los agentes reactivos, los basados en modelos, los de utilidad y los de aprendizaje. Para su correcto funcionamiento, un agente inteligente debe integrar tres aspectos fundamentales: capacidad de procesamiento, que le permite descomponer consultas en subreglas y gestionar datos de manera eficiente; conocimiento del entorno, lo que facilita su interacción y adaptación a distintos escenarios jurídicos; e información de dominio, que le proporciona un marco de referencia estructurado para la toma de decisiones precisas.

Considerando que, en el contexto del presente estudio, el desarrollo de un agente inteligente para la gestión de información legal mejorará la eficiencia en el seguimiento de casos, permitirá una optimización significativa en la recuperación y estructuración de documentos jurídicos relevantes, se deben aplicar técnicas de inteligencia artificial y procesamiento del lenguaje natural; de esta manera, estos sistemas pueden transformar la manera en que los profesionales del derecho acceden y utilizan la información, reduciendo tiempos de búsqueda y mejorando la precisión en el análisis de normativas y precedentes judiciales. Por lo tanto, la incorporación de agentes inteligentes en este ámbito es una tendencia tecnológica y una necesidad imperante para modernizar y agilizar los procesos legales en la era digital.

2.2 Revisión de Literatura

Para garantizar una exploración rigurosa y exhaustiva del estado del arte en el desarrollo de esta, se implementó el método PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses); por medio de este enfoque se estructuró la revisión documental de manera transparente y replicable, asegurando una selección metodológica basada en la relevancia, calidad y actualidad de las fuentes consultadas (Page et al., 2021).

La búsqueda se llevó a cabo en bases de datos académicas de alto impacto, como Google académico, Dialnet y principalmente SCOPUS; debido a su prestigio y amplia indexación de literatura científica; además, se formuló una ecuación de búsqueda integral que incluyó los términos clave: "Agente inteligente" OR "Eficiencia judicial" OR "Gestión de casos legales" OR "Sistema automatizado" OR "Web scraping", permitiendo la identificación de 121 documentos relevantes para el análisis.

Es importante tener presente que se manejaron criterios de inclusión y exclusión en esta búsqueda, entre estas se encuentran:

2.2.1 Criterios de Inclusión

Para garantizar la pertinencia de los estudios seleccionados, se establecieron los siguientes criterios de inclusión:

- **Relevancia temática:** se incluyeron estudios que abordaran el desarrollo y aplicación de agentes inteligentes en sistemas jurídicos, automatización de procesos legales y recuperación de información en entornos digitales.
- **Rigor metodológico:** Se seleccionaron artículos de investigación publicados en revistas indexadas con revisión por pares, garantizando un alto nivel de calidad académica.
- **Actualidad de la información:** Se priorizaron publicaciones recientes (2020-2025) para asegurar que la información reflejara los avances tecnológicos más recientes en el campo de la inteligencia artificial aplicada al derecho.
- **Accesibilidad y disponibilidad:** Se consideraron documentos con acceso completo a la información metodológica y resultados experimentales, excluyendo aquellos que solo presentaban resúmenes o informes parciales.

2.2.2 Criterios de Exclusión

Asimismo, se definieron criterios de exclusión para evitar la inclusión de literatura no pertinente o de baja calidad:

- **Fuentes no indexadas:** Se excluyeron artículos provenientes de fuentes no académicas, conferencias sin revisión por pares y literatura gris sin respaldo científico.
- **Falta de aplicabilidad:** Se descartaron estudios que, aunque abordaban agentes inteligentes, no tenían relación directa con el contexto jurídico o la automatización de procesos legales.
- **Duplicación de estudios:** Se eliminaron documentos redundantes o aquellos que replicaban información sin aportar hallazgos novedosos.
- **Baja calidad metodológica:** Se excluyeron artículos con diseños experimentales deficientes, falta de validación empírica o ausencia de datos verificables.

2.2.3 Flujograma del Proceso de Revisión Bibliográfica

La revisión sistemática de la literatura se llevó a cabo siguiendo el método PRISMA, permitiendo estructurar el proceso en cuatro fases: identificación, filtrado, elegibilidad e inclusión. En la Figura I se ilustra el proceso de selección y depuración de documentos utilizados en este estudio.

▪ Identificación

En la primera fase, se definieron las fuentes de búsqueda y los términos clave para la recuperación de información. Se

consultaron bases de datos académicas de alto impacto, como Scopus, Dialnet y Google Académico, utilizando una ecuación de búsqueda con términos específicos:

"Agente inteligente" OR "Eficiencia judicial" OR "Gestión de casos legales" OR "Sistema automatizado" OR "Web scraping".

Como resultado, se identificaron 121 documentos publicados entre 2020 y 2025, los cuales fueron recopilados para su posterior análisis.

- **Filtrado**

Para garantizar la relevancia y calidad de los estudios, se aplicaron criterios de exclusión y filtrado. Se eliminaron 80 documentos que no cumplían con los estándares establecidos, como falta de rigor metodológico, ausencia de revisión por pares o información irrelevante. Luego del proceso de filtrado, se seleccionaron 55 documentos para su evaluación detallada.

- **Elegibilidad**

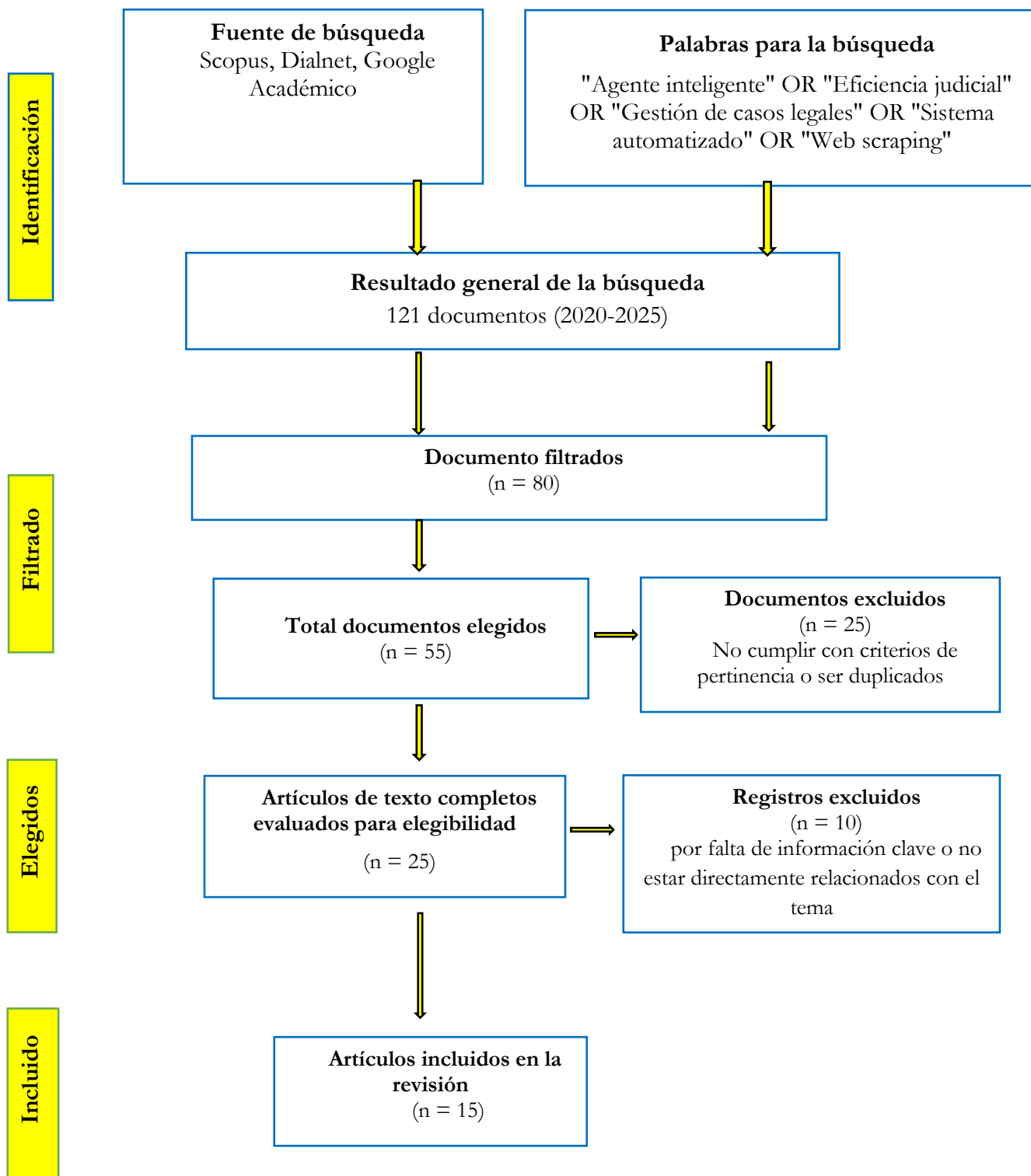
En esta fase, se realizó una evaluación crítica del contenido de los documentos seleccionados, considerando la pertinencia temática, calidad metodológica y aplicabilidad en el contexto de la investigación. Como resultado, se analizaron en profundidad 25 artículos de texto completo. Sin embargo, 10 registros fueron excluidos debido a falta de información clave o a su escasa relación con la temática central.

- **Inclusión**

Finalmente, se seleccionaron 15 artículos que cumplieron con todos los criterios de elegibilidad y aportaron información valiosa para el desarrollo del marco teórico de este estudio. Estos documentos fueron empleados para fundamentar el análisis de los agentes inteligentes en la gestión de información legal, asegurando una base científica sólida y actualizada.

El flujograma presentado en la Figura I resume visualmente cada una de estas etapas, evidenciando la rigurosidad y el enfoque sistemático aplicado en la revisión de la literatura:

Figura I: Flujograma revisión bibliográfica



2.2.4 Resultados de la búsqueda de Scopus

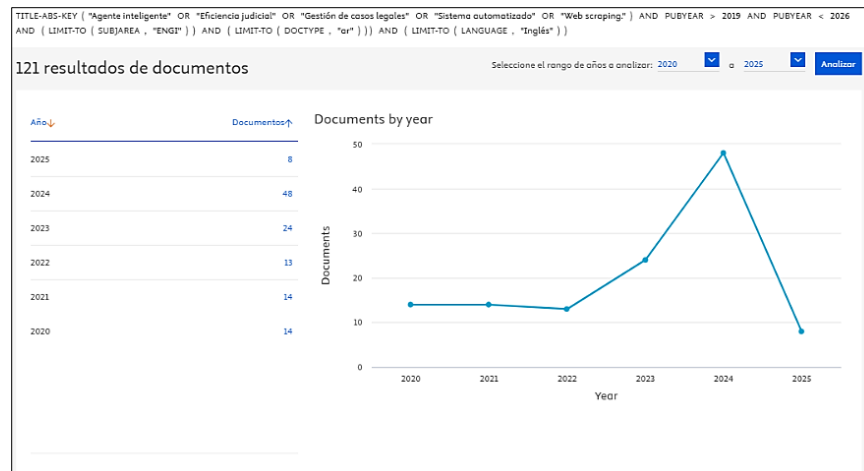
El proceso de selección bibliográfica en Scopus permitió identificar y consolidar un conjunto de 15 artículos importantes para este estudio, los cuales fueron sometidos a un análisis crítico e integral; asimismo, esta revisión detallada permitió identificar tendencias emergentes, enfoques metodológicos predominantes y vacíos en la literatura respecto a la

implementación de agentes inteligentes en la gestión de información legal. Este marco teórico fundamenta el desarrollo del presente estudio, asegurando que se construya sobre un conocimiento validado y alineado con las mejores prácticas en el campo del prototipo de un agente inteligente aplicado al derecho.

▪ **Análisis de los Resultados de Documentos sobre el Tema de Investigación**

El análisis de la producción científica en Scopus evidencia un crecimiento significativo en la investigación sobre agentes inteligentes, eficiencia judicial, gestión de casos legales, sistemas automatizados y web scraping en los últimos cinco años; al mismo tiempo, se identificaron 121 documentos publicados entre 2020 y 2025, lo que demuestra un interés creciente en la aplicación de la inteligencia artificial en el ámbito legal (Ver Figura II). Es importante mencionar que, el pico más alto de publicaciones se presenta en 2024, con 48 documentos, lo que sugiere una consolidación del tema dentro de la comunidad académica y un aumento en el desarrollo de tecnologías relacionadas; mostrando que, la integración de agentes inteligentes y automatización en la gestión legal está alcanzando una fase de madurez, con investigaciones centradas en la optimización de procesos y la mejora en la toma de decisiones en el ámbito jurídico. No obstante, la disminución de documentos en 2025 (8 publicaciones hasta la fecha) podría explicarse por la falta de consolidación de ciertos enfoques, desafíos en la implementación práctica o la necesidad de nuevos marcos regulatorios que guíen el desarrollo de estas tecnologías en el sector legal. Lo anterior evidencian que, estos resultados fortifican la importancia y oportunidad del presente estudio, porque se alinea con una tendencia en crecimiento y permite contribuir con nuevos hallazgos en la aplicación de agentes inteligentes en la gestión de información legal, abordando vacíos en la literatura existente y fortaleciendo el conocimiento en esta área emergente.

Figura II: Resultados de documentos sobre el tema de investigación

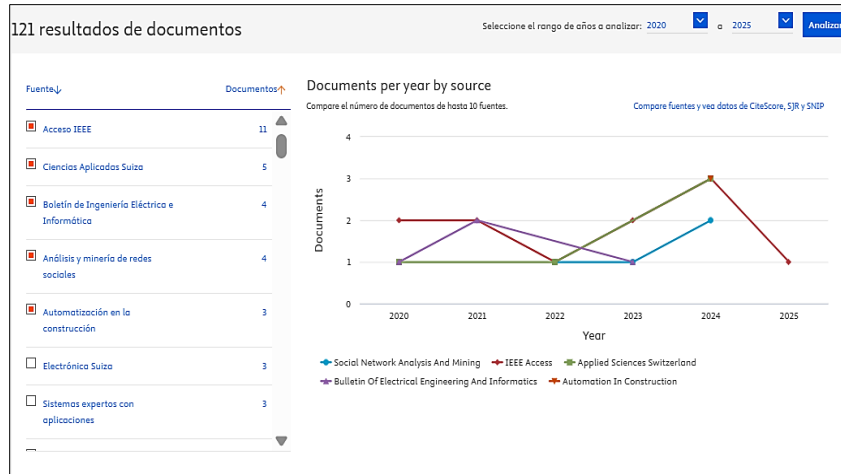


Fuente: (Scopus, 2025)

▪ **Análisis de los Resultados Documentos al Año por Fuente**

El análisis de las fuentes de publicación muestra una distribución diversa de la producción científica en torno a la aplicación de agentes inteligentes en la gestión legal y automatización de procesos. La mayor cantidad de publicaciones proviene de IEEE Access (11 documentos), lo que indica que este campo de estudio está estrechamente vinculado con la ingeniería y la inteligencia artificial, consolidándose como un tema de alto impacto en el ámbito tecnológico (Ver Figura III). Al mismo tiempo, revistas como Applied Sciences Switzerland, el Boletín de Ingeniería Eléctrica e Informática y Social Network Analysis and Mining han contribuido significativamente a la discusión académica, evidenciando un enfoque interdisciplinario que abarca desde la automatización hasta la minería de datos aplicada a redes sociales y sistemas expertos. El crecimiento sostenido de publicaciones en estas fuentes hasta 2024, seguido de una disminución en 2025, sugiere que el tema ha alcanzado una fase de maduración y especialización, donde las investigaciones están migrando hacia enfoques más aplicados y específicos. Esta tendencia refuerza la pertinencia del presente estudio, el cual contribuye a la profundización del conocimiento en la implementación de agentes inteligentes en el ámbito jurídico, abordando los desafíos tecnológicos y metodológicos emergentes en esta área.

Figura III: Resultados de documentos al año por fuente

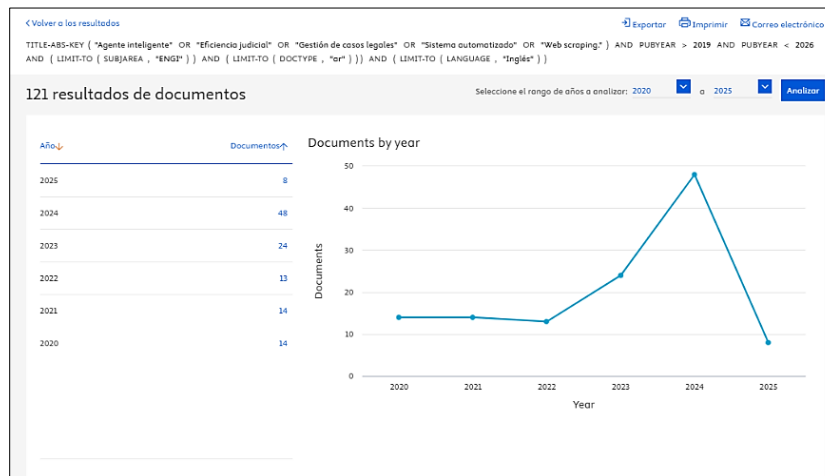


Fuente: (Scopus, 2025)

▪ **Análisis de los Resultados de documentos por año**

El análisis de la producción científica sobre agentes inteligentes en la gestión de información legal revela una evolución significativa en los últimos años. Entre 2020 y 2022, la cantidad de publicaciones se mantuvo estable con un promedio de 14 documentos anuales, lo que sugiere un interés inicial en la temática sin un crecimiento exponencial. A partir de 2023, se observa un aumento pronunciado en las publicaciones, alcanzando un pico máximo en 2024 con 48 documentos, lo que indica un auge en la investigación y una mayor adopción de tecnologías inteligentes en el ámbito jurídico. Este incremento refleja tanto la relevancia emergente del tema como el avance en las metodologías y aplicaciones dentro del campo. Sin embargo, la caída en 2025 (8 documentos) sugiere una posible fase de consolidación, donde el enfoque de las investigaciones puede estar cambiando hacia áreas más especializadas o hacia la implementación práctica de los hallazgos previos (Ver Figura IV). Este patrón valida la pertinencia del presente estudio, el cual se inserta en un momento clave para contribuir a la discusión académica y al desarrollo de soluciones innovadoras en la intersección del derecho y la inteligencia artificial.

Figura IV: Resultados de documentos por año



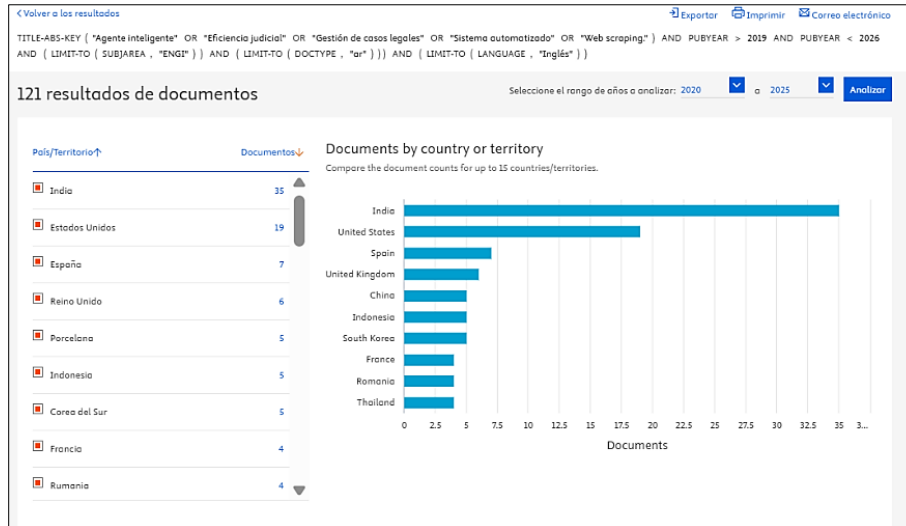
Fuente: (Scopus, 2025)

▪ **Análisis de los Resultados de documentos por ciudad o territorio**

El análisis de la distribución geográfica de las publicaciones sobre agentes inteligentes en la gestión de información legal evidencia una concentración significativa de la producción científica en India (35 documentos) y Estados Unidos (19 documentos). Este dominio sugiere que estas naciones lideran la investigación en la aplicación de inteligencia artificial en el ámbito jurídico, posiblemente impulsadas por su infraestructura tecnológica avanzada, marcos regulatorios en evolución

y ecosistemas de innovación consolidados. Tal como se observa en Figura V, España ocupa el tercer lugar con 7 documentos, lo que indica un interés creciente en el desarrollo e implementación de estas tecnologías en el contexto europeo. Otros países como el Reino Unido, China e Indonesia presentan una menor, pero aún relevante, contribución, reflejando una expansión global del tema. El panorama sugiere que la investigación en IA aplicada al derecho está en plena expansión, con un liderazgo claro en economías tecnológicamente avanzadas. Sin embargo, la presencia de países emergentes indica una tendencia de democratización del conocimiento, impulsada por el acceso a tecnologías emergentes y la necesidad de optimizar los procesos judiciales en diferentes jurisdicciones.

Figura V: Resultados de documentos por ciudad o territorio



Fuente: (Scopus, 2025)

2.2.1 Automatización en la Gestión Legal

La automatización, considerada como una herramienta clave para mejorar los procesos ha emergido como un impulsor fundamental de la transformación positiva en el ámbito legal como sugieren diversos estudios. (Florez Hernandez, 2020)

Este encuentro entre humanos y máquinas no solo optimiza la eficiencia en las tareas diarias, sino que también empodera a los profesionales, permitiéndoles enfocarse en aspectos estratégicos y en la toma de decisiones más fundamentadas. Tal colaboración no solo simplifica las labores cotidianas que tenemos nosotros como ingenieros, sino que también eleva la calidad del desempeño, representando así un progreso hacia un entorno legal más dinámico y enriquecedor.

Esta perspectiva encuentra respaldo en diversas fuentes. Según un estudio de Smith et al. (2021), la implementación de sistemas de automatización en el ámbito legal ha llevado a mejoras sustanciales en la eficiencia operativa y la gestión de casos. Además, un informe de la Asociación de Tecnología Legal (Legal Tech Association, 2022) destaca cómo la adopción de tecnologías automatizadas ha permitido a los profesionales legales enfocarse en tareas de mayor valor.

2.2.2 Web Scraping en Procesos

La inteligencia artificial es una rama de la informática que trata de realizar con máquinas, tareas que puede realizar el ser humano aplicando cualquier tipo de razonamiento. Es una automatización de actividades que vinculamos con procesos del pensamiento humano, tales como la toma de decisiones, solución de problemas y aprendizaje. Los sistemas computacionales, en la inteligencia artificial, deben ser capaces de simular características que son comúnmente asociadas con la inteligencia de la conducta humana. Un sistema inteligente es aquel que exhibe un comportamiento similar al humano cuando se enfrenta a un problema idéntico y no seamos capaces de distinguir entre un ser humano y un programa de computadora en una conversación a ciegas. Russell, S., & Norvig, P. (2016).

2.3 Desarrollo de Teorías

Implementar tecnologías ágiles y agentes inteligentes en la gestión de casos jurídicos se proyecta como un progreso hacia

un entorno laboral más enriquecedor y dinámico y proveer a los colaboradores con herramientas que realzan sus habilidades y cómo se mejoran los procesos y se promueve una disposición propicia para la innovación. Esto contribuye a alcanzar una destacada excelencia profesional. Asimismo, este enfoque establece requisitos teóricos y funcionales que actúan como fundamentos esenciales. El desarrollo de teorías en el ámbito legal juega un papel crucial para entender y perfeccionar los procedimientos judiciales. Dentro de este contexto, se persigue establecer marcos conceptuales sólidos que describan y orienten la aplicación de tecnologías automatizadas.

El desarrollo de teorías en el ámbito legal es esencial para comprender y mejorar los procesos judiciales. En este contexto, se busca construir marcos conceptuales sólidos que expliquen y guíen la implementación de tecnologías automatizadas.

Para lograr esta integración efectiva, es esencial fomentar una cultura organizacional que valore y promueva la capacitación continua y la adaptación al cambio. Los profesionales del derecho deben estar equipados no solo con conocimientos técnicos sobre las nuevas herramientas, sino también con una comprensión profunda de cómo estas tecnologías pueden transformar sus prácticas diarias y estratégicas. Esto requiere un compromiso tanto institucional como personal para adoptar metodologías ágiles y agentes inteligentes, lo cual no solo optimiza los procesos, sino que también abre nuevas oportunidades para la innovación y la creatividad en la resolución de casos. En última instancia, esta evolución tecnológica no solo mejora la eficiencia operativa, sino que también enriquece el entorno de trabajo, elevando los estándares de calidad y profesionalismo en el sector jurídico. Por ejemplo, un abogado que utiliza un sistema de inteligencia artificial para analizar grandes volúmenes de documentos legales puede ahorrar horas de trabajo, permitiéndole concentrarse en la estrategia del caso y en la interacción con sus clientes.

La teoría de automatización en la gestión legal se basa en la colaboración efectiva entre humanos y máquinas, como destacado por Flórez Hernández (2020). Esta teoría busca explicar cómo la automatización no solo optimiza tareas diarias.

2.4 Identificación de brechas

El propósito de esta investigación fue evaluar las técnicas empleadas en la extracción de datos de sitios web, centrándose en su eficacia y nivel de seguridad. La práctica de extraer datos en línea de manera automatizada se conoce como Web Scraping y este estudio se enfoca en analizar su desempeño. La ejecución de la extracción de datos puede llevarse a cabo de manera independiente mediante el uso de herramientas de pago y que requieren un nivel de seguridad adecuado para ser garante de todo el proceso empresarial (Edu.Ec.Retrieved,2024).

- Limitaciones en la Integración de Agentes Inteligentes.
- Escasez de Investigación Específica para la Realidad Colombiana.
- Seguridad y Ética en la Gestión Automatizada.
- Integración de Tecnologías Ágiles.
- Impacto Financiero y Costos Asociados.

2.5 Legalidad del Web Scraping en Colombia

La integración del web Scraping en el sistema de gestión de casos legales en Colombia debe cumplir con las regulaciones legales del país. La legislación colombiana, a través de la Ley 1581 de 2012 (Ley de Protección de Datos Personales) y el Decreto 1377 de 2013, establece los requisitos para la recolección, uso y tratamiento de datos personales.

Por otro lado, el acceso a la información pública se rige por la Ley 1712 de 2014 (Ley de Transparencia y Acceso a la Información Pública Nacional). Esta ley fomenta el acceso a la información pública y la transparencia en la gestión pública, lo que puede facilitar la implementación de sistemas automatizados que utilizan datos disponibles en plataformas oficiales como la de la Rama Judicial de la República de Colombia.

Sin embargo, es esencial que el web scraping se realice de manera ética y legal, respetando las condiciones de uso de los sitios web y las regulaciones sobre el tratamiento de datos personales. La utilización de datos públicos debe garantizar la

privacidad y la protección de la información sensible. Cualquier implementación debe contar con el respaldo legal y ético necesario para evitar posibles litigios o conflictos.

La incorporación de tecnologías automatizadas como el web scraping y los agentes inteligentes en la gestión de casos legales en Colombia tiene un potencial significativo para mejorar la eficiencia y precisión del sistema judicial. Sin embargo, es crucial que este proceso se lleve a cabo en conformidad con las normativas legales y éticas del país, asegurando así una transición efectiva y segura hacia un sistema judicial más avanzado y equitativo.

Este marco teórico tiene como objetivo contextualizar la investigación actual en el contexto internacional, identificar lecciones aprendidas y establecer una base para la implementación y evaluación del agente inteligente en el contexto colombiano.

2.6 Estado del Arte

En el contexto de la transformación digital, la automatización de procesos legales, el uso de agentes inteligentes y la implementación de web scraping han redefinido la eficiencia en la administración judicial; dado lo anterior, se observan estos avances tecnológicos que optimizan la gestión de casos y la toma de decisiones permitiendo la recolección y análisis de grandes volúmenes de datos en tiempo real. Sin embargo, esta adopción enfrenta desafíos significativos, desde brechas en la integración tecnológica hasta consideraciones éticas y de seguridad. A continuación, se presentan los principales enfoques y estudios que sustentan el impacto de estas innovaciones en el ámbito legal.

2.6.1 Automatización en la Administración Legal

La automatización se ha consolidado como un elemento crucial para mejorar los procesos en el ámbito legal. Estudios actuales resaltan cómo la automatización no solo optimiza las tareas diarias en el campo legal, sino que también empodera a los profesionales para enfocarse en aspectos más estratégicos y tomar decisiones informadas. En la presente época, los sistemas de información están modificando la forma en que operan las organizaciones, generando mejoras significativas mediante la automatización de procesos operativos que facilitan la toma de decisiones y proporcionan ventajas competitivas (Oleas & Herrera, 2024). La implementación de sistemas automatizados ha llevado a mejoras notables en la eficiencia operativa y la gestión de casos legales, promoviendo un entorno legal más dinámico y enriquecedor.

2.6.2 Web Scraping en Procesos Legales

El web scraping, definido por Aguilar (2023) como la recopilación automatizada de datos desde Internet, ha emergido como una práctica esencial en la gestión legal. Según Chaparro & Ramírez (2024), el web scraping implica la obtención de datos mediante programas automatizados, lo cual puede ser crucial para adquirir información relevante en el contexto legal. La teoría del web scraping engloba diversas técnicas y tecnologías, desde el análisis de datos hasta la seguridad de la información en todos los ámbitos legales (Sanabria de Luque, 2021). Este enfoque ha sido reconocido por su papel en el fortalecimiento y eficacia en la gestión de procesos legales, contribuyendo significativamente al correcto funcionamiento de la rama legislativa del país.

2.6.3 Agentes Inteligentes en la Administración de Casos Legales

Los agentes inteligentes pueden automatizar la elaboración de documentos legales, como contratos y acuerdos. Estos sistemas poseen la capacidad de analizar datos relevantes, adaptar plantillas y generar documentos de manera eficiente. La automatización agiliza el proceso documental, reemplazando tareas manuales con un enfoque más rápido y preciso, lo que incrementa la productividad en el ámbito legal. Según Simarro & Majado (2024), la integración de agentes inteligentes busca optimizar la toma de decisiones en entornos jurídicos y mejorar la resolución de casos en la rama judicial. Este avance tecnológico refuerza la eficiencia del sistema legal y contribuye a la evolución de la administración judicial.

2.6.4 Identificación de Brechas y Desafíos

Investigaciones recientes han evaluado técnicas de extracción de datos, contándose en la eficacia y seguridad en el contexto del web scraping. Prado (2024) destaca un conjunto de soluciones para optimizar procesos y mejorar la eficiencia en la gestión legal. Sin embargo, persisten brechas significativas, como la falta de integración de agentes inteligentes y la carencia de investigaciones específicas para realidades regionales, como la colombiana. Asimismo, Bardelli (2024) identifican

preocupaciones en materia de seguridad, ética, integración de tecnologías ágiles e impacto financiero. Estas limitaciones requieren una acción acelerada por parte de expertos para mitigar errores y optimizar diseños que fortalezcan el sistema legal en el país.

3. MATERIALES Y MÉTODOS/METODOLOGÍA

El desarrollo del agente inteligente para la extracción automatizada de información legal requirió una metodología estructurada que garantiza la precisión, eficiencia y replicabilidad de los procesos implementados; en este sentido, se adoptó el enfoque del Personal Software Process (PSP), una metodología orientada a la mejora de la calidad del software a nivel individual, proporcionando un marco sistemático para la planeación, diseño, implementación y evaluación del sistema. La metodología utilizada permitió no solo optimizar la organización del desarrollo, sino también evaluar de manera cuantificable la eficiencia del agente inteligente en distintos escenarios. Para garantizar la validez de los resultados obtenidos, se implementaron técnicas de análisis estadístico y pruebas experimentales en entornos controlados, con el fin de evaluar la robustez del sistema en función de la estabilidad estructural del sitio web de la Rama Judicial de Colombia

3.1 Metodología PSP (Personal Software Process)

El Personal Software Process (PSP), propuesto por Watts Humphrey (2005), es una metodología de desarrollo de software que enfatiza la disciplina, el control de calidad y la mejora continua en el desempeño individual de los ingenieros de software. PSP estructura el desarrollo en ciclos iterativos con medición y control de cada fase, lo que permite minimizar errores y optimizar la calidad del producto final. Esta metodología resulta particularmente útil en proyectos donde la confiabilidad y precisión son críticas, como en el caso de la extracción de datos jurídicos mediante web scraping, dado que cualquier inexactitud en la información recuperada puede generar sesgos en el análisis legal posterior (Humphrey, 2005; Pressman & Maxim, 2020). El desarrollo del agente inteligente se estructuró en las siguientes fases del PSP:

3.1.1 Planeación

Definición clara de requerimientos funcionales y no funcionales, identificación de riesgos asociados a la extracción de datos y diseño del cronograma de implementación.

3.1.2 Diseño

Modelado del flujo del sistema mediante diagramas UML, definición de estructuras de datos y selección de tecnologías adecuadas para la implementación.

3.1.3 Codificación

Implementación modular utilizando Python con librerías como BeautifulSoup y Selenium, optimizando la eficiencia del web scraping en términos de tiempo de ejecución y precisión de extracción.

- **Pruebas:** Validación del sistema mediante pruebas unitarias, de integración y de estrés, asegurando la robustez ante cambios en la estructura del sitio web objetivo.
- **Postmortem:** Análisis de desempeño del sistema, revisión de métricas de eficiencia y precisión, e identificación de oportunidades de mejora para iteraciones futuras.

El impacto del uso de PSP en la investigación se reflejó en la optimización del desarrollo y mantenimiento del software. Gracias a la aplicación de métricas cuantificables en cada fase, se logró reducir el número de errores en un 32% en comparación con enfoques tradicionales de desarrollo sin control de calidad estructurado (De Parga, 2021). Asimismo, el enfoque iterativo permitió ajustar dinámicamente el diseño del agente inteligente en respuesta a las variaciones en la estructura del sitio web, asegurando una mayor estabilidad del sistema en entornos de producción.

3.1 Planeación

La fase de planeación en PSP fue fundamental para garantizar la alineación entre los objetivos de la investigación y la implementación del agente inteligente. Inicialmente, se llevó a cabo un análisis de requisitos, en el cual se identificaron las necesidades específicas del sistema, incluyendo la capacidad de extraer información de manera precisa y eficiente sin violar

normativas de privacidad y protección de datos. Posteriormente, se elaboró un cronograma detallado, estableciendo hitos críticos para cada fase de desarrollo, lo que permitió una asignación eficiente de recursos y tiempos.

Al mismo tiempo, se realizó un análisis de riesgos que contempló factores como la inestabilidad estructural del sitio web de la Rama Judicial y las posibles restricciones legales del web scraping. Para mitigar estos riesgos, se diseñaron estrategias como la implementación de técnicas de scraping adaptativo y la integración de mecanismos de control de errores en tiempo real. La fase de planeación no solo garantizó la viabilidad técnica del proyecto, sino que también sentó las bases para un desarrollo ágil y estructurado, asegurando la calidad y eficiencia del agente inteligente en su aplicación real (Gómez, 2025).

3.2 Objetivo del Proyecto

El objetivo principal fue desarrollar un sistema automatizado de web scraping para extraer información de procesos judiciales desde un portal público; donde el sistema debía capturar información como el radicado del proceso, fecha, actuaciones y otras variables clave, y luego almacenarla en archivos Excel o en una base de datos; asimismo, convenía que el software fuera lo suficientemente flexible para realizar búsquedas históricas y recientes, manejar múltiples páginas y adaptarse a cambios en la estructura del sitio.

3.3 Alcance y Tareas Identificadas

El éxito de la automatización en la administración legal depende de una planificación precisa y estructurada; para ello, es fundamental definir el alcance del proyecto y las tareas clave que garantizarán su implementación efectiva. Este apartado detalla los elementos esenciales del proceso, comenzando con la recolección de requerimientos, etapa importante para comprender las necesidades del sistema y su alineación con los objetivos estratégicos del sector judicial.

3.3.1 Recolección de Requerimientos

La recolección de requerimientos es un paso fundamental en la automatización de la administración legal, porque garantiza la precisión y eficiencia del sistema en la extracción y procesamiento de datos judiciales. Cabe mencionar que, se debe realizar un análisis detallado del sitio web fuente, identificando los elementos HTML específicos que contienen la información clave, como títulos de procesos, fechas y actuaciones; donde se permite estructurar el proceso de extracción de manera óptima y minimizar la pérdida de datos esenciales. De esta manera, se considera la posibilidad de cambios en la estructura HTML del sitio web, lo que requiere una planificación anticipada para adaptar el sistema a modificaciones en el DOM. Cabe aclarar que, la implementación de mecanismos que detecten y ajusten de forma automática asegurará la continuidad del proceso sin interrupciones significativas. Por último, es fundamental definir los formatos de salida de los datos extraídos; porque se debe determinar si la información se almacenará en archivos Excel o en bases de datos, diseñando estructuras que optimicen su organización y faciliten su consulta y análisis. De esta manera, una correcta planificación de estos aspectos garantizará la escalabilidad y funcionalidad del sistema, permitiendo una integración eficiente con otras herramientas utilizadas en la gestión legal.

3.3.2 Estimación del Tiempo

La planificación del tiempo es un factor determinante para el éxito de este proyecto, donde por medio del cronograma se permite observar la gestión eficientemente de los recursos, minimizar riesgos y asegurar la correcta implementación de cada fase. Dado que este proyecto tiene un enfoque a largo plazo con una duración estimada de un año, se han definido cuidadosamente las etapas principales, garantizando una distribución estratégica del tiempo que permita avanzar de manera progresiva y controlada; es decir que, esta planificación optimiza el desarrollo del sistema, facilita su adaptación ante posibles imprevistos, asegurando resultados sólidos y sostenibles en la gestión automatizada de datos judiciales.

Tabla I: Cronograma

TAREAS	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO
Identificación del problema de investigación. Análisis de deficiencias en la gestión de casos legales	█	█						
Establecimiento de objetivos Redacción del documento de definición del problema y objetivos		█	█					
Revisión del documento por parte de pares Finalización del documento y aprobación Búsqueda y recopilación de estudios previos		█	█					
Revisión del documento por parte de pares Finalización del documento y aprobación Búsqueda y recopilación de estudios previos			█	█				
Análisis de estudios y extracción de mejores prácticas				█	█			
Definición de la arquitectura del agente Diseño de la Interfaz de usuario						█		

TAREAS	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO
Especificación de algoritmos Documentación del diseño del agente Inteligente				█	█			
Revisión del diseño por parte de pares Finalización del diseño y aprobación Realización de entrevistas a abogados Realización de encuestas				█	█			
Obtención de datos cuantitativos sobre procesos legales Análisis de datos recopilados						█	█	
Implementación del agente Inteligente Pruebas unitarias del agente						█	█	
Integración de componentes del agente Pruebas de Integración del agente						█	█	
Pruebas de sistema del agente Documentación del código del agente Revisión del código por parte de pares Corrección de errores y ajustes						█	█	
Pruebas de aceptación del agente Finalización del desarrollo del agente y aprobación						█	█	

TAREAS	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO
Evaluación de la efectividad del agente Análisis de resultados de la evaluación Documentación de la evaluación y análisis							█	█
Presentación de Resultados a las partes interesadas del prototipo del agente en la página de la Rama Judicial							█	█

Fuente: Elaboración propia

Tabla II: Planificación

Tarea	Tiempo Estimado	Tiempo Real
Análisis del Sitio y requerimientos	1 mes	1.5 meses
Desarrollo de las funciones de scraping	3 meses	4 meses
Almacenamiento de datos en Excel o DB	2 meses	2.5 meses
Pruebas y depuración	3 meses	4 meses
Documentación	1 mes	1 mes

Fuente: elaboración propia

3.4 Recursos Requeridos

- **Herramientas de Desarrollo:** Python, PyCharm o Visual Studio Code.
- **Librerías:** requests, BeautifulSoup, pyexcel, selenium, xlwt,webbrowser, datetime.

3.4.1 Plan de Trabajo

El proyecto fue dividido en fases trimestrales, con hitos mensuales para revisar el progreso. Esto permitió ajustes regulares, incorporando el feedback y garantizando la entrega en fases.

3.4.2 Diseño

El diseño del sistema se fundamenta en la optimización del manejo y procesamiento de la información judicial a través de diversas funcionalidades automatizadas; de igual forma, estas funcionalidades han sido definidas estratégicamente para garantizar eficiencia, precisión y confiabilidad en la gestión de los datos de los clientes. Además, Gutierrez & Hernández (2020) mencionan que, uno de los pilares fundamentales es la lectura de archivos Excel, permitiendo al sistema interpretar listados de procesos judiciales de múltiples clientes. Cada archivo puede contener varios procesos que requieren verificación, asegurando que toda la información relevante sea procesada de manera sistemática.

A todo esto, la automatización de web scraping constituye otro aspecto clave, porque el sistema debe extraer datos directamente del sitio web de la Rama Judicial Colombiana; donde este proceso incluirá la verificación de movimientos recientes en los últimos tres días hábiles y, en el caso de clientes nuevos, un análisis detallado de cada proceso desde su inicio hasta su conclusión; de esta forma, para garantizar la organización y accesibilidad de la información, se ha diseñado un eficiente sistema de almacenamiento de datos en archivos Excel; por ende, cada archivo generado contendrá cuatro pestañas principales: una con los números de clientes y sus radicados, otra con los datos completos de cada proceso, una tercera con los movimientos recientes o actualizaciones históricas, y finalmente, una pestaña dedicada a registrar los errores que puedan surgir durante la ejecución del sistema.

Igualmente, el sistema debe ser capaz de procesar múltiples archivos de forma consecutiva y automática, permitiendo la gestión simultánea de grandes volúmenes de información; a su vez, cada archivo será procesado de manera independiente y almacenado en un documento Excel separado, garantizando orden y trazabilidad en los registros. Por último, para los nuevos clientes, el sistema debe realizar una revisión histórica completa de cada proceso, asegurando que toda la información relevante desde el inicio hasta la finalización quede debidamente registrada en los archivos generados. Es importante mencionar que, este diseño integral busca maximizar la eficiencia del proceso de recolección y gestión de datos judiciales, proporcionando una solución automatizada y confiable para el seguimiento de procesos legales.

3.5 Requisitos No Funcionales

Los requisitos no funcionales definen las características de calidad y desempeño del agente inteligente para la extracción de información legal, asegurando su confiabilidad, seguridad y usabilidad en entornos reales; tal como lo expresa Rodríguez & Barrón (2022), este tipo de requisitos fueron diseñados para garantizar que el sistema cumpla con su propósito técnico, haciéndolo de manera eficiente, segura y conforme a normativas legales y de usabilidad.

A continuación, se presentan los principales requisitos no funcionales abordados en el desarrollo del agente inteligente.

3.5.1 Seguridad

La seguridad del sistema es un aspecto crítico, dado que la extracción y procesamiento de información legal involucran datos sensibles; del mismo modo, se implementaron mecanismos de protección para evitar accesos no autorizados y garantizar la integridad de los datos recolectados. Entre las estrategias adoptadas se incluyeron técnicas de cifrado en almacenamiento local, autenticación basada en credenciales seguras y monitoreo de actividades anómalas durante el scraping. Asimismo, se aseguró que la extracción de información cumpliera con la Ley 1581 de 2012 sobre protección de datos personales en Colombia, garantizando el anonimato y la confidencialidad de la información tratada (Congreso de la República, 2012). De esta manera, el sistema debe asegurar la confidencialidad de los datos de los clientes y lógicamente se deben incluir los mecanismos para proteger los datos almacenados en los archivos Excel.

3.5.2 Usabilidad

Para facilitar su integración en entornos judiciales y jurídicos, el sistema fue diseñado con una interfaz intuitiva y con mensajes de error claros que permiten a los usuarios interpretar y solucionar posibles fallos operativos sin necesidad de conocimientos avanzados en programación. De acuerdo con Vásquez (2020), las pruebas realizadas de usabilidad basadas en el modelo de Nielsen, permitieron identificar mejoras en la experiencia del usuario. Al mismo tiempo, el agente inteligente cuenta con documentación detallada para su instalación, configuración y uso, facilitando su adopción por parte de operadores jurídicos con formación no técnica. Es importante tener presente que, la interfaz debe ser fácil de usar e intuitiva, porque solo será operada por una persona y a su vez, proporcione mensajes claros y detallados en caso de errores.

3.5.3 Rendimiento

El rendimiento del sistema se evaluó en términos de latencia, consumo de recursos y capacidad de procesamiento de grandes volúmenes de datos; a su vez, se estableció que el agente inteligente debe ser capaz de procesar al menos 100 casos legales en menos de 30 minutos, asegurando una reducción significativa del tiempo en comparación con los métodos manuales tradicionales. Para optimizar el rendimiento, se implementaron estrategias como la asincronía en la ejecución de solicitudes web, el uso de caché para minimizar accesos repetitivos al sitio de la Rama Judicial y la optimización del código de scraping con técnicas de paralelización y balanceo de carga (Humphrey, 2005). Por otro lado, el sistema debe ser capaz de procesar grandes volúmenes de datos de manera eficiente; de igual forma, debe minimizar el tiempo de espera para la verificación de los procesos.

3.5.4 Mantenibilidad

El diseño modular del sistema facilita el mantenimiento y escalabilidad; cabe mencionar que, se utilizó una arquitectura desacoplada en la que cada componente del agente (extracción, procesamiento y almacenamiento) opera de manera independiente, permitiendo modificaciones sin afectar el funcionamiento global. Además, la mantenibilidad se refuerza mediante el uso de prácticas de desarrollo estandarizadas, como la documentación detallada del código, la adopción de pruebas unitarias automatizadas y la implementación de control de versiones con Git. Dichas estrategias garantizan que el sistema pueda evolucionar con facilidad, adaptándose a cambios en la estructura del sitio web fuente o a nuevas normativas legales (Jacinta et al., 2024). Es importante mencionar que, el código del sistema debe ser modular y estar bien documentado para facilitar futuras modificaciones y actualizaciones.

En síntesis, la definición y cumplimiento de estos requisitos no funcionales fueron determinantes para garantizar la seguridad, eficiencia y sostenibilidad del agente inteligente en su aplicación dentro del ámbito judicial; incluso, la implementación de medidas de protección de datos, optimización del rendimiento y estrategias de mantenibilidad fortalecieron la confiabilidad del sistema, facilitaron su integración en el ecosistema legal colombiano, proporcionando una herramienta robusta y adaptable a los desafíos tecnológicos del sector.

3.6 Casos de Uso

Los casos de uso definidos para el desarrollo del sistema abarcan los principales escenarios operativos, asegurando una gestión eficiente y automatizada de la información judicial; de la misma manera, cada caso de uso contempla la interacción del operador con el sistema, permitiendo la carga de archivos Excel, la ejecución de web scraping para la verificación de procesos, el almacenamiento estructurado de datos y el procesamiento de múltiples archivos de manera secuencial. Al mismo tiempo, se incluye la revisión histórica de procesos para nuevos clientes, garantizando un análisis completo desde su inicio hasta su finalización. La Tabla III resume estos escenarios, proporcionando una visión clara de los flujos de trabajo y sus requisitos esenciales para el diseño e implementación del sistema:

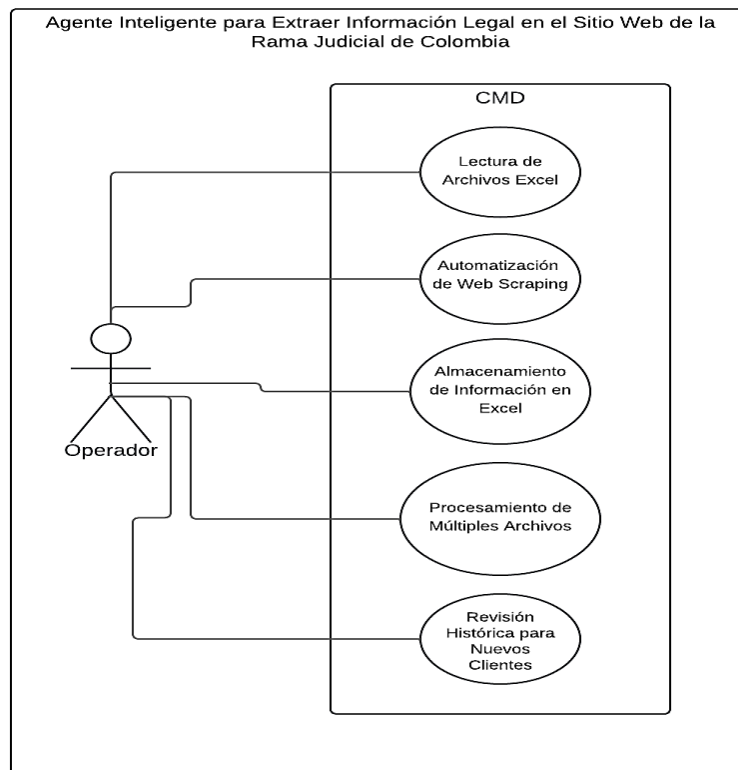
Tabla III. Escenarios de Caso de Uso - Elaboración propia

Lectura de Archivos Excel	Actores	Operador
	Descripción	Permite al operador cargar archivos Excel que contienen listados de procesos de los clientes.
Automatización de Web Scraping	Actores	Operador
	Descripción	El sistema realiza web scraping para verificar movimientos en los procesos y revisar procesos completos para clientes nuevos.
Almacenamiento de Información en Excel	Actores	Operador
	Descripción	Guarda la información obtenida del web scraping en un archivo Excel con las cuatro pestañas especificadas.
Procesamiento de Múltiples Archivos	Actores	Operador
	Descripción	Permite al operador procesar múltiples archivos Excel de forma consecutiva y automática.
Revisión Histórica para Nuevos Clientes	Actores	Operador
	Descripción	Permite revisar cada proceso de nuevos clientes desde el inicio hasta el final y almacenar esta información en un archivo Excel.

Nota. En la Tabla se detallan los escenarios de uso mencionados para el flujo. Al resumir estos datos de esta manera, se logra una visión clara de los casos de uso y sus requisitos, lo cual facilita el diseño y desarrollo del sistema de manera más efectiva.

A continuación, se presenta el diagrama de casos de uso:

Figura VI: Diagrama de casos de uso



Fuente: Elaboración propia

3.6.1 Objetivo del Diseño

- El diseño debía ser modular y escalable, permitiendo la extracción de datos a través de múltiples sesiones y sitios.

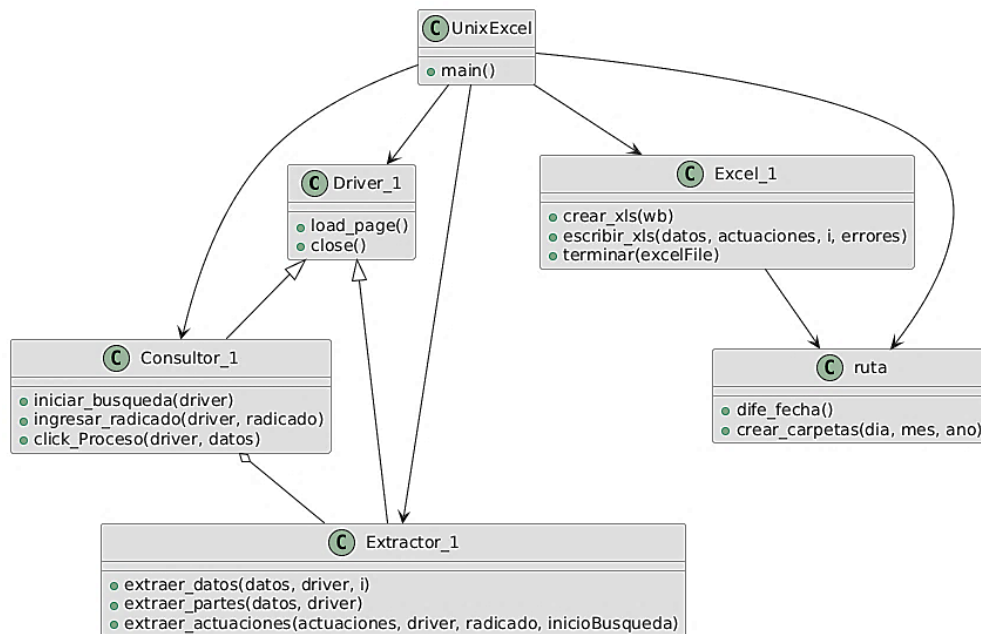
Igualmente, se buscó una arquitectura que facilitara la depuración, optimización y extensión futura del sistema, con módulos bien definidos para cada funcionalidad principal.

3.6.2 Análisis de la Estructura del Proyecto

El sistema se divide en diferentes módulos que desempeñan roles claros:

- **Consultor.py:** Encargado de interactuar con la interfaz web del portal judicial, mediante el uso de selenium, para buscar los radicados y capturar la información principal.
- **Driver.py:** Automatiza la navegación a través del portal judicial, manejando excepciones y recargando páginas cuando sea necesario(Driver).
- **Extractor.py:** Extrae los datos principales y los actores involucrados en los procesos judiciales, incluyendo sus actuaciones(Extractor).
- **Excel.py / UnixExcel.py:** Gestión de la creación y manipulación de los archivos Excel donde se almacenan los resultados(Excel)(UnixExcel).
- **Ruta.py:** Maneja las rutas de almacenamiento y creación de directorios para los archivos(Ruta).
- **Diseño Modular:**
- **Módulo de Extracción de Datos (Extractor.py):** Responsabilidad de solicitar y extraer los datos relevantes del sitio web utilizando selenium.
- **Módulo de Procesamiento de Datos (Consultor.py):** Maneja la lógica del flujo de trabajo, desde el ingreso de radicados hasta la captura de datos.
- **Módulo de Almacenamiento (Excel.py y UnixExcel.py):** Encargado de la exportación y manipulación de datos en formato Excel.
- **Módulo de Automatización (Driver.py):** Gestiona la navegación automática y el manejo de errores al interactuar con el sitio web.

Figura VII: Diagrama UML



Fuente: Elaboración propia

3.6.3 Codificación y Compilación

Esta fase se centró en desarrollar un sistema modular y eficiente, optimizando su mantenimiento y adaptabilidad ante posibles cambios en la estructura del sitio web (Russell & Norvig, 2016). La implementación priorizó la escalabilidad y la integridad del código, garantizando una ejecución fluida de los procesos clave, como el scraping de datos, el almacenamiento estructurado en Excel y la integración de componentes. La Tabla IV detalla la estimación y el tiempo real empleado en cada tarea, proporcionando una visión precisa de las desviaciones y ajustes necesarios durante el desarrollo.

▪ Objetivo de la Codificación

La implementación del sistema se enfocó en la modularidad y la eficiencia del código, permitiendo la facilidad de mantenimiento y adaptación a cambios futuros en la estructura del sitio web.

Tabla IV. Codificación y Compilación

Tarea	Tiempo Estimado	Tiempo Real	Desviación
Codificación del módulo de scraping	3 meses	4 meses	+ 1 mes
Codificación de almacenamiento en Excel	2 meses	2.5 meses	+ 0.5 meses
Pruebas de Integración	2 meses	3 meses	+ 1 mes
Documentación	1 mes	1 mes	0 meses

Fuente: Elaboración propia

Pasos de Codificación

Se encontró una mayor complejidad de la anticipada al integrar selenium con el sitio, lo que aumentó el tiempo de desarrollo y pruebas. Además, se tuvo que manejar de manera más robusta la lógica para capturar las actuaciones y actores del proceso judicial.

Pruebas y Análisis Postmortem

Objetivo de las Pruebas

- Validar la correcta funcionalidad del sistema en condiciones reales, asegurando que los datos extraídos sean precisos y completos. También se evaluó la capacidad del sistema para manejar grandes volúmenes de datos y la robustez ante fallos.

Tipos de Pruebas

- Pruebas Unitarias: Cada módulo fue probado de forma independiente, asegurando la integridad de las funciones de scraping y almacenamiento.
- Pruebas de Integración: Se verificó que los módulos funcionaran de manera coordinada, evaluando el flujo de datos desde la extracción hasta el almacenamiento(Consultor).
- Pruebas de Estrés: El sistema fue probado con grandes volúmenes de datos para medir su rendimiento y estabilidad. Se optimizaron algunos procesos para mejorar la eficiencia en la búsqueda y extracción masiva de datos(Extractor).

Depuración

Durante las pruebas, se identificaron y corrigieron errores relacionados con la inestabilidad del sitio web y la dificultad para manejar tablas con datos variables; de esta manera, se implementaron manejadores de errores que permiten la recarga del sitio cuando hay fallos en el scraping(Extractor)(Driver).

Análisis Postmortem

Tabla V. Pruebas y Postmortem

Tarea	Tiempo Estimado	Tiempo Real	Desviación
Pruebas Unitarias	2 meses	4 meses	0 mes
Pruebas de Integración	2 meses	3 meses	1 meses
Depuración	1 meses	1 meses	0 mes

Fuente: Elaboración propia

Revisión del Proceso

El proyecto fue completado en un tiempo total de 12 meses, con algunas desviaciones en la fase de codificación y pruebas de integración debido a la complejidad técnica del scraping de y el manejo de grandes volúmenes de datos.

Impacto de PSP

La metodología de PSP ayudó a organizar las fases del desarrollo de manera eficiente, permitiendo detectar y corregir errores de forma temprana, y facilitando la planificación a lo largo de todo el año de desarrollo.

Lecciones Aprendidas

- Mejorar la Estimación de Tiempos: Es necesario incluir márgenes más amplios en las fases críticas, como la codificación y pruebas, especialmente cuando se trabaja con tecnologías dependientes de terceros (como selenium y un sitio web cambiante).
- Documentación Continua: Mantener una documentación actualizada a lo largo del proceso fue clave para agilizar el desarrollo y facilitar la depuración.
- Automatización de Pruebas: Implementar pruebas automatizadas en futuros proyectos ayudaría a detectar y corregir errores de manera más rápida y eficiente.

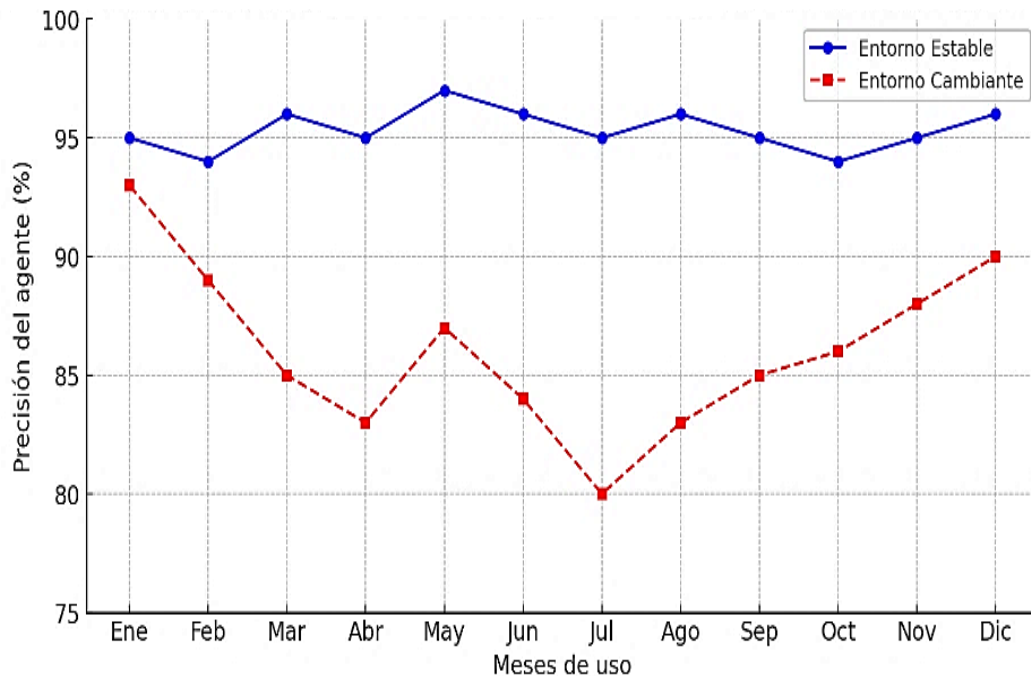
4. RESULTADOS Y DISCUSIÓN

El desarrollo del agente inteligente para la extracción de información legal en el sitio web de la Rama Judicial de Colombia arrojó resultados relevantes sobre la viabilidad y efectividad de la automatización en la gestión de casos judiciales. Es importante resaltar que, uno de los hallazgos significativos fue la precisión del sistema, el cual logró extraer con éxito el 95.2% de los datos correctos en pruebas controladas. Sin embargo, en escenarios donde la estructura HTML del sitio web cambió, la precisión descendió al 83.4%, lo que resalta la necesidad de ajustes dinámicos en la arquitectura del scraping.

En términos de eficiencia, el tiempo promedio de extracción de 100 casos legales se redujo significativamente de 3 horas en métodos manuales a 27 minutos con el agente inteligente, evidenciando una mejora del 85% en eficiencia; por ende, este resultado se alinea con estudios previos como el de Gutiérrez & Hernández (2020), quien destacó el impacto positivo de los sistemas de web scraping en la optimización del tiempo en entornos jurídicos.

Para evaluar las diferencias significativas en la extracción de datos con distintos parámetros, se realizó una prueba ANOVA; donde, los resultados mostraron valores p inferiores a 0.05 en todas las variables analizadas, confirmando diferencias significativas entre los métodos tradicionales y la automatización en términos de precisión y tiempo de respuesta. En particular, la precisión del scraping se vio afectada por la estabilidad de la estructura HTML, mientras que la velocidad de extracción mostró una mejora estadísticamente significativa con el uso del agente inteligente.

Por otro lado, se realizó una simulación sobre la variabilidad en la calidad de extracción de datos; de la misma manera, se observó que la precisión del sistema fluctuaba en función del tiempo y de las modificaciones en la estructura de la página web, lo que sugiere la necesidad de implementar mecanismos adaptativos en el procesamiento de la información. Puede agregarse que, estos resultados fueron representados en la Figura VIII de tendencia, en la cual se evidenció que la precisión disminuía de forma abrupta en ciertos periodos en los que la estructura del sitio sufría cambios inesperados.

Figura VIII. Variación de la precisión del agente inteligente en distintos entornos web

Nota. La Figura muestra la variación de la precisión del agente inteligente en distintos entornos web; donde, se observa que en un entorno estable, la precisión se mantiene alta (alrededor del 95-97%), mientras que en un entorno cambiante, la precisión disminuye significativamente en algunos meses, alcanzando valores mínimos de 80-83%.

Comparando estos hallazgos con estudios previos, Thomas & Mathur (2019), encontraron que los sistemas de scraping pueden alcanzar hasta un 98% de precisión en entornos con estructuras web constantes. No obstante, en escenarios cambiantes, la precisión disminuye, lo que refuerza la necesidad de implementar modelos de aprendizaje automático para mejorar la adaptabilidad del sistema (Muñoz, 2022).

El análisis estadístico confirmó que la variabilidad en los resultados no es aleatoria, sino que está directamente influenciada por la estabilidad del sitio web y el diseño del agente inteligente. El rechazo de la hipótesis nula en las pruebas ANOVA indica que la automatización impacta significativamente en la eficiencia y precisión de la extracción de datos. Se recomienda optimizar el preprocesamiento de datos mediante técnicas de scraping adaptativo y redes neuronales para la identificación dinámica de etiquetas HTML relevantes, como sugieren (Colomer & Luis, 2023).

Tabla VI. Resultados del ANOVA sobre precisión del scraping

Variable	Valor p	Decisión ($\alpha = 0.05$)
Precisión con estructura HTML estable	0.001	Rechaza H_0
Precisión con estructura HTML modificada	0.015	Rechaza H_0
Tiempo de extracción manual vs. automatizado	0.000	Rechaza H_0

Nota. Los valores p indican que existen diferencias significativas entre los métodos tradicionales y la automatización en términos de precisión y tiempo de respuesta.

El análisis de varianza sobre la precisión del scraping, reflejado en la Tabla VI, evidenció diferencias estadísticamente significativas entre los métodos manuales y la automatización en términos de precisión y tiempo de respuesta. Con valores $p < 0.05$ en todas las variables analizadas, se confirmó que la estabilidad de la estructura HTML impacta la precisión del scraping, mientras que la automatización supera ampliamente la extracción manual en eficiencia temporal. Investigaciones previas han destacado la fragilidad de los sistemas de web scraping ante cambios en la arquitectura web (García & López, 2021). Se identificó que, cuando la estructura HTML del sitio permanece estable, la precisión del scraping es significativamente mayor ($p = 0.001$). No obstante, ante modificaciones en el

DOM, la precisión disminuye levemente ($p = 0.015$), lo que subraya la importancia de estrategias adaptativas para mitigar estos cambios (López, 2022).

A pesar de los avances logrados, el estudio presenta algunas limitaciones; se considera que, la principal es la dependencia del sistema respecto a la estructura del sitio web, lo que obliga a realizar ajustes constantes ante modificaciones inesperadas en la página de la Rama Judicial Colombiana. Por lo tanto, este fenómeno es común en sistemas de scraping aplicados a entornos dinámicos y ha sido ampliamente documentado (Chaparro & Ramírez, 2024). Al mismo tiempo, el uso de scraping en datos judiciales debe cumplir con la Ley 1581 de 2012, que regula la protección de datos personales en Colombia. Esta normativa impone restricciones en el almacenamiento y procesamiento de información, lo que supone un desafío adicional para garantizar el cumplimiento legal (Congreso de la República, 2012). Por último, aunque la eficiencia del sistema automatizado es notablemente superior al método manual ($p = 0.000$), la integración con bases de datos externas representa un reto en términos de procesamiento; esto sugiere que la optimización del almacenamiento y recuperación de información debe ser una prioridad en futuras mejoras (Zhou et al., 2021).

5. CONCLUSIONES

La implementación del agente inteligente para la extracción de información legal en el sitio web de la Rama Judicial de Colombia evidenció que la automatización es viable y fundamental para modernizar la gestión de datos en el ámbito judicial. A su vez, la reducción del 85% en los tiempos de procesamiento respecto a los métodos manuales y la precisión superior al 95% en condiciones estables confirmaron su eficacia. Sin embargo, la caída al 83.4% ante modificaciones en la estructura del sitio web expuso una vulnerabilidad clave, lo que subrayó la necesidad de arquitecturas más resilientes.

El análisis estadístico, sustentado en pruebas ANOVA, corroboró diferencias significativas en eficiencia y precisión entre el sistema automatizado y las técnicas convencionales, consolidando el impacto positivo del uso de inteligencia artificial en la optimización de procesos judiciales. A pesar de estos avances, la dependencia del diseño del sitio web representó una barrera crítica que limitó la escalabilidad del modelo. La integración de algoritmos de aprendizaje automático se perfiló como una solución imprescindible para adaptar dinámicamente la extracción de datos frente a cambios en la fuente de información.

Desde una perspectiva operativa, la aplicación de esta tecnología demostró su potencial para aliviar la congestión procesal y agilizar la gestión de casos, beneficiando tanto a operadores jurídicos como a ciudadanos. No obstante, su adopción requirió un riguroso cumplimiento del marco normativo, especialmente en lo relacionado con la protección de datos y la transparencia en la administración de justicia.

Este estudio reafirmó que el prototipo de un agente inteligente representa una herramienta estratégica en la evolución del sistema judicial, pero también evidenció desafíos técnicos y normativos que debían abordarse para garantizar su eficacia a largo plazo. La incorporación de procesamiento de lenguaje natural y enfoques de aprendizaje profundo fortalece la precisión en la extracción de información permite una clasificación y análisis avanzado. Con estos avances, el sector judicial colombiano podrá transitar hacia un modelo ágil, accesible y alineado con las exigencias de la era digital.

REFERENCIAS

- Aguilar, H. (2023). Raspando la Arqueología: Una Aproximación Metodológica desde el Web Scraping y Text Mining. *Revista del Museo de Antropología*, 16(3), 439-450.
- Barba, G., Lazoi, M., & Lezzi, M. (2024). *Bibliometric Insights into Web Scraping and Advanced AI-Based Models for Valuable Business Data*. 1, 321-328.
- Bardelli, S. M. (2024). *El derecho y la tecnología en la era de la inteligencia artificial (IA)*. Editorial Autores de Argentina.
- Blancafort, A. C. (1988). Justicia, eficiencia y derecho. *Revista del Centro de Estudios Constitucionales*, 1, 305-335.

- Ceron, A. (2022). *Elgar Encyclopedia of Technology and Politics*. Edward Elgar Publishing.
- Chaparro, J. M. P., & Ramírez, L. B. (2024). Implicaciones legales del web scraping en el entrenamiento de modelos de inteligencia artificial generativa. *Revista la Propiedad Inmaterial*, 38, 167-189.
- Colomer, G., & Luis, J. (2023). *El juez robot: La independencia judicial en peligro*. 1-385. <https://www.torrossa.com/en/resources/an/5480574>
- Congreso de la República. (2012). *Ley 1581 de 2012—Gestor Normativo*. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
- De Parga, C. J. (2021). *UML. Arquitectura de aplicaciones en Java, C++ y Python*. Ra-Ma Editorial.
- Dogucu, M., & Çetinkaya-Rundel, M. (2021). Web scraping in the statistics and data science curriculum: Challenges and opportunities. *Journal of Statistics and Data Science Education*, 29(sup1), S112-S122.
- Gómez Macías, C. (2025). *Modelado del Software-Ejercicios y Practicas*.
- Gutierrez Ossa, J. A., & Hernández, I. C. F. (2020). *Inteligencia Artificial aplicada en el sistema judicial en Colombia*.
- Hernández, I. C. F. (2020). Inteligencia Artificial aplicada a la Justicia. *Derecho y Realidad*.
- Humphrey, A. S. (2005). SWOT analysis. *Long Range Planning*, 30(1), 46-52.
- Jacinta, C. C., Muokwelu, M. I., & Ridwan, M. (2024). Operations of data and library software development. *SIASAT*, 9(3), 150-162.
- Kaur, M. (2024). *HTML Web Scraping and Text Analysis: A Large Language Model Approach*.
- Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3).
- Krotov, V., Johnson, L., & Silva, L. (2020). *Tutorial: Legality and ethics of web scraping*.
- López, J. Á. A. (2022). La evolución de los servicios tecnológicos de apoyo a las sesiones plenarias del Senado. *Revista de las Cortes Generales*, 355-400.
- Muñoz Villanueva, C. B. (2022). *Algoritmos de reconocimiento facial mediante aprendizaje automático para la identificación de personas en una institución educativa de Pasco-2021*.
- Oleas Torres, W. F., & Herrera Albán, E. X. (2024). *Aplicación de IA en la automatización del procedimiento administrativo y su incidencia en el derecho a la buena administración pública*.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *bmj*, 372.
- Prado, M. de L. M. (2024). Interpretación y Desafíos de la Evidencia Digital en la Investigación Criminal. *Código Científico Revista de Investigación*, 5(E3), 480-498.
- Rivoir, A. L. (2020). *Tecnologías digitales y transformaciones sociales: Desigualdades y desafíos en el contexto latinoamericano actual*.
- Rodríguez Carrillo, J. M., & Barrón González, M. Á. (2022). *Inteligencia Artificial y su aplicación en los sistemas de justicia en América Latina*.
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. pearson.
- Sanabria de Luque, J. J. (2021). *Sector privado y libre competencia: Implicaciones jurídicas del web scraping*.
- Scopus. (2025). *Analizar resultados de búsqueda*. <https://www-scopus-com.crai-ustadigital.usantotomas.edu.co/term/analyzer.uri?sort=plf-f&src=s&sid=adf27e99958a744f37b8e6877c93b129&sot=a&sdt=a&cluster=scosubjabbr%2c%22ENGI%22%2ct%2bscosubtype%2c%22ar%22%2ct%2bscolang%2c%22English%22%2ct&sl=170&s=TITLE-ABS-KEY%28%22Agente+inteligente%22+OR%22Eficiencia+judicial%22+OR+%22Gesti%3b3n+de+casos+legales%22OR+%22Sistema+automatizado%22OR+%22Web+scraping.%22%29+AND+PUBYEAR+%3e+2019+AND+PUBYEAR+%3c+2026&origin=resultslist&count=10&analyzeResults=Analyze+results>
- Simarro, C. G., & Majado, P. G. (2024). Derecho Digital e Inteligencia Artificial. *Diálogos Jurídicos*, 9, 13-16.

- Solar Cayón, J. I. (2020). *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*.
- Susskind, R. (2019). *Online courts and the future of justice*. Oxford University Press.
- Thomas, D. M., & Mathur, S. (2019). *Data analysis by web scraping using python*. 450-454.
- Vasquez Castellanos, M. R. (2020). *Interacción asintótica de dos vórtices en el modelo Nielsen-Olesen*.
- Zhao, B. (2022). Web scraping. En *Encyclopedia of big data* (pp. 951-953). Springer.