

# **ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE INGENIERÍA DE SISTEMAS**

**DESARROLLO DE UN SISTEMA DE TOMA DE DECISIONES PARA  
APOYO A LOS INVESTIGADORES ECUATORIANOS BASADO EN  
MOTORES DE BÚSQUEDA Y RECOMENDACIÓN**

**AUTOMATIZACIÓN DEL PROCESO DE EXTRACCIÓN,  
DEPURACIÓN Y ALMACENAMIENTO DE DATOS DE SCOPUS  
MEDIANTE LA API DE ELSEVIER**

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO  
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO/A EN  
SOFTWARE**

**JHON FERNANDO SANGOPANTA NAULA**

jhon.sangopanta@epn.edu.ec

**DIRECTOR: PH.D. LORENA KATHERINE RECALDE CERDA**

lorena.recalde@epn.edu.ec

**DQM, Julio 2024**

## **CERTIFICACIONES**

Yo, Jhon Fernando Sangopanta Naula declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

Jhon Fernando Sangopanta Naula

Certifico que el presente trabajo de integración curricular fue desarrollado por Jhon Fernando Sangopanta Naula, bajo mi supervisión.

Lorena Katherine Recalde Cerda  
DIRECTOR

## **DECLARACIÓN DE AUTORÍA**

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

JHON FERNANDO SANGOPANTA NAULA

LORENA KATHERINE RECALDE CERDA

JOFFRE ALEXANDER CÓNDOR TIPÁN

ROMMEL PAÚL MASABANDA CASTRO

DANNY RUBÉN CABRERA AGUILAR

## **DEDICATORIA**

A mis padres, por su amor incondicional y su constante apoyo a lo largo de esta maravillosa etapa de mi vida. Su presencia constante y su guía han sido fundamentales para alcanzar mis metas. Gracias por enseñarme el verdadero valor del esfuerzo y la perseverancia, y por ser mi fuente inagotable de inspiración y fortaleza. Su ejemplo y aliento me han mostrado que con dedicación y amor, todo es posible. Este logro es tan suyo como mío, y siempre llevaré con orgullo el legado de su apoyo y dedicación.

A mis hermanos, que a pesar de la distancia siempre han estado presentes, por su cariño y por ser mi refugio en los momentos difíciles.

A mi tío, por abrirme las puertas de su hogar y brindarme su apoyo incondicional.

Finalmente, a mis abuelitos en el cielo, por su amor eterno y sabiduría. Aunque no estén físicamente presentes, su legado y sus enseñanzas siguen guiándome y dándome fuerzas. Esta dedicación también es para ustedes, por todo lo que me han dado y por estar siempre en mi corazón.

## **AGRADECIMIENTO**

Agradezco profundamente a mi tutora, la Dra. Lorena Recalde, por su inquebrantable apoyo, paciencia y dedicación. Su valiosa gestión de recursos y orientación ha sido fundamental para la realización de este proyecto.

A los miembros del Laboratorio de Analítica Avanzada de Datos - ADA, por brindarnos un espacio de trabajo excepcional y por su colaboración en el desarrollo de este proyecto.

A mis profesores, por su enseñanza y su orientación invaluable durante mi formación académica. Sus conocimientos y consejos han sido una guía constante en mi desarrollo profesional.

A mi familia, por su amor incondicional y su constante apoyo. Su confianza y aliento han sido una fuente constante de inspiración a lo largo de este viaje.

A mis amigos de la universidad, por su amistad, compañía y por su apoyo a lo largo de la carrera sus conocimientos y experiencias compartidas han sido de gran motivación no solo para mi formación académica sino también para mi desarrollo personal.

A mis compañeros de trabajo, por su colaboración y su amistad. Trabajar juntos ha hecho de este proyecto una experiencia enriquecedora y gratificante.

# Índice general

<b>1. INTRODUCCIÓN</b>	<b>1</b>
1.1. Planteamiento del problema . . . . .	1
1.2. Justificación teórica . . . . .	2
1.3. Justificación metodológica . . . . .	3
1.4. Objetivos . . . . .	3
1.4.1. Objetivo General . . . . .	3
1.4.2. Objetivos específicos . . . . .	3
1.5. Alcance . . . . .	4
1.6. Marco Teórico . . . . .	4
1.6.1. Scopus . . . . .	4
1.6.2. Redes de Co-Autoría . . . . .	5
1.6.3. Bases de Datos Orientadas a Grafos . . . . .	5
1.6.4. TF-IDF . . . . .	6
1.6.5. Cosine Similarity . . . . .	6
1.6.6. SCRUM . . . . .	6
1.6.7. Arquitectura Hexagonal . . . . .	8
1.6.8. Herramientas Utilizadas . . . . .	9
<b>2. METODOLOGÍA</b>	<b>10</b>
2.1. Sprint 0 . . . . .	10
2.2. Sprint 1 . . . . .	14
2.2.1. Introducción . . . . .	14
2.2.2. Objetivos del Sprint . . . . .	14
2.2.3. Planificación . . . . .	15
2.2.4. Implementación . . . . .	18
2.2.5. Revisión y Retrospectiva . . . . .	24

2.3. Sprint 2 . . . . .	25
2.3.1. Introducción . . . . .	25
2.3.2. Objetivos . . . . .	25
2.3.3. Planificación . . . . .	25
2.3.4. Implementación . . . . .	28
2.3.5. Revisión y Retrospectiva . . . . .	31
2.4. Sprint 3 . . . . .	32
2.4.1. Introducción . . . . .	32
2.4.2. Objetivos . . . . .	32
2.4.3. Planificación . . . . .	32
2.4.4. Implementación . . . . .	34
2.4.5. Revisión y Retrospectiva . . . . .	43
2.5. Sprint 4 . . . . .	44
2.5.1. Introducción . . . . .	44
2.5.2. Objetivos . . . . .	44
2.5.3. Planificación . . . . .	44
2.5.4. Implementación . . . . .	46
2.5.5. Revisión y Retrospectiva . . . . .	49
2.6. Sprint 5 . . . . .	50
2.6.1. Introducción . . . . .	50
2.6.2. Objetivos . . . . .	50
2.6.3. Planificación . . . . .	50
2.6.4. Implementación . . . . .	52
2.6.5. Revisión y Retrospectiva . . . . .	60
<b>3. RESULTADOS, CONCLUSIONES Y RECOMENDACIONES</b>	<b>61</b>
3.1. Resultados . . . . .	61
3.1.1. Capacidades de Centinela . . . . .	61
3.1.2. Prueba con usuarios finales . . . . .	62
3.2. Conclusiones . . . . .	64
3.3. Recomendaciones . . . . .	65
<b>4. REFERENCIAS BIBLIOGRÁFICAS</b>	<b>66</b>
<b>5. ANEXOS</b>	<b>69</b>

# Índice de figuras

1.1. Diagrama general propuesto para la integración de ResNet y ResearchDecide, mostrando la estructura y la interconexión de los componentes principales del sistema. . . . .	2
2.1. Proyecto en Azure Boards . . . . .	12
2.2. Arquitectura Hexagonal aplicada en el modulo del motor de búsqueda . . . . .	13
2.3. Arquitectura hexagonal aplicada al modulo de consenso . . . . .	14
2.4. Criterios de aceptación de la historia de usuario HU-SE-01 . . . . .	17
2.5. Criterios de aceptación de la historia de usuario HU-SE-02 . . . . .	17
2.6. Planificación de tareas del Sprint 1 . . . . .	18
2.7. Mockup de la página de inicio . . . . .	19
2.8. Estructura de carpetas de la página de inicio . . . . .	20
2.9. Página de inicio . . . . .	21
2.10. Mockup de la página de resultados . . . . .	22
2.11. Mockup de la página de detalle del Artículo . . . . .	23
2.12. Routing de la página de detalle de Artículo . . . . .	23
2.13. Página de detalle de Artículo . . . . .	24
2.14. Criterios de Aceptación HU-SE-03 . . . . .	27
2.15. Criterios de Aceptación HU-SE-04 . . . . .	27
2.16. Planificación de tareas del sprint 2 . . . . .	27
2.17. Mockup de los detalles de un autor . . . . .	28
2.18. Mockup de los artículos publicados por un autor . . . . .	29
2.19. Interfaz en Typescript para la creación de la red de coautoría . . . . .	29
2.20. Red de coautoría . . . . .	30
2.21. Tabla de artículos publicados por un autor . . . . .	31
2.22. Criterios de aceptación HU-SE-02 . . . . .	34



2.23. Criterios de aceptación HU-SE-05 . . . . .	34
2.24. Criterios de aceptación HU-SE-06 . . . . .	34
2.25. Archivo Dockerfile . . . . .	35
2.26. Archivo Docker Compose . . . . .	35
2.27. Salida del comando docker ps . . . . .	36
2.28. Aplicaciones de Django . . . . .	36
2.29. Modelo de la base de datos . . . . .	37
2.30. Capa de dominio de la aplicación . . . . .	37
2.31. Modelo Author implementado con Neomodel . . . . .	38
2.32. Repositorio AuthorRepository . . . . .	39
2.33. Servicio AuthorService . . . . .	40
2.34. Caso de uso FindAuthorsByQuery . . . . .	40
2.35. Vista AuthorViews . . . . .	41
2.36. Documentación de la API con Swagger . . . . .	42
2.37. Diagrama de secuencia de la funcionalidad de búsqueda de autores . . . . .	43
2.38. Base de datos Neo4j . . . . .	43
2.39. Diagrama de secuencia para la generación del Corpus . . . . .	46
2.40. Diagrama de secuencia para la generación del Modelo de TF-IDF . . . . .	47
2.41. Diagrama de secuencia para la obtención de los Artículos Relevantes . . . . .	48
2.42. Diagrama de secuencia para la obtención de la red de coautoría de un autor . . . . .	49
2.43. Diagrama de secuencia para la obtención de los Artículos en los que ha colaborado un autor . . . . .	49
2.44. Criterios de Aceptación HU-SE-07 . . . . .	50
2.45. Criterios de Aceptación HU-SE-08 . . . . .	51
2.46. Historias de Usuario en Azure Boards para el Sprint 5 . . . . .	52
2.47. Flujo de la Automatización . . . . .	52
2.48. Estructura de las Variables de Entorno . . . . .	53
2.49. Modelo Topic . . . . .	54
2.50. Diagrama de Secuencia de la Integración con la API de Elsevier . . . . .	55
2.51. Ejemplo de JSON de un Artículo . . . . .	56
2.52. Diagrama de Secuencia de la Integración con la API de Author Retrieval . . . . .	57
2.53. Configuración del Sistema de Logging . . . . .	58
2.54. Login de Administrador . . . . .	58
2.55. Panel de Administración . . . . .	59

2.56. Panel de Logs . . . . .	60
3.1. Documentación detallada . . . . .	61
3.2. Documentación de la API generada con Swagger . . . . .	62
3.3. Resultados de la pregunta 1 del cuestionario SUS . . . . .	63
3.4. Escala de Jeff Sauro . . . . .	64

# Índice de Tablas

1.1. Herramientas utilizadas . . . . .	9
2.1. Historias de Usuario del sprint 1 . . . . .	16
2.2. Historias de Usuario del Sprint 2 . . . . .	26
2.3. Historias de Usuario del sprint 3 . . . . .	33
2.4. Historias de Usuario del sprint 4 . . . . .	45
2.5. Historias de Usuario del Sprint 5 . . . . .	51
3.1. Cuestionario SUS . . . . .	63

## RESUMEN

Este documento aborda la creciente necesidad de colaboración y eficiencia en la investigación científica, reflejada en el aumento de documentos y la complejidad de las interacciones entre expertos. Para enfrentar estos desafíos, se desarrollaron dos aplicaciones complementarias: ResNet, un motor de búsqueda que visualiza redes de coautoría entre investigadores, y ResearchDecide, un sistema de recomendación para la formación de equipos de investigación.

La integración de estas herramientas dio lugar a Centinela, una solución integral diseñada para optimizar la colaboración científica. Centinela combina las capacidades de ResNet y ResearchDecide, ofreciendo una plataforma que facilita tanto la visualización de redes de investigación como la gestión de equipos científicos. Para el frontend, se utilizó Angular, proporcionando una interfaz dinámica y receptiva. El backend se construyó con Django Rest Framework, asegurando una API robusta y escalable, y se implementó una arquitectura hexagonal para facilitar la escalabilidad y el mantenimiento del sistema.

El proyecto se centra en automatizar la extracción, limpieza y almacenamiento de datos, mejorando la integridad y consistencia de la información, reduciendo errores humanos y aumentando la eficiencia. Centinela representa un avance significativo en la integración de herramientas de investigación, abordando eficazmente los problemas de colaboración y gestión de datos en el ámbito científico.

**PALABRAS CLAVE** - redes de coautoría, arquitectura limpia, integración, elsevier, automatización

## ABSTRACT

This document addresses the growing need for collaboration and efficiency in scientific research, highlighted by the increase in documents and the complexity of interactions among experts. To tackle these challenges, two complementary applications were developed: ResNet, a search engine that visualizes co-authorship networks among researchers, and ResearchDecide, a recommendation system for forming research teams.

The integration of these tools led to Centinela, a comprehensive solution designed to optimize scientific collaboration. Centinela combines the capabilities of ResNet and ResearchDecide, providing a platform that facilitates both the visualization of research networks and the management of scientific teams. Angular was used for the frontend, offering a dynamic and responsive interface, while the backend was built with Django Rest Framework to ensure a robust and scalable API. An hexagonal architecture was also implemented to support scalability and maintainability.

The project focuses on automating data extraction, cleaning, and storage processes to improve data integrity and consistency, reduce human error, and enhance efficiency. Centinela represents a significant advance in integrating research tools, effectively addressing issues of collaboration and data management in the scientific field.

**KEYWORDS** - co-authorship networks, clean architecture, integration, elsevier, automation

# Capítulo 1

## INTRODUCCIÓN

### 1.1. Planteamiento del problema

Un análisis de la Universidad San Francisco de Quito (USFQ) revela que la investigación científica en Ecuador ha avanzado significativamente en los últimos 100 años, con 30.205 documentos en 27 áreas temáticas y colaboraciones de 84 países [1].

En respuesta al aumento de la producción científica y a la creciente necesidad de colaboración entre expertos, se desarrollaron dos aplicaciones que intentan abordar esta problemática, ResNet [2] y ResearchDecide [3] [4].

ResNet es una herramienta que emplea modelos de minería de datos para buscar y mostrar redes de investigadores asociados con instituciones ecuatorianas y sus áreas académicas. Su principal objetivo es representar visualmente las redes de coautoría entre investigadores ecuatorianos, o aquellos que han participado en publicaciones con afiliaciones ecuatorianas, utilizando datos extraídos de Scopus [5].

Por otro lado, ResearchDecide se centra en la complejidad y la duración del proceso para lograr un consenso, en cuanto a toma de decisiones y la afiliación de miembros en un equipo de investigación.

En este proyecto, se propone combinar ambas plataformas para aprovechar al máximo sus funcionalidades. De esta integración nace Centinela, una aplicación robusta y capaz de abordar el problema en cuestión.

Centinela estará compuesto por cuatro componentes principales, cada uno con una función específica. Aunque cada componente puede operar de manera independiente, es fundamental que se comuniquen entre sí para su correcto funcionamiento. En la Figura 1.1, se ilustra la estructura y las interacciones entre estos componentes.

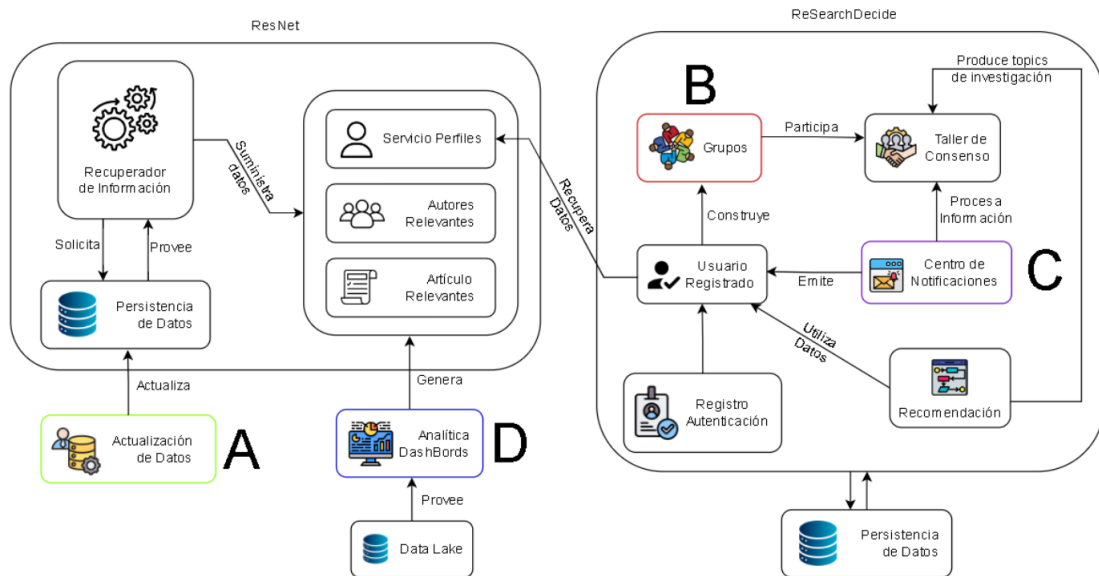


Figura 1.1: Diagrama general propuesto para la integración de ResNet y ResearchDecide, mostrando la estructura y la interconexión de los componentes principales del sistema.

Debido a que actualmente en la plataforma de ResNet los procesos de extracción, limpieza y almacenamiento de datos se hacen de manera manual, como parte del Componente A, se plantea la automatización de dichos procesos para asegurar la integridad de los datos, así como su consistencia. La automatización permitirá reducir errores humanos, aumentar la eficiencia y garantizar que los datos se manejen de manera uniforme.

## 1.2. Justificación teórica

La creciente producción científica a nivel mundial y la necesidad de colaboración entre investigadores demandan herramientas innovadoras para facilitar la conexión y el intercambio de conocimiento. Centinela, una plataforma que integra ResNet y ResearchDecide, responde a esta necesidad al proporcionar una visión integral de las redes de coautoría en Ecuador y al facilitar la formación de nuevos equipos de investigación.

Basándonos en la teoría de redes sociales [6], planteamos que la estructura de las redes de coautoría influye significativamente en la productividad científica y en la difusión del conocimiento. Al analizar estas redes, Centinela permitirá identificar patrones de colaboración, detectar vacíos de conocimiento y facilitar la formación de nuevos grupos de investigación.

La automatización del proceso de extracción de datos, fundamentada en los principios de la minería de datos, garantizará la calidad y la consistencia de la información, lo que a su vez permitirá realizar análisis más precisos y confiables. En última instancia, Centinela

contribuirá a fortalecer el sistema científico y a fomentar la colaboración en el Ecuador.

### **1.3. Justificación metodológica**

SCRUM se utilizará como la metodología principal para este proyecto debido a su capacidad para gestionar el desarrollo de software de manera iterativa y adaptativa [7]. SCRUM facilita el trabajo en ciclos cortos, conocidos como sprints [8], que permiten la entrega continua de valor y una rápida adaptación a los cambios en los requisitos. Esto es especialmente útil en entornos dinámicos, donde los requisitos pueden evolucionar con frecuencia.

SCRUM proporciona visibilidad constante del progreso del proyecto y fomenta la colaboración y comunicación entre los miembros del equipo y los stakeholders. Su enfoque en la mejora continua y la gestión temprana de riesgos ayuda a identificar y solucionar problemas de manera proactiva. Al integrar SCRUM en el proyecto, se asegura una mayor eficiencia en la entrega de productos funcionales y una respuesta efectiva a los desafíos emergentes.

### **1.4. Objetivos**

#### **1.4.1. Objetivo General**

Integrar en una única plataforma los sistemas informáticos (independientes) Resnet y Research Decide, desarrollados previamente. Además, se busca incorporar un módulo de administrador para la gestión eficiente del sistema, permitiendo, de forma automática, la extracción y limpieza de los datos obtenidos desde Scopus.

#### **1.4.2. Objetivos específicos**

1. Analizar y comprender a detalle los sistemas Resnet y ReSerchDecide para identificar sus funcionalidades, requerimientos y posibles áreas de mejora.
2. Diseñar e implementar una arquitectura de integración que permita la interoperabilidad entre los diferentes sistemas, garantizando una coherencia de datos y compatibilidad de las plataformas.
3. Implementar la automatización del proceso de extracción, limpieza y almacenamiento de datos de Scopus con el fin de garantizar la actualización constante del sistema.



## **1.5. Alcance**

El alcance principal del proyecto se enfoca en la integración de los sistemas Resnet y ResearchDecide, para luego servir como base del desarrollo de los otros componentes de dicho proyecto como se puede ver en la Figura 1.1. El objetivo principal es asegurar la interoperabilidad entre ambos sistemas y llevar a cabo una renovación completa de la interfaz de usuario, centralizándola en una única plataforma. Esto implica la unificación de la experiencia del usuario y la optimización de la eficiencia operativa al consolidar funcionalidades clave en un solo punto de acceso.

También se prevé automatizar el proceso del Componente A, que incluye la extracción, limpieza y almacenamiento de datos, mediante el uso de las APIs proporcionadas por Elsevier [9] en su portal para desarrolladores.

Para el desarrollo de este componente, no se contemplan mejoras en los algoritmos existentes de ResNet, ya que el enfoque estará exclusivamente en la integración de ResNet y ResearchDecide en Centinela. Las optimizaciones y ajustes a los algoritmos de ResNet no se incluirán en este proyecto. Finalmente un sistema de autenticación robusto no forma parte de este componente.

## **1.6. Marco Teórico**

### **1.6.1. Scopus**

Scopus es una destacada base de datos bibliográfica y herramienta de indexación científica ampliamente utilizada en el ámbito académico y de la investigación. Desarrollada por la editorial Elsevier, Scopus abarca diversas disciplinas académicas, incluyendo ciencias sociales, ciencias de la salud, ingeniería, ciencias naturales y ciencias exactas.

Esta plataforma proporciona un extenso repositorio de información, que incluye resúmenes y citas de artículos científicos publicados en revistas revisadas por expertos, conferencias y libros. Su alcance internacional abarca publicaciones de diversas partes del mundo, lo que la convierte en una herramienta valiosa para investigadores que buscan acceder a una amplia gama de literatura científica.

Scopus no solo indexa la información, sino que también ofrece herramientas analíticas que permiten evaluar el impacto de las publicaciones y la productividad de los investigadores. Asimismo, su capacidad para rastrear las citas entre artículos facilita el seguimiento

del desarrollo de las investigaciones y la identificación de tendencias en diversas áreas del conocimiento.

### **1.6.2. Redes de Co-Autoría**

Las redes de coautoría se construyen a partir de las publicaciones científicas, donde cada autor se representa como un nodo en la red y las colaboraciones entre autores se representan como enlaces entre los nodos correspondientes. Estas redes proporcionan una representación visual y matemática de las relaciones de colaboración en la comunidad científica. Al analizar estas redes, es posible identificar la estructura de la colaboración, encontrar comunidades de autores que colaboran estrechamente, identificar autores influyentes y recomendar nuevas colaboraciones potenciales.

Además, las redes de coautoría se utilizan para estudiar tendencias en la producción científica, identificar áreas de investigación interconectadas y evaluar la difusión del conocimiento en diferentes campos. En resumen, estas redes proporcionan una herramienta poderosa para comprender la dinámica de la colaboración científica y para facilitar la identificación de oportunidades de colaboración entre investigadores.

### **1.6.3. Bases de Datos Orientadas a Grafos**

Las bases de datos orientadas a grafos son sistemas de gestión de bases de datos diseñados específicamente para almacenar, recuperar y gestionar datos que se pueden representar como grafos. En lugar de utilizar el modelo de datos relacional tradicional, las bases de datos orientadas a grafos utilizan estructuras de datos de grafo para representar y almacenar la información.

En un grafo, los datos se modelan como nodos (también conocidos como vértices) que están conectados por relaciones (también conocidas como aristas). Cada nodo y relación pueden contener propiedades que describen los atributos de los datos. Este enfoque es especialmente útil para modelar y consultar datos que tienen relaciones complejas y que requieren un análisis de redes.

Las bases de datos orientadas a grafos son ampliamente utilizadas en una variedad de aplicaciones, incluyendo redes sociales, recomendaciones personalizadas, análisis de redes, sistemas de recomendación, bioinformática, gestión del conocimiento y análisis de datos interconectados. Algunas bases de datos orientadas a grafos populares incluyen Neo4j, Amazon Neptune, y JanusGraph, entre otras.

#### **1.6.4. TF-IDF**

TF-IDF (Term Frequency-Inverse Document Frequency) es una técnica de procesamiento de lenguaje natural que se utiliza para evaluar la relevancia de un término en un documento o corpus de documentos. Esta técnica se utiliza comúnmente en la recuperación de información y la minería de texto.

La técnica TF-IDF se basa en dos conceptos principales: la frecuencia del término (TF) y la frecuencia inversa del documento (IDF). La frecuencia del término se refiere al número de veces que un término aparece en un documento, mientras que la frecuencia inversa del documento se refiere a la frecuencia con la que aparece un término en todo el corpus de documentos.

La fórmula para calcular el valor TF-IDF de un término es:

$$\text{TF-IDF} = \text{TF} \cdot \log\left(\frac{N}{\text{DF}}\right)$$

Donde TF es la frecuencia del término en el documento, N es el número total de documentos en el corpus y DF es el número de documentos que contienen el término.

El valor TF-IDF de un término es alto si aparece con frecuencia en un documento específico, pero rara vez aparece en otros documentos del corpus. Esto indica que el término es importante y relevante para el documento en cuestión.

La técnica TF-IDF se utiliza comúnmente en la recuperación de información para clasificar y ordenar documentos según su relevancia para una consulta de búsqueda específica. También se utiliza en la minería de texto para identificar patrones y tendencias en grandes conjuntos de datos de texto.

#### **1.6.5. Cosine Similarity**

La similitud de coseno mide cuán similares son dos vectores en un espacio de producto interno, evaluando el ángulo entre ellos. Calcula el coseno de este ángulo para determinar la dirección relativa de los vectores. Esta métrica se utiliza frecuentemente en el análisis de texto para comparar documentos según la frecuencia de palabras o frases [10].

#### **1.6.6. SCRUM**

Al iniciar un proyecto, es crucial que el equipo involucrado esté al tanto de sus roles, actividades y los plazos para la entrega de resultados. Scrum es un marco de trabajo que

facilita este proceso, permitiendo además acelerar la entrega de valor al cliente a través de iteraciones cortas [7].

En 2001, un grupo de desarrolladores en Salt Lake City creó el Manifiesto Ágil, que se resume en los siguientes cuatro valores:

1. Valorar a los individuos y su interacción por encima de los procesos y herramientas.
2. Valorar el software que funciona por encima de la documentación exhaustiva.
3. Valorar la colaboración con el cliente por encima de la negociación contractual.
4. Valorar la respuesta al cambio por encima del seguimiento de un plan.

El marco de trabajo SCRUM, dentro del contexto del Manifiesto Ágil, se fundamenta en aspectos como la flexibilidad, el factor humano, la colaboración y el desarrollo iterativo, con el objetivo principal de asegurar buenos resultados [7]. Scrum posee herramientas que permiten resolver problemas y mejorar la administración del proyecto, los cuales son:

- **Product Backlog:** Lista dinámica de características, requisitos, arreglos y mejoras que se deben completar.
- **Sprint Backlog:** Lista de elementos que el equipo de desarrollo debe completar durante el tiempo definido del sprint [8].
- **Incremento:** Entregable que surge al final del sprint.

Scrum define tres roles principales: Product Owner, Scrum Master y Team. Estos roles son esenciales en cada equipo, que se caracterizan por ser auto-organizados y multifuncionales [11].

- **Product Owner:** Encargado de maximizar el valor del trabajo y el perfil que habla con el cliente.
- **Scrum Master:** Responsable de que las técnicas y eventos de scrum sean aplicadas.
- **Team:** Es un grupo de profesionales que trabajan juntos para entregar un incremento de producto terminado al final de cada sprint. El equipo de desarrollo es auto-organizado y multidisciplinario, lo que significa que posee todas las habilidades necesarias para crear el producto sin depender de personas externas al equipo.

### 1.6.7. Arquitectura Hexagonal

La arquitectura Hexagonal o también conocida como la arquitectura de puertos y adaptadores, es un patrón de arquitectura de software que nos ayuda a tener un código mantenible, escalable y testable. Su característica principal es separar la lógica del dominio y que la misma no se acople a nada externo como frameworks, bases de datos, etc.

La arquitectura hexagonal se divide de 3 capas:

1. **Infraestructura:** También conocida como puertos es la capa que actúa como el punto de entrada hacia la aplicación. La misma que permite la comunicación desde y hacia el exterior.
2. **Aplicación:** También conocida como adaptadores, es el puente entre la capa de dominio y la capa de infraestructura. Esta capa sirve para transformar la comunicación entre actores externos (Capa de Infraestructura) y la lógica de la aplicación (Capa de Dominio), consiguiendo así que ambas capas sean independientes.
3. **Dominio:** Esta capa es el core de la aplicación, la cual va a contener toda la información y lógica del negocio.

Estas capas vienen acompañadas de una regla de dependencia que va desde afuera hacia adentro, teniendo en cuenta que la capa más interna es la del dominio y la más externa es la infraestructura. La regla de dependencia es muy simple.

## 1.6.8. Herramientas Utilizadas

<b>Nombre</b>	<b>Descripción</b>	<b>Versión</b>
Pycharm	IDE de desarrollo de Python para web y ciencia de datos [12].	2024.1.4
Django Rest	Conjunto de herramientas para el desarrollo de APIs [13].	3.15.2
Visual Studio Code	Editor de código [14].	1.91
Neomodel	Object Graph Mapper (OGM) para la base de datos Neo4j [15].	5.3.0
Angular	Plataforma para crear aplicaciones de escritorio web y móviles [16].	16
Swagger	Especificación abierta para definir las API REST [17].	0.27.2
Docker	Tecnología en contenedores que permite crear y usar contenedores Linux [18].	23.0.3
Power Designer	Herramienta de modelado de datos y diseño de bases de datos [19].	16.6

Tabla 1.1: Herramientas utilizadas

## Capítulo 2

# METODOLOGÍA

Para el desarrollo del proyecto se hizo uso de SCRUM [7] como marco de trabajo a través de 4 sprints [8] (Tomando en cuenta el Sprint 0). Cada subsección representa un Sprint en concreto.

### 2.1. Sprint 0

En este sprint denominado el "*Sprint 0*", se definieron aspectos fundamentales para el éxito del proyecto, con el objetivo de preparar al equipo y el entorno para los sprints de desarrollo regulares.

- Se realizó un análisis exhaustivo del stack tecnológico que se usará para el desarrollo, tanto del lado del frontend como del backend, abordando cada uno de los dos backends (ResNet Y ResearchDecide) y los cuatro componentes diferentes del sistema, los cuales son:
  - Automatización del proceso de extracción, depuración y almacenamiento de datos de Scopus mediante la API de Elsevier.
  - Desarrollo de módulo que implemente funciones del tipo red social para los usuarios del sistema.
  - Implementación de estrategia(s) de mediación o facilitación del consenso (sistema como mediador para el grupo de investigadores).
  - Desarrollo del módulo de Analítica: Dashboards y estadísticas para la toma de decisiones.

El propósito de este análisis es garantizar que tanto el frontend como el backend utilicen las tecnologías más apropiadas para satisfacer los requisitos técnicos y de negocio; y así, asegurar, y para así asegurar una integración efectiva y eficiente entre todas las partes del sistema.

En consecuencia, se tomaron las siguientes decisiones:

1. Para la integración del sistema en el lado del frontend se optó por utilizar Angular como framework de desarrollo, debido a que se puede reutilizar componentes de el sistema ResNet [2].
2. Para el lado del backend se propuso utilizar Django como framework. Con Django, se espera aprovechar su robustez, seguridad y su capacidad para desarrollar rápidamente aplicaciones web escalables y de alto rendimiento. Además, Django ofrece una amplia gama de características integradas, como autenticación de usuarios, administración de bases de datos, y un ORM<sup>1</sup> potente, lo que lo hace una elección sólida para el desarrollo del backend del sistema.
3. Para mantener las mismas dependencias de desarrollo y simplificar la gestión del entorno, se utilizará Docker [18]. Esto permite encapsular aplicaciones y sus dependencias en contenedores, asegurando consistencia entre entornos.
4. Además, Docker Compose se empleará para orquestar el servidor backend y la base de datos, facilitando la configuración y gestión del entorno, y simplificando el despliegue del sistema en diferentes entornos.
5. La elección de Neo4j se basa en la decisión de no alterar el modelo de datos actual. Dado que Neo4j está especialmente diseñado para modelos de datos conectados y relacionales, su adopción evita la necesidad de reestructurar el modelo existente, ahorrando tiempo y recursos mientras garantiza la continuidad del desarrollo del sistema.
6. El control de versiones se hará mediante Git y para alojar dichas versiones se utilizará GitHub. Mediante una organización en donde se alojará el código fuente de los 4 componentes.
7. Para la gestión del proyecto, se utilizó SCRUM en conjunto con Azure Boards [20]. SCRUM proporciona un marco ágil para la planificación y ejecución de pro-

---

<sup>1</sup>ORM (Object-Relational Mapping) es una herramienta que facilita la comunicación y traducción de datos entre bases de datos relacionales y objetos en programación orientada a objetos.



yectos, mientras que Azure Boards [20] ofrece herramientas integradas para facilitar todas las etapas del ciclo de vida del desarrollo de software. Esta combinación garantiza una gestión eficiente y colaborativa del proyecto, mejorando la transparencia y la entrega oportuna de resultados. En la Figura 2.1 se muestra el proyecto en Azure Boards.

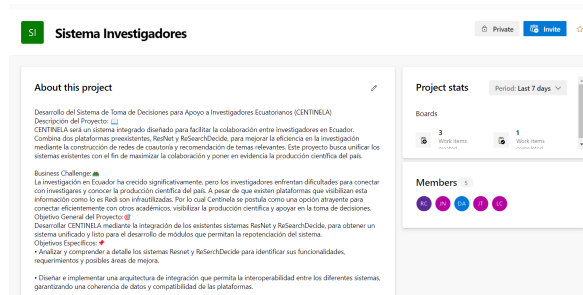


Figura 2.1: Proyecto en Azure Boards

8. Se optó por Visual Studio Code como editor de código para el frontend y Py-Charm como IDE de desarrollo especializado para el backend. Esta elección proporcionó un entorno flexible y eficaz para la colaboración en el proyecto.
- Se establecieron horarios para las reuniones diarias, las cuales se llevarán a cabo los días lunes, martes, miércoles y jueves en el horario de 14:00 a 16:00 en un espacio proporcionado por la directora del proyecto, Lorena Recalde, en el edificio de Sistemas-EPN. Durante este intervalo, se realizaron los daily meetings y se llevó a cabo el trabajo en equipo.
  - Se decidió adoptar el inglés como el idioma oficial para la codificación, los commits y toda la documentación técnica dentro de nuestro equipo. Esta medida busca estandarizar nuestras prácticas con los estándares internacionales, facilitar la colaboración con otros equipos y mejorar la accesibilidad a recursos técnicos de alta calidad.
  - Se decidió utilizar Arquitectura Hexagonal [21] como patrón de arquitectura, definiendo así una estructura de carpetas y respetando los principios de esta arquitectura. Cabe destacar que no se ha implementado arquitectura hexagonal en su totalidad, ya que se optó por una adaptación de la misma, obteniendo así una arquitectura de dos capas.
    - **Backend:** Para el backend, en cuanto a la estructura de carpetas, se definió que por cada módulo en general se tendrá las tres capas de arquitectura como son infraestructura, aplicación y dominio como se muestra en la Figura 2.2.

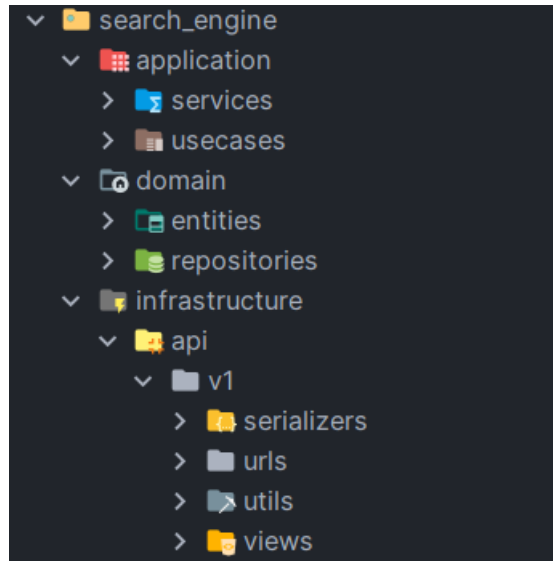


Figura 2.2: Arquitectura Hexagonal aplicada en el modulo del motor de búsqueda

- **Frontend:** Basándonos en el concepto de puertos y adaptadores, en el frontend también hemos optado por implementar la Arquitectura Hexagonal [21] como patrón de arquitectura, con algunas adaptaciones específicas para este contexto. Dado que el frontend se enfoca en la parte gráfica, el adaptador se denominará "presentación", donde se gestionará toda la interfaz gráfica. Esta capa de presentación será el punto de entrada hacia el núcleo (core) de la aplicación frontend, facilitando así una separación clara entre la lógica de negocio y la interfaz de usuario, y promoviendo una estructura modular y fácilmente mantenible (Figura 2.3.)

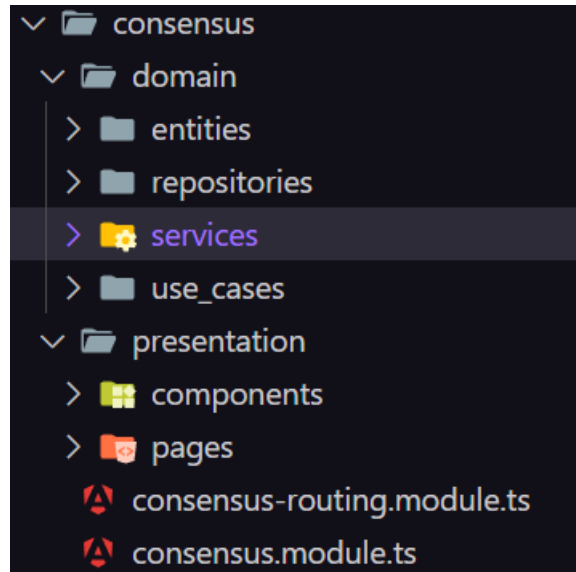


Figura 2.3: Arquitectura hexagonal aplicada al modulo de consenso

## 2.2. Sprint 1

### 2.2.1. Introducción

Durante este sprint, nuestro objetivo principal es crear diseños detallados que integren las funcionalidades de las aplicaciones Resnet (ver Sección 1.1) y Research Decide en una plataforma unificada. Además, nos centraremos en identificar y proponer mejoras significativas para esta aplicación combinada. Utilizaremos Figma para desarrollar los mockups, aplicando los patrones de diseño más adecuados para asegurar una experiencia óptima para el usuario final. Es importante destacar que esta fase será iterativa, permitiendo ajustes y cambios continuos durante todo el desarrollo, por lo que la propuesta final no será definitiva.

### 2.2.2. Objetivos del Sprint

- Diseñar el flujo de navegación de la aplicación combinada.
- Crear mockups de las interfaces de usuario de la aplicación.
- Desarrollar la estructura principal del motor de búsqueda.
- Implementar el enrutamiento de la aplicación.
- Implementar la página de inicio de la aplicación.

### **2.2.3. Planificación**

Para este sprint, hemos seleccionado las historias de usuario con su respectivo código enfocadas principalmente en la integración gráfica de las plataformas mencionadas en la Sección 1.1, que se muestran en la Tabla 2.1. Se debe tomar en cuenta que esta parte del desarrollo se enfoca en los usuarios que no necesariamente están registrados en la plataforma, por lo que se ha priorizado la visualización de información relevante para estos usuarios.

Identificador	Historia de Usuario	Tareas
HU-SE-01	Como usuario no registrado deseo poder encontrar artículos relevantes dado un tema de investigación para poder acceder rápidamente a información útil y actualizada que apoye mi estudio o trabajo.	<ul style="list-style-type: none"> <li>■ Componente para listar artículos</li> <li>■ Paginación para extraer resultados limitados</li> <li>■ Componente para filtrar la información por años</li> <li>■ Enrutamiento para redirigir hacia la página del artículo</li> <li>■ Página para visualizar información general del artículo</li> </ul>
HU-SE-02	Como usuario no registrado deseo poder ver los investigadores que tengan colaboraciones en artículos con afiliaciones ecuatorianas para mantenerme informado sobre sus investigaciones y campos de estudio.	<ul style="list-style-type: none"> <li>■ Interfaz para poder seleccionar distintos tipos de búsqueda</li> <li>■ Componente para visualizar lista de autores que cumplan con los criterios de búsqueda</li> <li>■ Página para visualizar información general del autor</li> <li>■ Componente de paginación para extraer resultados limitados</li> <li>■ Enrutamiento para redirigir hacia la página del autor</li> </ul>

Tabla 2.1: Historias de Usuario del sprint 1

Todas las historias de usuario se han dividido en tareas más pequeñas y manejables. También en las Figuras 2.4 y 2.22 se muestran los criterios de aceptación de cada historia de usuario, que se utilizarán para verificar si la historia de usuario se ha completado correctamente.

**Visualización de Investigadores:**

1. El usuario debe poder acceder a una lista de investigadores que han colaborado en artículos con afiliaciones ecuatorianas.
2. La lista debe incluir al menos el nombre del investigador, su afiliación principal.

**Información de Colaboraciones:**

1. Cada investigador en la lista debe mostrar claramente el número de colaboraciones en artículos con afiliaciones ecuatorianas.
2. Debe ser posible ver los detalles de las colaboraciones, incluyendo el título del artículo y la fecha de publicación.

**Filtros y Búsqueda:**

1. El usuario debe poder filtrar la lista de investigadores por área de estudio, institución o universidad, y año de publicación.
2. El usuario tener la opción de buscar investigadores por nombre o por palabras clave relacionadas con su campo de estudio.

**Acceso a Información de Artículos:**

1. Debo poder hacer clic en un investigador para ver una lista detallada de los artículos en los que ha colaborado con afiliaciones ecuatorianas.
2. Cada artículo debe mostrar información relevante como resumen, coautores, y enlace a la publicación completa si está disponible.

Figura 2.4: Criterios de aceptación de la historia de usuario HU-SE-01

**Resultados de Búsqueda:**

1. Al ingresar un tema de investigación, debo recibir una lista de artículos relevantes.
2. Los resultados deben ser presentados de manera clara, mostrando al menos el título del artículo, autores principales y fecha de publicación

**Filtros y Ordenación:**

1. Debo poder aplicar filtros para refinar los resultados de búsqueda por criterios como fecha de publicación, tipo de artículo (revista, conferencia, etc.), y relevancia.
2. Debo poder ordenar los resultados por relevancia, fecha de publicación, o número de citas.

**Acceso a Detalles del Artículo:**

1. Al hacer clic en un resultado de búsqueda, debo poder ver detalles completos del artículo, incluyendo el resumen completo, la lista completa de autores, afiliaciones, y enlaces a la publicación original o al texto completo si está disponible.

Figura 2.5: Criterios de aceptación de la historia de usuario HU-SE-02

Estas historias de usuario se han priorizado en función de su importancia y complejidad, y se han asignado a los miembros del equipo de desarrollo. La Figura 2.6 muestra las tareas asignadas a cada miembro del equipo.

Title	State	Assigned To	F
✓  Como usuario no registrado deseo poder encontrar artículo...	● New	:	
✓  Componente para listar Articulos	● To Do	JHON FERNAN...	
✓  Paginación para extraer resultados limitados	● To Do	JHON FERNAN...	
✓  Componente para filtrar la información por años	● To Do	JOFFRE ALEXA...	
✓  Enrutamiento para redirigir hacia la página del articulo	● To Do	JOFFRE ALEXA...	
✓  Pagina para visualizar la información general de un articu...	● To Do	JHON FERNAN...	
✓  Como usuario no registrado deseo poder ver los investigad...	● Done		
✓  Desarrollar la interfaz para poder seleccionar distintos tip...	● To Do	JHON FERNAN...	
✓  Componente para poder visualizar la lista de autores que...	● To Do	:	JOFFRE ALEXA...
✓  Pagina para el perfil de Investigador	● To Do	DANNY RUBEN...	
✓  Restricción para usuarios no registrados	● To Do	DANNY RUBEN...	
✓  Componente de paginación para extraer resultados limit...	● To Do	JHON FERNAN...	
✓  Enrutamiento para dirigirse a la pagina de un autor selec...	● To Do	JOFFRE ALEXA...	

Figura 2.6: Planificación de tareas del Sprint 1

## 2.2.4. Implementación

Para la implementación de las historias de usuario, se ha utilizado la herramienta de diseño Figma, que permite crear prototipos de alta y baja fidelidad. A continuación, se presentan los mockups de las interfaces de usuario de la aplicación que se han desarrollado durante este sprint. Cabe destacar que se siguió un proceso iterativo, por lo que los mockups presentados no son definitivos y pueden sufrir cambios en futuras iteraciones. Además el diseño se baso en el patrón de diseño Mobile First [22], que consiste en diseñar primero la versión móvil de la aplicación y luego adaptarla a dispositivos de mayor tamaño.

### Página de inicio

La página de inicio de la aplicación es la primera pantalla que verá el usuario al ingresar a la plataforma. En esta pantalla, se mostrarán las opciones de búsqueda y un conjunto de

datos sobre la información que tiene Centinela. Así como también una barra de navegación que permitirá al usuario acceder a otras secciones de la aplicación. La Figura 2.7 muestra el diseño de la página de inicio.



Figura 2.7: Mockup de la página de inicio

Como se mencionó en la Sección 2.1 se utilizará una adaptación de arquitectura hexagonal para desarrollar la estructura de la aplicación, dando como resultado la estructura de carpetas que se muestra en la Figura 2.8. Puesto que la página de inicio es la primera pantalla que verá el usuario al ingresar a la plataforma, se ha decidido crear una carpeta específica para esta sección. También debido a que Angular es un framework que se maneja con componentes, se ha creado un componente para la página de inicio, el cual se encuentra en la carpeta de pages. A su vez se ha creado el archivo de tipo module para la página de inicio, el mismo nos permitirá importar los componentes necesarios para la página de inicio, sin tener que importarlos en el archivo principal de la aplicación. Este enfoque permitirá que el modulo de la página de inicio sea independiente del resto de la aplicación,



que junto con la arquitectura hexagonal facilitara la escalabilidad y mantenimiento de la aplicación.

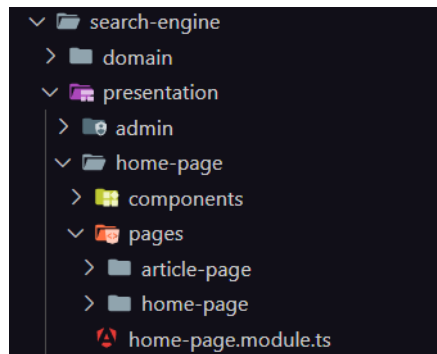


Figura 2.8: Estructura de carpetas de la página de inicio

El resultado de la implementación de la página de inicio se muestra en la Figura 2.9, en la que se muestra la barra de navegación, el formulario de búsqueda y la información que tiene Centinela. El formulario de búsqueda cuenta con 3 opciones, la primera es para buscar por palabra clave a un autor, la segunda es para buscar autores relevantes dado un tópico de investigación y la tercera es para buscar artículos relevantes dado un tópico. Debajo del motor de búsqueda, se presenta un resumen que indica la cantidad de artículos, autores y tópicos registrados en Centinela. Al final, se incluye un mapa que muestra la cantidad de artículos publicados en cada provincia del Ecuador.



Figura 2.9: Página de inicio

## Página de resultados

La página de resultados es la pantalla que se mostrará una vez que el usuario haya realizado una búsqueda en la aplicación. En esta pantalla, se mostrarán los resultados de la búsqueda, que incluirán una lista de autores y artículos relevantes, dependiendo del tipo de búsqueda que haya realizado el usuario. La Figura 2.10 muestra el diseño de la página de resultados. Cabe destacar que esta página va a ser dinámica, es decir, se modificará en función de los resultados de la búsqueda. También contiene el componente de paginación, que permitirá al usuario navegar entre los resultados de la búsqueda. Este último está disponible para el tipo de búsqueda de autores por palabra clave y para el tipo de búsqueda de artículos relevantes por tópico, ya que estos tipos de búsqueda pueden devolver un gran número de resultados. Y al ser una aplicación web, se ha optado por mostrar 10 resultados por página para garantizar una buena experiencia de usuario y reducir tiempos de carga.



Figura 2.10: Mockup de la página de resultados

Como se mencionó en las Figuras 2.4 y 2.22 los criterios de aceptación de las historias de usuario HU-SE-01 y HU-SE-02 cada resultado tendrá que ser redirigido a una pantalla en donde se visualizará la información detallada del autor o del artículo. Para este caso en específico se muestra el diseño para la página que tendrá la información detallada de un autor.

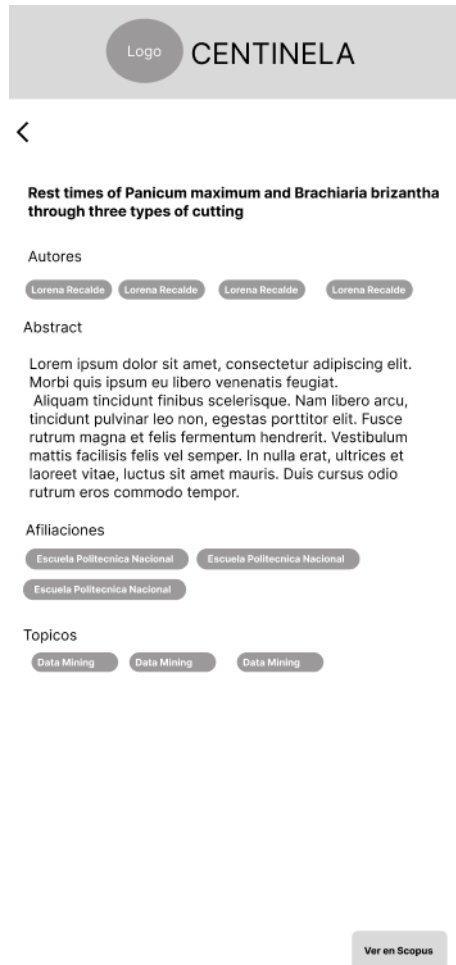


Figura 2.11: Mockup de la página de detalle del Artículo

Para el enrutamiento de la aplicación se ha utilizado el módulo de enrutamiento de Angular, el cual nos permite definir las rutas de la aplicación y asociarlas con los componentes correspondientes. En la Figura 2.12 se muestra el enrutamiento hacia el componente que contendrá la información general de un artículo, el mismo que como parámetro recibirá el id del Artículo que se desea visualizar.

```

1 children: [
2   {
3     path: '', // Esta ruta coincide con la página principal
4     component: SearchResultComponent, // El componente app-author-retrieve se mostrará solo en la página principal
5   },
6   {
7     path: 'article/:scopusId',
8     component: ArticlePageComponent
9   },
10 ]

```

Figura 2.12: Routing de la página de detalle de Artículo

Resultado de la implementación de la página de información general de un Artículo se

muestra en la Figura 2.13. En la que se muestra la información general de un Artículo, como el título, resumen, autores, afiliaciones, palabras clave y el año de publicación, así como también un botón que permitirá al usuario acceder al Artículo completo en la página de Scopus.



Figura 2.13: Página de detalle de Artículo

## 2.2.5. Revisión y Retrospectiva

Durante este sprint, hemos logrado completar todas las historias de usuario planificadas, lo que nos ha permitido avanzar significativamente en el desarrollo de la aplicación. Hemos creado los mockups de las interfaces de usuario de la aplicación, desarrollado la estructura principal del motor de búsqueda, implementado el enrutamiento de la aplicación y la página de inicio. Además, hemos implementado la página de resultados y la página de información general de un artículo. En general, el equipo ha trabajado de manera eficiente y colaborativa, lo que ha permitido cumplir con los objetivos del sprint en el tiempo previsto. Sin embargo, hemos identificado algunas áreas de mejora que podrían ayudarnos a optimizar nuestro trabajo en futuros sprints.

- **Comunicación:** Aunque hemos mantenido una comunicación constante a través de reuniones diarias y canales de mensajería, es importante mejorar la comunicación entre los miembros del equipo para garantizar que todos estén al tanto de los avances del proyecto.
- **Planificación:** Aunque hemos logrado completar todas las historias de usuario planificadas, es importante revisar y ajustar nuestra planificación para futuros sprints,

teniendo en cuenta la complejidad y el tiempo requerido para cada tarea.

- **Colaboración:** Hemos trabajado de manera colaborativa y eficiente durante este sprint, lo que ha permitido cumplir con los objetivos del proyecto. Es importante seguir fomentando la colaboración entre los miembros del equipo para garantizar el éxito del proyecto.
- **Retroalimentación:** Es importante recopilar y analizar la retroalimentación de los usuarios para identificar áreas de mejora y realizar ajustes en futuras iteraciones del proyecto.

Cabe destacar que las tareas referentes a la integración de la aplicación con la API de Scopus, así como la implementación de funcionalidades del lado del servidor, se han dejado para futuros sprints, ya que requieren un mayor tiempo de desarrollo y dependen de la finalización de otras tareas.

## **2.3. Sprint 2**

### **2.3.1. Introducción**

Durante el segundo sprint se continuó con el desarrollo de la integración de las aplicaciones mencionadas en el sprint anterior. Para este sprint en específico se trabajó en los detalles sobre los autores, así como su red de coautoría, artículos publicados y tópicos de interés.

### **2.3.2. Objetivos**

- Implementar las ventanas para cada uno de los detalles de los autores.
- Integrar la fuerza de colaboración entre autores.
- Desarrollar la funcionalidad de artículos publicados por autor.
- Integrar enrutamiento para las diferentes vistas de los autores.

### **2.3.3. Planificación**

Para este sprint se seleccionó un total de 2 historias de usuario, divididas en un total de 12 tareas a realizar, las cuales fueron asignadas a los integrantes del equipo de desarrollo.

En la Tabla 2.2 se detallan las historias de usuario seleccionadas para este sprint.

Identificador	Historia de Usuario	Tareas
HU-SE-03	Como usuario no registrado, deseo poder ver la red de co-autoría de un autor para visualizar un grafo con los autores con los que ha colaborado, así como la fuerza de esas colaboraciones.	<ul style="list-style-type: none"> <li>■ Diseñar e implementar la interfaz y el grafo interactivo</li> <li>■ Definir y aplicar un esquema visual para la fuerza de la colaboración</li> <li>■ Implementar funcionalidades interactivas y filtros</li> <li>■ Agregar un panel de detalles con información adicional sobre el autor</li> </ul>
HU-SE-04	Como usuario no registrado quiero poder ver los artículos de un investigador para conocer su trabajo y las publicaciones en las que ha contribuido.	<ul style="list-style-type: none"> <li>■ Diseñar la interfaz y enrutamiento para la lista de artículos</li> <li>■ Implementar la tabla de artículos y la funcionalidad de redirección</li> </ul>

Tabla 2.2: Historias de Usuario del Sprint 2

Al igual que en el sprint anterior, cada historia de usuario se dividió en tareas más pequeñas, las cuales se asignaron a los integrantes del equipo de desarrollo. Así también se definieron los criterios de aceptación que se muestran en la Figura 2.14 y en la Figura 2.15. para cada una con el fin de asegurar la calidad del desarrollo y que cada miembro del equipo tuviera claro lo que se esperaba de la tarea asignada.

**Visualización del Grafo de Coautoría:**

- El usuario no registrado debe poder acceder a una página o sección donde se muestre un grafo interactivo de la red de coautoría de un autor específico.
- El grafo debe mostrar nodos que representen a los autores y enlaces que representen las colaboraciones entre ellos

**Información del Nodo:**

- Al pasar el cursor sobre un nodo (autor) en el grafo, debe aparecer información básica sobre el autor, como su nombre completo y afiliación.
- Al hacer clic en un nodo, se debe mostrar un panel con detalles adicionales del autor, como el número de publicaciones y las áreas de investigación principales.

**Fuerza de la Colaboración:**

- Los enlaces entre nodos deben variar en grosor o color para indicar la fuerza de la colaboración, basada en el número de publicaciones conjuntas o la duración de la colaboración.
- Debe haber una leyenda que explique el significado de los diferentes grosores o colores de los enlaces.

**Interactividad:**

- El grafo debe ser interactivo, permitiendo al usuario hacer zoom y desplazar la vista para explorar diferentes partes de la red de coautoría.
- El usuario debe poder filtrar las colaboraciones mostradas por año o relevancia.

Figura 2.14: Criterios de Aceptación HU-SE-03

**Visualización de Artículos:**

- El usuario debe poder acceder a una página o sección donde se muestran los artículos de un investigador específico.
- La lista de artículos debe incluir el título del artículo y la fecha de publicación.

**Detalle del Artículo:**

- Al hacer clic en un artículo de la lista, el usuario debe ser redirigido a una página de detalles que muestre información adicional del artículo, como el resumen, los coautores, y enlaces al texto completo si están disponibles.

Figura 2.15: Criterios de Aceptación HU-SE-04

A continuación, en la Figura 2.16, se muestran las tareas de cada historia de usuario. Es así que en Azure Boards se crearon los *work items* para cada tarea y se asignaron a los miembros del equipo.

Title	State	Assigned To
<input checked="" type="checkbox"/> Como usuario no registrado, deseo poder ver la red de coautoría de un autor para visualizar un grafo con los autores con los que ha colaborado, así como la fuerza de esas colaboraciones	New	
<input checked="" type="checkbox"/> Diseñar e implementar la interfaz y el grafo interactivo	To Do	JOFFRE ALEXA...
<input checked="" type="checkbox"/> Definir y aplicar un esquema visual para la fuerza de la colaboración	To Do	JOFFRE ALEXA...
<input checked="" type="checkbox"/> Implementar funcionalidades interactivas y filtros	To Do	JHON FERNAN...
<input checked="" type="checkbox"/> Agregar un panel de detalles con información adicional sobre el autor	To Do	JHON FERNAN...
<input checked="" type="checkbox"/> Como usuario no registrado quiero poder ver los artículos de un investigador para conocer su trabajo y las publicaciones en las que ha contribuido.	New	
<input checked="" type="checkbox"/> Diseñar la interfaz y enrutamiento para la lista de artículos	To Do	JHON FERNAN...
<input checked="" type="checkbox"/> Implementar la tabla de artículos y la funcionalidad de redirección	To Do	JOFFRE ALEXA...

Figura 2.16: Planificación de tareas del sprint 2



## 2.3.4. Implementación

Al igual que en el sprint anterior, se utilizó Figma para el diseño de los mockups de las vistas que posteriormente se implementarán utilizando Angular. En la Figura 2.17 se muestra el diseño de la vista de los detalles de un autor. Así como los detalles que va tener su red de coautoría.

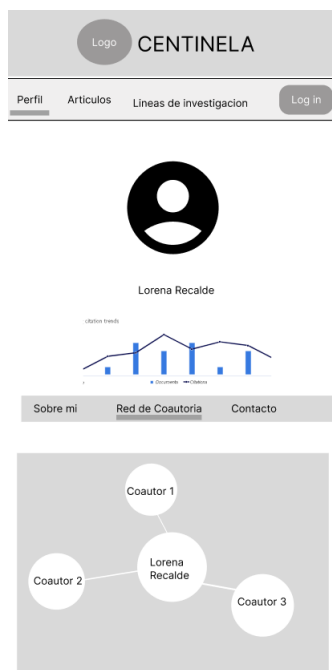


Figura 2.17: Mockup de los detalles de un autor

Como se muestra en la Figura 2.17, se diseñó la vista de los detalles de un autor, en la cual se muestra la red de coautoría del autor seleccionado. En la red de coautoría se muestra un grafo con los autores con los que ha colaborado, y la fuerza de colaboración con cada autor se refleja en el grosor de la línea que los une.

En la Figura 2.18 se muestra el diseño de la vista de los artículos publicados por un autor. Los mismos están representados en una tabla con la información de cada uno y contendrán la interactividad que permitirá al usuario redirigir al mockup de la Figura 2.11.



Figura 2.18: Mockup de los artículos publicados por un autor

Para implementar la red de coautoría se utilizó la librería D3.js, la cual permite la creación de gráficos interactivos en la web. Para esto primero debemos definir los nodos y las relaciones entre ellos, en este caso los autores y la fuerza de colaboración entre ellos. En la Figura 2.19 se muestra la interfaz en Typescript que servirá para la creación de la red de coautoría.

```

1
2 export interface Link {
3   source: number;
4   target: number;
5   collabStrength: number;
6 }
7
8 export interface Coauthors {
9   links: Link[]
10  nodes: AuthorNode[]
11  affiliations: { scopusId: number, name: string }[]
12 }
13 export interface AuthorNode {
14   scopus_id: number
15   initials: string
16   first_name: string
17   last_name: string
18   weight: number
19 }
20

```

Figura 2.19: Interfaz en Typescript para la creación de la red de coautoría

Debido a que el actual componente no se encargará de la creación de la red, sino que se encargará de la visualización de la misma, no se ahondará en la implementación de la red de coautoría, ya que se encuentra fuera del alcance de este documento. El resultado

final de la red de coautoría se muestra en la Figura 2.20.



Figura 2.20: Red de coautoría

Cabe mencionar que para armar la red de coautoría que se muestra en la Figura 2.20 se utilizó un conjunto de datos de prueba, ya que de momento la aplicación no cuenta con la información necesaria para armar la red de coautoría.

Ahora, en la Figura 2.21, se muestra la tabla de los artículos publicados por un autor. Esta tabla también contendrá la interactividad que permitirá al usuario redirigir al mockup de la Figura 2.11.



Figura 2.21: Tabla de artículos publicados por un autor

### 2.3.5. Revisión y Retrospectiva

Como resultado de este sprint se logró implementar las ventanas para cada uno de los detalles de los autores, la integración de la fuerza de colaboración entre autores, la funcionalidad de los artículos publicados por autor y el enrutamiento para las diferentes vistas de los autores. Se logró cumplir con los objetivos planteados para este sprint, sin embargo, se presentaron algunos problemas durante el desarrollo de las tareas asignadas. Uno de los problemas que se presentó fue la falta de información necesaria para la creación de la red de coautoría, ya que la aplicación no cuenta con la información necesaria para armar la red de coautoría. En todo caso, los servicios que se encargarán de construir la red de coautoría ya se encuentran implementados y listos para recibir la información necesaria. Además la implementación de la extracción de la información desde scopus se realizará en el siguiente Sprint.

## **2.4. Sprint 3**

### **2.4.1. Introducción**

A partir de este Sprint, se comenzará a implementar la integración con Scopus y toda lógica relacionada con el servidor.

Para este Sprint específico, se establecerá el ambiente de desarrollo con las herramientas necesarias para la implementación. Además, se rediseñará la base de datos orientada a grafos utilizando Neo4j, Django y Neomodel como ORM, basándonos en el diseño en Neo4j propuesto en Resnet. Se trabajará en la implementación de los modelos en Neomodel.

Finalmente, se pretende exponer el uso de Arquitectura Hexagonal para la implementación de la funcionalidad de búsqueda de autores. Cabe destacar que éste será el único sprint en el que se ahondará en la Arquitectura Hexagonal; los demás sprints no tendrán ese nivel de detalle.

### **2.4.2. Objetivos**

- Levantar el ambiente de desarrollo.
- Implementar y rediseñar los modelos de la base de datos en Neo4j a través de Neomodel.
- Desarrollar la funcionalidad de búsqueda de autores.

### **2.4.3. Planificación**

Para este sprint se ha seleccionado las historias de usuario que se muestran en la Tabla 2.3.

<b>Identificador</b>	<b>Historia de Usuario</b>	<b>Tareas</b>
HU-SE-02	Como usuario no registrado deseo poder ver los investigadores que tengan colaboraciones en artículos con afiliaciones ecuatorianas para mantenerme informado sobre sus investigaciones y campos de estudio.	<ul style="list-style-type: none"> <li>■ Crear servicios para el modelo de autor</li> <li>■ Diseñar el endpoint que maneje las solicitudes del autor</li> </ul>
HU-SE-05	Como desarrollador de software, quiero migrar nuestro modelo de base de datos actual a Neomodel, para que podamos aprovechar las capacidades del ORM Neomodel para simplificar la gestión de datos y mejorar la integración con nuestro proyecto Django.	<ul style="list-style-type: none"> <li>■ Definir los modelos, relaciones y propiedades de los nodos.</li> <li>■ Hacer las migraciones desde Django a Neo4j.</li> </ul>
HU-SE-06	Como desarrollador, quiero configurar un entorno de desarrollo replicable, para asegurar que podamos contar con un entorno completamente aislado y consistente que optimice el flujo de trabajo.	<ul style="list-style-type: none"> <li>■ Crear el contenedor de Docker para la aplicación de backend</li> <li>■ Crear un orquestador de contenedores</li> <li>■ Orquestar los servicios de la aplicación y la base de datos</li> </ul>

Tabla 2.3: Historias de Usuario del sprint 3

A continuación en las Figuras 2.22, 2.23 y 2.24 se muestran los criterios de aceptación para las historias de usuario seleccionadas.

#### Acceptance Criteria



- El sistema debe permitir la búsqueda de autores utilizando cualquier atributo relevante, como nombre, apellido, ID de Scopus, o cualquier otro identificador único o atributo significativo
- El sistema debe permitir la búsqueda de autores utilizando cualquier atributo relevante, como nombre, apellido, ID de Scopus, o cualquier otro identificador único o atributo significativo
- Si no se encuentran resultados que coincidan con los criterios de búsqueda, el sistema debe devolver una lista vacía
- Los resultados de la búsqueda deben ser devueltos con paginación para manejar grandes volúmenes de datos y mejorar la eficiencia

Figura 2.22: Criterios de aceptación HU-SE-02

#### Acceptance Criteria



- El modelo de datos actual debe ser migrado a Neomodel sin pérdida de información ni alteraciones en la estructura de los datos.
- Los modelos en Neomodel deben reflejar fielmente la estructura y las relaciones definidas en el modelo de datos actual.
- La integración entre Neomodel y el proyecto Django debe estar configurada correctamente.

Figura 2.23: Criterios de aceptación HU-SE-05

#### Acceptance Criteria



- El entorno de desarrollo debe estar configurado utilizando Docker y Docker Compose para asegurar la replicabilidad
- Debe existir un archivo `Dockerfile` para definir la imagen del contenedor y un archivo `docker-compose.yml` para la configuración de los servicios.
- Los contenedores deben estar configurados para comunicarse entre sí de manera segura y eficiente
- Toda información sensible debe ser guardada en variables de entorno

Figura 2.24: Criterios de aceptación HU-SE-06

### 2.4.4. Implementación

Para el desarrollo del Sprint 3 se empezó con la configuración del entorno de desarrollo, se utilizó Docker para la creación de contenedores y Docker Compose para la orquestación de los servicios. Para dockerizar la aplicación se creó un archivo Dockerfile en la raíz del proyecto, en el cual se especifican las instrucciones necesarias para la creación de la imagen de la aplicación, tal como se muestra en la Figura 2.25.

En la Figura 2.26 presenta el archivo de configuración de Docker Compose. El mismo se encarga de orquestar los servicios de la aplicación y la base de datos. También se asegura de persistir los datos de la base de datos en un volumen de Docker. De la misma manera, se especifican las variables de entorno necesarias para la configuración y el correcto funcionamiento de la aplicación.

Una vez configurados los archivos de Docker y Docker Compose, se procedió a levantar los contenedores con el comando `docker-compose up --build`. Este comando se encarga

```
Dockerfile

FROM python:3.11-bookworm

WORKDIR /app

COPY requirements.txt .

RUN pip install --no-cache-dir -r requirements.txt

COPY . .
EXPOSE 8001

CMD ["python", "manage.py", "runserver", "0.0.0.0:8001", "--noreload"]
```

Figura 2.25: Archivo Dockerfile

```
compose.yaml

services:
  web:
    build: .
    command: python manage.py runserver 0.0.0.0:8001
    volumes:
      - ./centinela_logs:/centinela_logs
      - ./app
    restart: always
    ports:
      - "8001:8001"
    environment:
      - NEO4J_PASSWORD=${NEO4J_PASSWORD}
      - MONGO_DB_NAME=${MONGO_DB_NAME}
      - MONGO_DB_USERNAME=${MONGO_DB_USERNAME}
      - MONGO_DB_PASSWORD=${MONGO_DB_PASSWORD}
      - MONGO_DB_HOST=${MONGO_DB_HOST}
      - MONGO_DB_PORT=${MONGO_DB_PORT}
    links:
      - "neo4j:neo4j"
      - "mongo:mongo"

  neo4j:
    container_name: neo4j
    image: neo4j:latest
    restart: always
    ports:
      - "7474:7474"
      - "7687:7687"
    environment:
      - NEO4J_AUTH=neo4j/${NEO4J_PASSWORD}
      - NEO4J_apoc_export_file_enabled=true
      - NEO4J_apoc_import_file_enabled=true
      - NEO4J_apoc_import_file_use__neo4j__config=true
      - NEO4J_PLUGINS=["apoc", "graph-data-science"]
    volumes:
      - ./neo4j_db/data:/data
      - ./neo4j_db/logs:/logs
      - ./neo4j_db/import:/var/lib/neo4j/import
      - ./neo4j_db/plugins:/plugins
```

Figura 2.26: Archivo Docker Compose



de construir y levantar los servicios de la aplicación y la base de datos. Para verificar que los contenedores se encuentran en ejecución se utiliza el comando `docker ps`. En la Figura 2.27 se muestra la salida del comando `docker ps`.

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NAMES
8be816bcd986	search_engine_backend-web	"python manage.py ru..."	5 days ago	Up About a minute	0.0.0.0:8001->8001/tcp	search_engine_back
end-web-1						
6737fb98afb3	neo4j:latest	"tini -g -- /startup..."	5 days ago	Up About a minute	0.0.0.0:7474->7474/tcp, 7473/tcp, 0.0.0.0:7687->7687/tcp	neo4j
2b74dd808f5c	mongo:latest	"docker-entrypoint.s..."	5 days ago	Up About a minute	0.0.0.0:27017->27017/tcp	mongo

Figura 2.27: Salida del comando `docker ps`

Con el entorno de desarrollo configurado, se procedió a la creación de las aplicaciones de Django. Django denomina aplicaciones a un conjunto de funcionalidades que se pueden reutilizar en diferentes proyectos. En este caso se crearon 3 aplicaciones: `dashboards`, `scopus_integration` y `search_engine`, mismas que se pueden observar en la Figura 2.28.

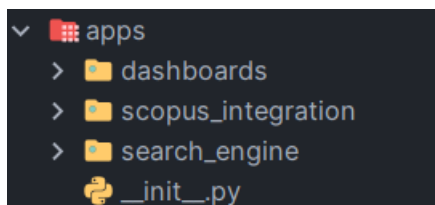


Figura 2.28: Aplicaciones de Django

Las aplicaciones tendrán la estructura de carpetas en base a la Arquitectura Hexagonal, como se mencionó en la Sección 2.1. Cada aplicación tendrá su propia capa de dominio, aplicación e infraestructura. Así como cada una tendrá una función específica en la aplicación, `dashboards` se encargará de la visualización de los datos, `scopus_integration` de la integración con Scopus y `search_engine` de la funcionalidad del motor de búsqueda.

Una vez finalizada la creación de las aplicaciones se procedió a hacer el rediseño de la base de datos previo a la implementación de los modelos en Neonmodel. Como se mencionó en la Sección 2.4.1, se tomó como base el diseño en Neo4j propuesto en Resnet. Únicamente se agregó nuevos atributos a los nodos y relaciones existentes. Específicamente, se agregaron atributos al nodo de `Autor` y a la relación `COAUTHORED`. En la Figura 2.29 se muestra el diseño de la base de datos.

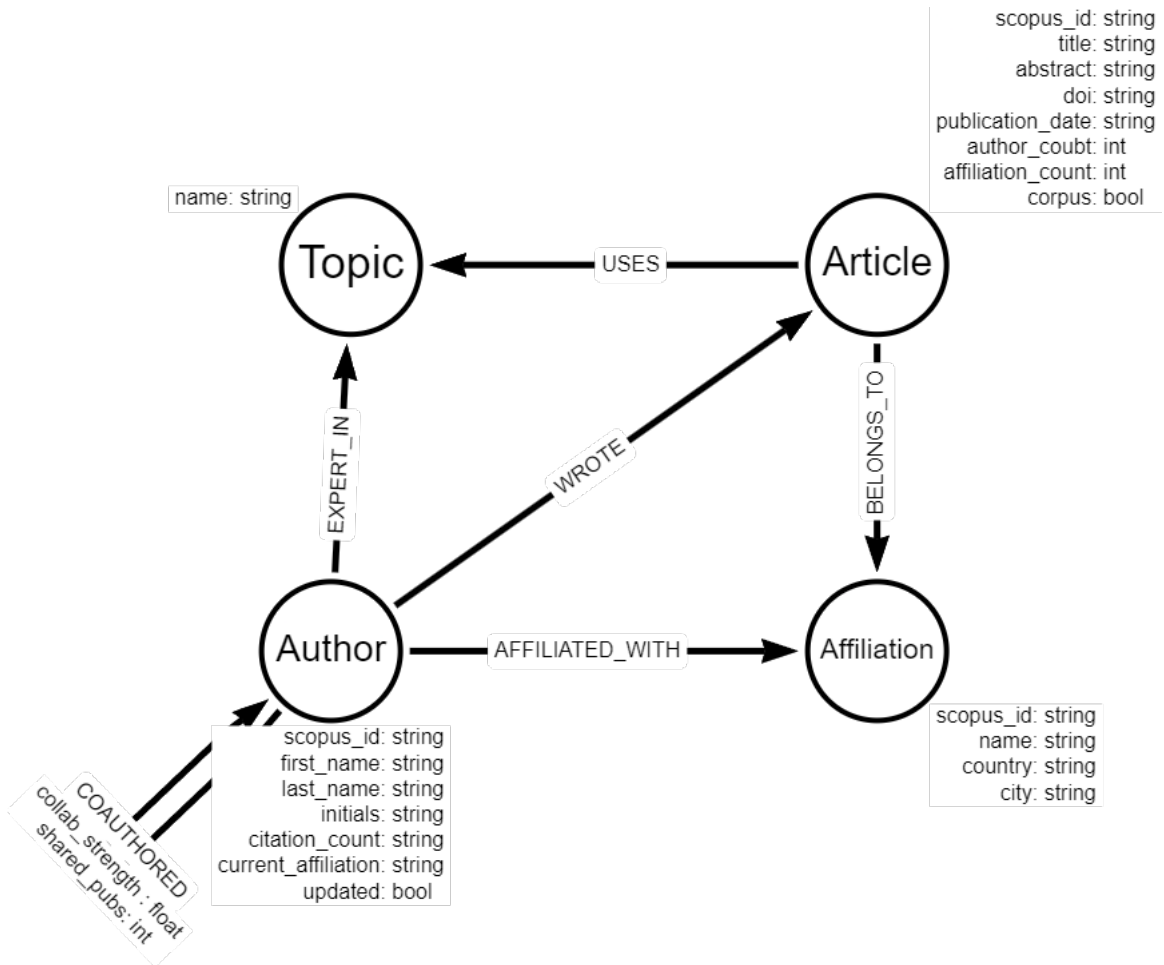


Figura 2.29: Modelo de la base de datos

Con estas actualizaciones en el modelo, se procedió a la implementación de los mismos en Neomodel. Como se mencionó en el Sprint 0, para el desarrollo de la aplicación se utilizará la Arquitectura Hexagonal. Bajo este enfoque se crearon los modelos de la base de datos en la capa de dominio. En la Figura 2.30 se muestra la capa de dominio de la aplicación que se encargará del motor de búsqueda.

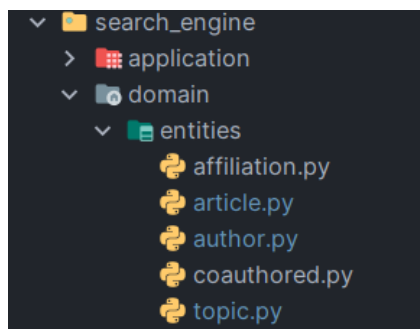
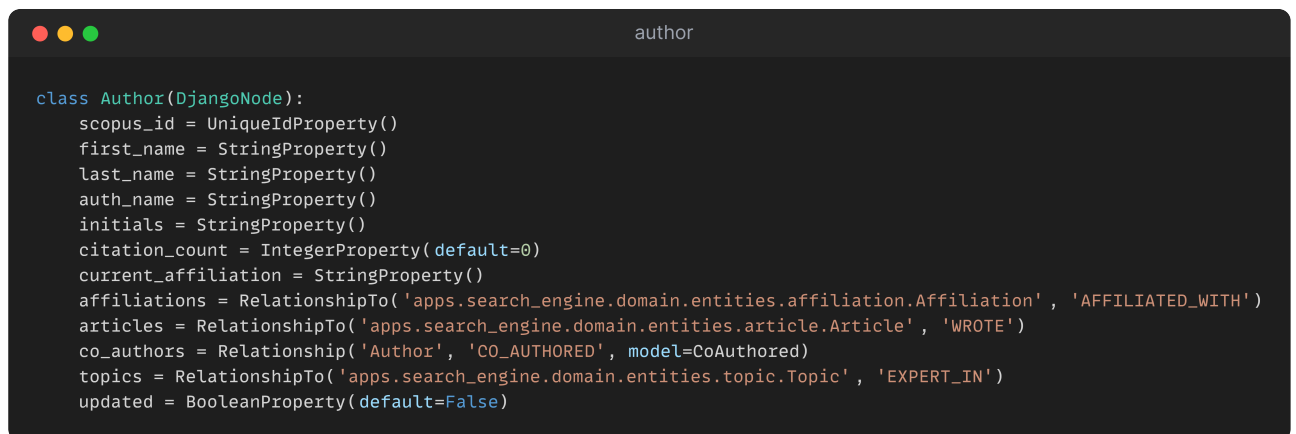


Figura 2.30: Capa de dominio de la aplicación

Basándonos en el diseño de la base de datos que se muestra en la Figura 2.29, se muestra el modelo *Author* implementado con Neomodel en la Figura 2.31. En términos generales, para asegurar la integridad de los datos se hace uso de las restricciones de unicidad de Neomodel, con el tipo de dato *UniqueIdProperty*. Esto será de gran utilidad para evitar la creación de nodos duplicados en la base de datos. Asimismo, se utilizan las propiedades de tipo *Relationship* y *RelationshipTo* para definir las relaciones entre los nodos.

The image shows a code editor window titled 'author' with a dark background. The code defines a Neomodel class for 'Author'. It includes properties for scopus\_id (Unique), first\_name, last\_name, auth\_name, initials, citation\_count (Integer), and current\_affiliation (String). It also defines relationships: 'affiliations' (RelationshipTo Affiliation), 'articles' (RelationshipTo Article), 'co\_authors' (Relationship to Author with CoAuthored model), and 'topics' (RelationshipTo Topic). An 'updated' Boolean property is also present.

```
class Author(DjangoNode):
    scopus_id = UniqueIdProperty()
    first_name = StringProperty()
    last_name = StringProperty()
    auth_name = StringProperty()
    initials = StringProperty()
    citation_count = IntegerProperty(default=0)
    current_affiliation = StringProperty()
    affiliations = RelationshipTo('apps.search_engine.domain.entities.affiliation.Affiliation', 'AFFILIATED_WITH')
    articles = RelationshipTo('apps.search_engine.domain.entities.article.Article', 'WROTE')
    co_authors = Relationship('Author', 'CO_AUTHORED', model=CoAuthored)
    topics = RelationshipTo('apps.search_engine.domain.entities.topic.Topic', 'EXPERT_IN')
    updated = BooleanProperty(default=False)
```

Figura 2.31: Modelo Author implementado con Neomodel

Finalmente, la migración de estos modelos a la base de datos se realiza con el comando `python manage.py install_labels`. Aunque este comando es opcional, es recomendable ejecutarlo para asegurar que los modelos se encuentren en la base de datos, o si en su defecto se requiere hacer una actualización de los modelos.

Siguiendo la Arquitectura Hexagonal, se procedió a implementar la funcionalidad de búsqueda de autores. Para ello, se creó el repositorio *AuthorRepository* en la capa de dominio, encargado de definir los métodos necesarios para la búsqueda de autores.

El uso del patrón de diseño *Repository* garantiza la separación entre la lógica de negocio y la lógica de persistencia. Posteriormente, en la capa de aplicación, se implementará la lógica de negocio correspondiente a cada una de las funciones definidas en el repositorio.

En la Figura 2.32 se muestra un extracto del repositorio *AuthorRepository* con algunos de los métodos que nos servirán para la búsqueda de autores.

```
author_repository

class AuthorRepository(ABC):
    @abstractmethod
    def find_authors_by_query(self, name: str, page_size=None, page=None) → List[object]:
        pass

    @abstractmethod
    def find_authors_by_affiliation_filter(self, filter_type: str, affiliations_ids: List[str],
                                         authors_ids: List[str]) → List[object]:
        pass

    @abstractmethod
    def find_community(self, authors_ids: List[str]):
        pass

    @abstractmethod
    def find_most_relevant_authors_by_topic(self, topic: str, authors_number: int):
        pass
```

Figura 2.32: Repositorio AuthorRepository

Una vez definidos los repositorios pasamos a la capa de aplicación, en la cual se implementa la lógica de negocio correspondiente a cada uno de los métodos definidos en el repositorio. La Figura 2.33 muestra un extracto del servicio *AuthService*, el mismo que hereda de la clase *AuthorRepository* con la implementación de la función *find\_authors\_by\_query*.

Dado que en este caso se está utilizando una base de datos orientada a grafos, se empleó el lenguaje de consulta Cypher para realizar la búsqueda de autores. Usar Cypher permite aprovechar las ventajas de las bases de datos orientadas a grafos, como la rapidez en la consulta de datos y la facilidad para encontrar relaciones entre los nodos.

Sin embargo, existe el riesgo de Cypher Injection, por lo que es fundamental tener cuidado al construir las consultas. Para todas las consultas, se utilizó el método *cypher\_query* de Neomodel, que permite ejecutar consultas Cypher directamente en la base de datos, enviando los parámetros, no en la cadena de texto, sino como argumentos de la función. Esto evita la inyección de Cypher en la base de datos, sanitizando las consultas y protegiéndolas de posibles inyecciones de Cypher.

Una vez con el servicio listo, se procedió a la implementación del caso de uso *FindAuthorsByQuery* en la capa de aplicación. En la Figura 2.34 se muestra el caso de uso que se encargará de llamar a los repositorios necesarios para la búsqueda de autores.

```
author_service

class AuthorService(AuthorRepository):
    def find_authors_by_query(self, name: str, page_size=1, page=1) → (List[object], int):
        custom_name = unicode(name).strip().lower()
        skip = (page - 1) * page_size
        query = """
            MATCH (au:Author)
            WHERE toLower(au.first_name) CONTAINS $custom_name OR
                toLower(au.last_name) CONTAINS $custom_name OR
                toLower(au.first_name) + " " + toLower(au.last_name) CONTAINS $custom_name OR
                toLower(au.last_name) + " " + toLower(au.first_name) CONTAINS $custom_name OR
                toLower(au.auth_name) CONTAINS $custom_name OR
                toLower(au.initials) CONTAINS $custom_name OR
                toLower(au.email) CONTAINS $custom_name OR
                au.scopus_id CONTAINS $custom_name
            RETURN count(au) as total
        """
        params = {'custom_name': custom_name}
        results, meta = db.cypher_query(query, params=params)

        total = results[0][0]
        query = """
            MATCH (au:Author)
            WHERE toLower(au.first_name) CONTAINS $custom_name OR
                toLower(au.last_name) CONTAINS $custom_name OR
                toLower(au.first_name) + " " + toLower(au.last_name) CONTAINS $custom_name OR
                toLower(au.last_name) + " " + toLower(au.first_name) CONTAINS $custom_name OR
                toLower(au.auth_name) CONTAINS $custom_name OR
                toLower(au.initials) CONTAINS $custom_name OR
                toLower(au.email) CONTAINS $custom_name OR
                au.scopus_id CONTAINS $custom_name
            RETURN au
            ORDER BY au.citation_count DESC
            SKIP $skip LIMIT $page_size
        """
        params = {'custom_name': custom_name, 'skip': skip, 'page_size': page_size}
        results, meta = db.cypher_query(query, params=params)
        authors = [Author.inflate(row[0]) for row in results]
        return authors, total
```

Figura 2.33: Servicio AuthorService

```
author_by_query_use_case

class AuthorByQueryUseCase:
    def __init__(self, author_repository: AuthorRepository):
        self.author_repository = author_repository

    def execute(self, name: str, page_size=10, page=1) → list[object]:
        return self.author_repository.find_authors_by_query(name, page_size=page_size, page=page)
```

Figura 2.34: Caso de uso FindAuthorsByQuery

A continuación, se procedió a la implementación de la capa de infraestructura. Esta capa se va a encargar de la comunicación con el exterior de la aplicación, en este caso con la API REST. Para ello se creó la vista *AuthorViews* que se encargará de recibir las peticiones HTTP y llamar a los casos de uso correspondientes. En la Figura 2.35 se muestra un extracto de la vista *AuthorViews* con la implementación del caso de uso *FindAuthorsByQuery*. Como se puede observar, la vista únicamente se encarga de recibir la petición HTTP y lla-

mar al caso de uso correspondiente. No maneja la lógica de negocio. Esto permite que la vista sea independiente de la lógica de negocio y de la lógica de persistencia. De esta manera, se garantiza la separación de responsabilidades y se facilita la reutilización de código, logrando que la aplicación sea más mantenible y escalable.

```
author_views

class AuthorViews(viewsets.ViewSet):
    # inject needed services
    author_service = AuthorService()
    affiliation_service = AffiliationService()

    @extend_schema(
        description="Find authors by query",
        responses=RetrieveAuthorSerializer(many=True),
        tags=['Authors'],
        parameters=[
            OpenApiParameter(name='query', type=str, required=True),
            OpenApiParameter(name='page_size', type=int, required=False),
            OpenApiParameter(name='page', type=int, required=False)
        ]
    )
    @action(detail=False, methods=['get'], url_path='find_by_query')
    def find_by_query(self, request, *args, **kwargs):
        try:
            query = request.query_params.get('query', '')
            page_size = int(request.query_params.get('page_size', 10))
            page = int(request.query_params.get('page', 1))
            author_by_query_use_case = AuthorByQueryUseCase(author_repository=self.author_service)
            authors, total = author_by_query_use_case.execute(name=query, page_size=page_size, page=page)
            serializer = RetrieveAuthorSerializer(authors, many=True)
            pagination_info = build_pagination_urls(request, page, page_size, serializer.data)
            data = serializer.data
            return Response({'total': total, 'next_page': pagination_info.get('next_page'),
                            'previous_page': pagination_info.get('previous_page'), 'data': data})
        except ValueError as ve:
            return Response({'error': str(ve)}, status=status.HTTP_400_BAD_REQUEST)
        except Exception as e:
            return Response({'error': str(e)}, status=status.HTTP_400_BAD_REQUEST)
```

Figura 2.35: Vista AuthorViews

También se implementó la serialización de los datos en la capa de infraestructura. Para ello se creó el serializador *AuthorSerializer* que se encargará de serializar los datos de los autores. Además, se utilizó Swagger para documentar la API REST. En la Figura 2.35 se puede visualizar el decorador<sup>2</sup> *extend\_schema* que permite la documentación automática de la API, en este caso se documenta el endpoint *find\_by\_query*. El resultado de la documentación se muestra en la Figura 2.36. La cual tiene una interfaz gráfica que permite probar los endpoints de la API.

<sup>2</sup>Un decorador es una función que recibe otra función y amplía su funcionalidad sin alterarla directamente.

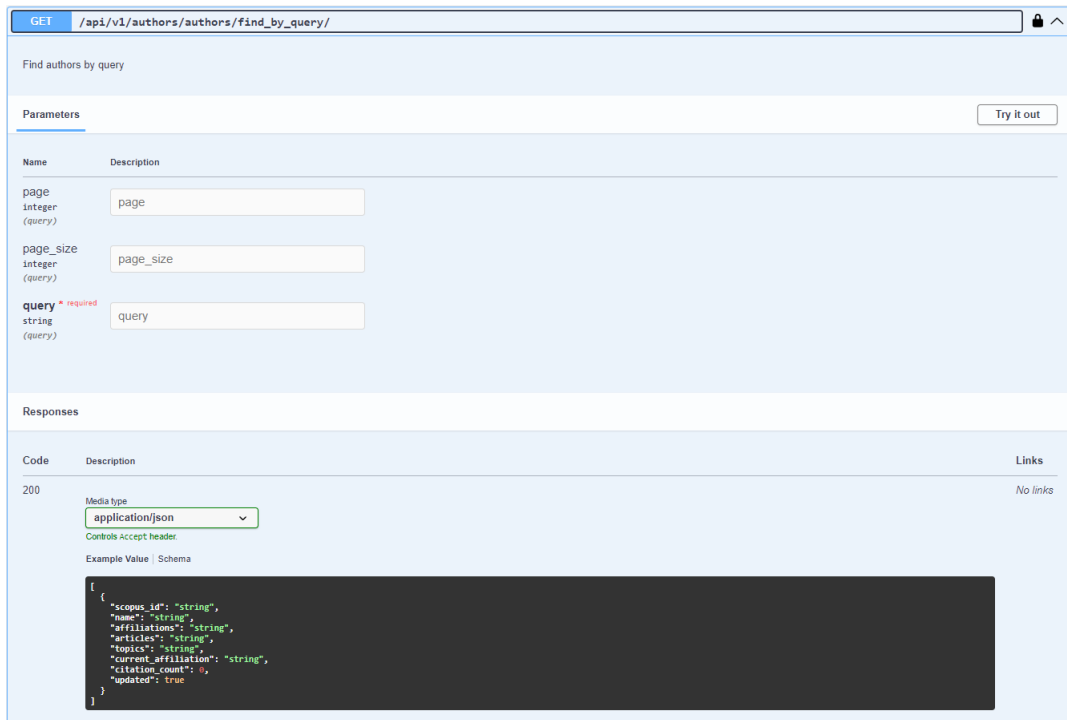


Figura 2.36: Documentación de la API con Swagger

Todo lo descrito anteriormente para la implementación de la funcionalidad de búsqueda de autores se puede resumir con un diagrama de secuencia. En la Figura 2.37 se muestra el diagrama de secuencia de la funcionalidad de búsqueda de autores. En resumen, con este diagrama de secuencia, se puede observar el flujo de la aplicación. La petición HTTP llega a la vista *AuthorViews*, la vista llama al caso de uso *FindAuthorsByQuery*, el caso de uso llama al repositorio *AuthorRepository* y este a su vez ejecuta la consulta en la base de datos. También podemos distinguir las capas por sus colores, por ejemplo la capa de dominio está representada en color naranja, la capa de aplicación en color amarillo y la capa de infraestructura en color verde. A partir de este punto se representará únicamente el flujo de la aplicación, sin entrar en detalles de la implementación.

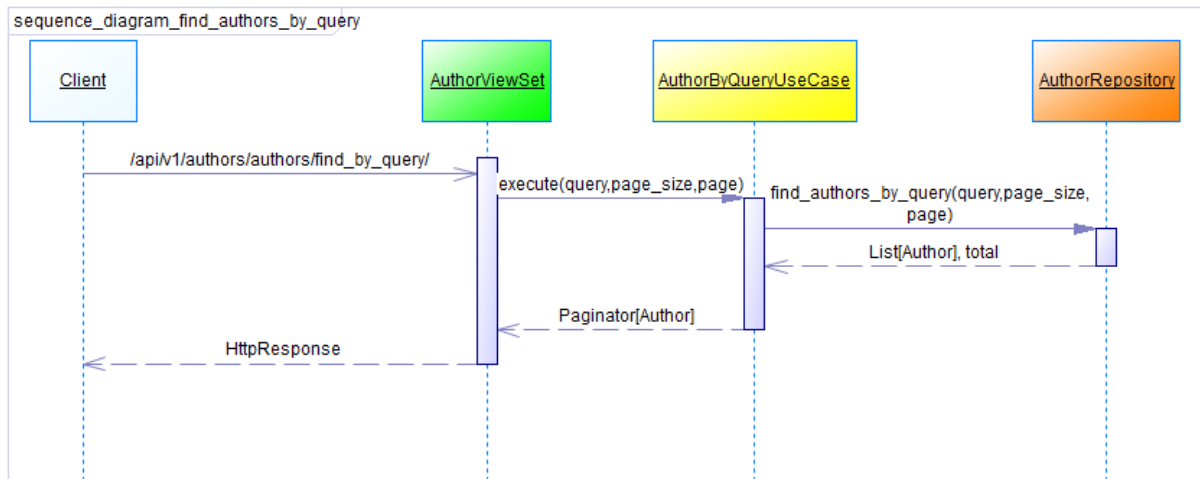


Figura 2.37: Diagrama de secuencia de la funcionalidad de búsqueda de autores

## 2.4.5. Revisión y Retrospectiva

En la revisión del Sprint 3 se verificó que se cumplieran los criterios de aceptación de las historias de usuario seleccionadas. Se comprobó que la funcionalidad de búsqueda de autores trabajara correctamente y que la documentación de la API REST estuviera actualizada. Además, se verificó que los modelos de la base de datos se encontraran en la base de datos. En la Figura 2.38 se muestra la base de datos con los nodos y relaciones creados.

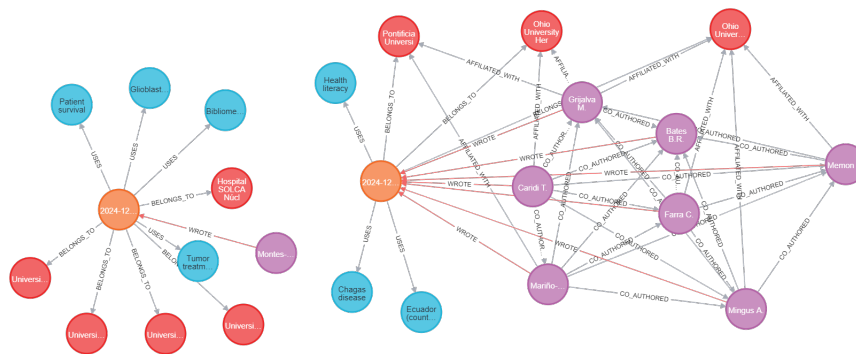


Figura 2.38: Base de datos Neo4j

Se considera que se cumplió con los objetivos planteados para este sprint, ya que se logró implementar la funcionalidad de búsqueda de autores y se rediseñó la base de datos en Neo4j utilizando Neomodel. Además, se logró levantar el ambiente de desarrollo con Docker y Docker Compose.

Sin embargo, se presentaron algunos problemas durante el desarrollo de las tareas asig-



nadas. Uno de los problemas más significativos fue la falta de documentación de Neomodel, lo que dificultó la implementación de los modelos en la base de datos. La falta de datos de prueba también fue un problema, ya que se tuvo que trabajar con datos ficticios para realizar las pruebas de la funcionalidad de búsqueda de autores.

## **2.5. Sprint 4**

### **2.5.1. Introducción**

El cuarto sprint tiene como objetivo alcanzar el estado final de Resnet, es decir, lograr que Centinela esté completamente funcional y lista para ser utilizada por los usuarios. En el contexto del motor de búsqueda, se espera que Centinela sea capaz de realizar las búsquedas que Resnet podía realizar en su estado final, alcanzando así el nivel de funcionalidad de la versión anterior.

### **2.5.2. Objetivos**

- Implementar la automatización para generar el Corpus y el Modelo de TF-IDF
- Implementar la funcionalidad de búsqueda de autores relevantes dado un tópico
- Implementar la funcionalidad de búsqueda de artículos relevantes dado un tópico
- Implementar la funcionalidad para obtener la red de coautoría de un autor
- Implementar la funcionalidad para obtener un Artículo dado su ID de Scopus
- Implementar la funcionalidad para obtener un Autor dado su ID de Scopus

### **2.5.3. Planificación**

Para este sprint se terminaron las tareas que quedaron pendientes en el Sprint 1 (Ver sección 2.2) y en el Sprint 2 (Ver sección 2.3). En la Tabla 2.4 se presentan las historias de usuario que se abordarán en este sprint con sus respectivas tareas.

Los criterios de aceptación para las historias de usuario HU-SE-01, HU-SE-03 y HU-SE-04 han sido definidos y se pueden consultar en las siguientes figuras: Figura 2.4, Figura 2.14 y Figura 2.15.

Identificador	Historia de Usuario	Tareas
HU-SE-01	Como usuario no registrado deseo poder encontrar artículos relevantes dado un tema de investigación para poder acceder rápidamente a información útil y actualizada que apoye mi estudio o trabajo.	<ul style="list-style-type: none"> <li>■ Generar el corpus <sup>3</sup> de datos</li> <li>■ Generar el modelo de TF-IDF a partir del corpus</li> <li>■ Crear el modelo de Artículo para la base de datos</li> <li>■ Implementar el filtro de artículos por año</li> <li>■ Crear un servicio de búsqueda de artículos</li> <li>■ Desarrollar un endpoint para manejar las solicitudes entrantes</li> </ul>
HU-SE-04	Como usuario no registrado quiero poder ver los artículos de un investigador para conocer su trabajo y las publicaciones en las que ha contribuido.	<ul style="list-style-type: none"> <li>■ Implementar un endpoint para recuperar el perfil de un investigador y sus publicaciones</li> <li>■ Crear un servicio de búsqueda de investigadores y sus artículos</li> </ul>
HU-SE-03	Como usuario no registrado, deseo poder ver la red de coautoría de un autor para visualizar un grafo con los autores con los que ha colaborado, así como la fuerza de esas colaboraciones.	<ul style="list-style-type: none"> <li>■ Implementar un endpoint para recuperar los datos de colaboración entre autores</li> <li>■ Crear un servicio que genere los datos del grafo de coautoría de un autor</li> </ul>

Tabla 2.4: Historias de Usuario del sprint 4

## 2.5.4. Implementación

Para la implementación de ciertas tareas en este Sprint, se optó por reutilizar el código de la versión anterior de ResNet, ya que se consideró más eficiente y rápido que comenzar desde cero. La optimización de funcionalidades existentes no está incluida en el alcance de este componente. Las funcionalidades que se reutilizarán son la generación del corpus y el modelo de TF-IDF, dado que estas tareas son las más demandantes en términos de tiempo y se determinó que no era necesario reimplementarlas. A estas funcionalidades solo se les añadirá una capa de infraestructura, ya que la lógica de negocio ya está implementada en la versión anterior de ResNet. Esto garantizará un punto de entrada para generar tanto el corpus como el modelo de TF-IDF, permitiendo su utilización en las búsquedas de autores y artículos.

Tarea: *HU-SE-01: Generar el corpus de datos*. Dado que las funcionalidades de buscar autores y artículos relevantes dependen del corpus de datos y de modelo de TF-IDF, estas tareas se realizarán en primer lugar. Para poder generar el corpus de datos, el endpoint `/api/v1/generate-corpus/` se encargará de recibir la petición mediante el método POST. Este llamará al caso de uso *GenerateCorpusUseCase* que se encargará de generar y ejecutar la lógica de negocio para guardar el corpus de datos. Este caso de uso toma datos de Autores, Artículos, y Tópicos de la base de datos Neo4j para crear los documentos y almacenarlos en un archivo Pickle. En la Figura 2.39 se muestra el diagrama de secuencia para la generación del corpus de datos.

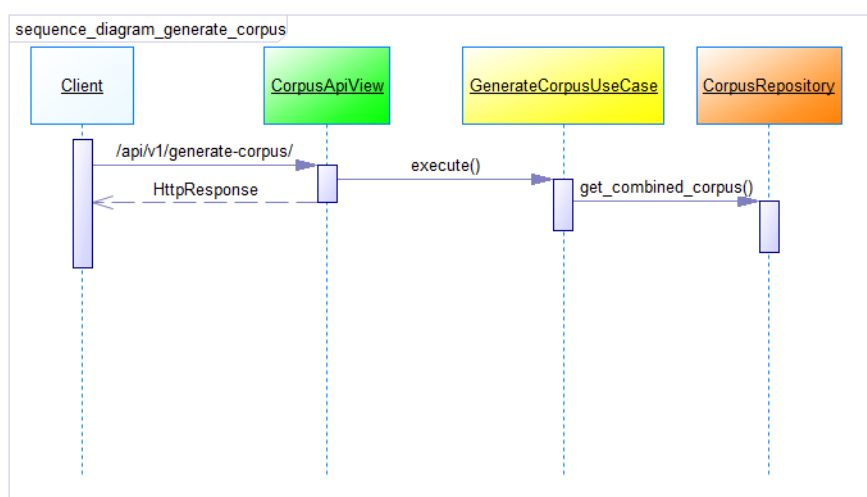


Figura 2.39: Diagrama de secuencia para la generación del Corpus

Tarea: *HU-SE-01: Generar el modelo de TF-IDF*. Una vez que el corpus de datos ha sido

generado, se procederá a generar el modelo de TF-IDF. Para ello, el endpoint `/api/v1/generate-model/` se encargará de recibir la petición mediante el método POST. Este llamará al caso de uso *GenerateTfidfUseCase* que se encargará de ejecutar la lógica de negocio para generar el modelo de TF-IDF, utilizando el corpus de datos previamente creado. Procesará este corpus para construir el modelo de TF-IDF y, finalmente, lo guardará en un archivo Pickle. En la Figura 2.40 se muestra el diagrama de secuencia para la generación del modelo de TF-IDF.

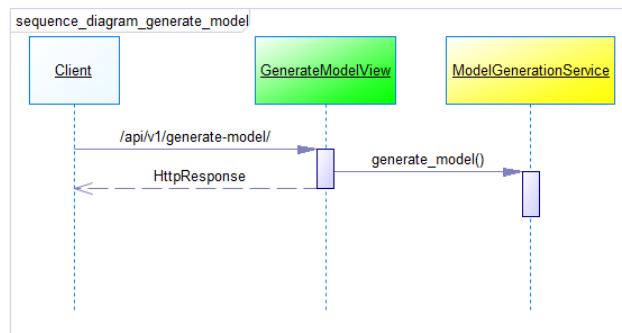


Figura 2.40: Diagrama de secuencia para la generación del Modelo de TF-IDF

Una vez que el corpus de datos y el modelo han sido generados se puede proceder a realizar las implementaciones tanto de autores relevantes como de artículos relevantes. Para la funcionalidad de Artículos relevantes, la ruta `/api/v1/articles/most-relevant-articles-by-topic/` (POST) recibirá en su cuerpo los siguientes datos:

- query: El Tópico por el cual se desea buscar los Artículos relevantes
- size: El límite de Artículos que se desea obtener, por defecto es 10
- page: La página de Artículos que se desea obtener, por defecto es 1
- type: Si se desea incluir o excluir años de publicación
- years: Años de publicación a incluir o excluir

Este endpoint llamará al caso de uso *GetMostRelevantArticlesByTopicUseCase* que se encargará de ejecutar la lógica de negocio para obtener los Artículos relevantes. El proceso comienza tomando el tópico y pasándolo al modelo TF-IDF para identificar los IDs de los artículos relevantes. A continuación, se consulta la base de datos Neo4j para obtener los datos de esos artículos. Utilizando los parámetros de la petición, se buscan y retornan los artículos relevantes en una lista paginada, todo esto dentro de una respuesta Http. En la

Figura 2.41 se muestra el diagrama de secuencia para la obtención de los Artículos relevantes.

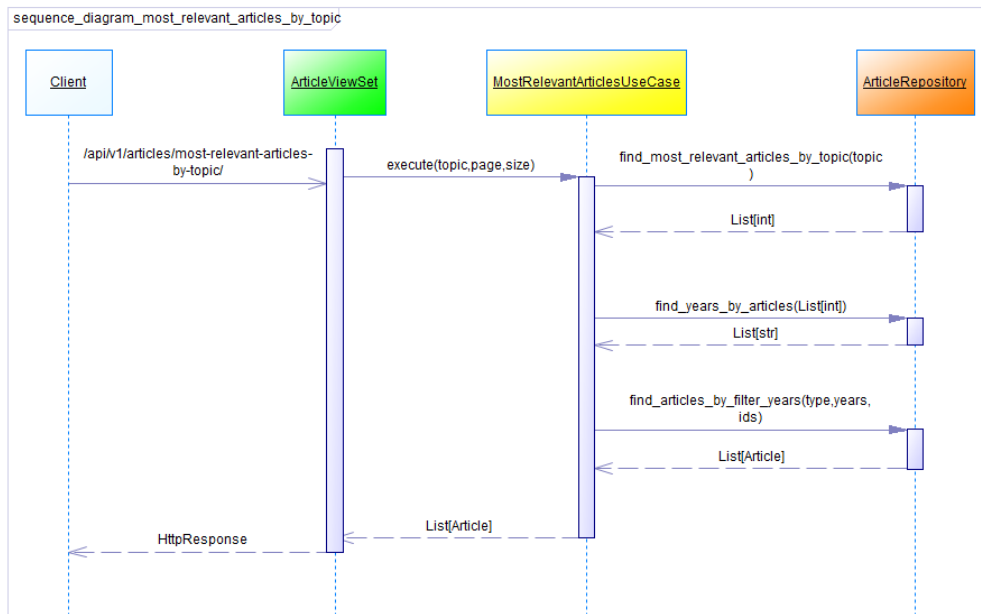


Figura 2.41: Diagrama de secuencia para la obtención de los Artículos Relevantes

HU-SE-03: Obtener la red de coautoría de un autor. Para la funcionalidad de obtener la red de coautoría de un autor, la ruta `/api/v1/coauthors/coauthors/id/coauthors_by_id/` (GET) recibirá en su URL el ID del autor del cual se desea obtener la red de coautoría. Este endpoint llamará al caso de uso `GetCoauthorsByAuthorIdUseCase` que se encargará de ejecutar la lógica de negocio para obtener la red de coautoría de un autor. El proceso comienza tomando el ID del autor y consultando la base de datos Neo4j para obtener los nodos (Autores) y links (`collab_strength`) de la red de coautoría de ese autor. A continuación, se consultan los datos de esos coautores y se retornan en una respuesta Http. En la Figura 2.42 se muestra el diagrama de secuencia para la obtención de la red de coautoría de un autor.

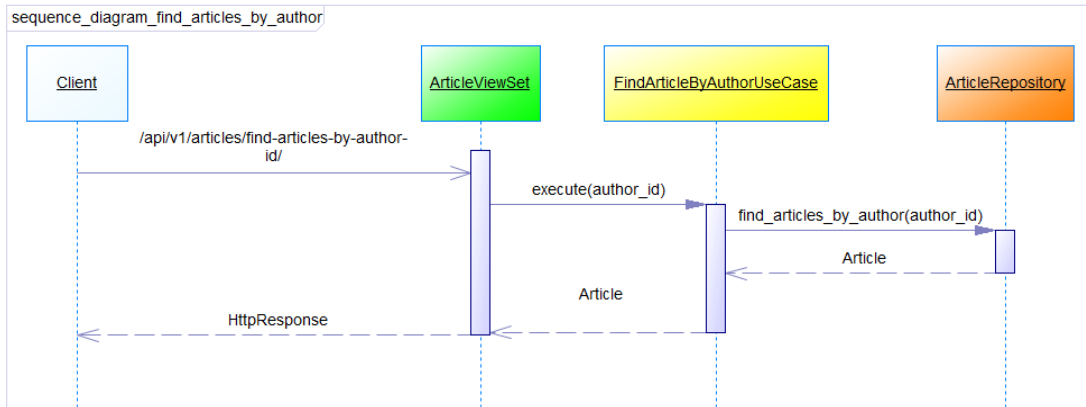


Figura 2.42: Diagrama de secuencia para la obtención de la red de coautoría de un autor

HU-SE-03: Para obtener los Artículos en los que ha colaborado un autor, en su URL la ruta `/api/v1/articles/find-articles-by-author-id/` (GET) recibirá el ID del autor del cual se desea obtener los artículos en los que ha colaborado. Este endpoint llamará al caso de uso *FindArticlesByAuthorIdUseCase* que se encargará de ejecutar la lógica de negocio para obtener los Artículos en los que ha colaborado un autor. En la Figura 2.43 se muestra el diagrama de secuencia para la obtención de los artículos en los que ha colaborado un autor.

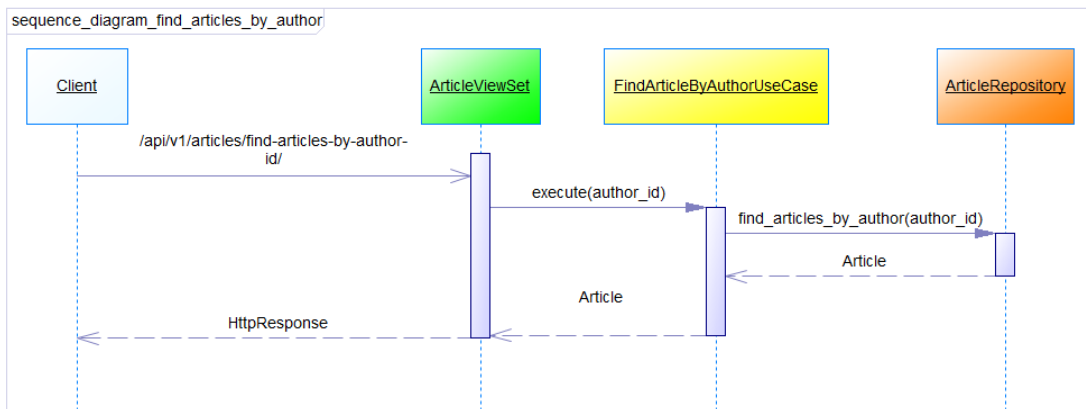


Figura 2.43: Diagrama de secuencia para la obtención de los Artículos en los que ha colaborado un autor

### 2.5.5. Revisión y Retrospectiva

Al finalizar el sprint, se logró implementar las funcionalidades de búsqueda de autores y artículos relevantes, así como la obtención de la red de coautoría de un autor y la obtención de los artículos en los que ha colaborado un autor, mismas que habían quedado pendientes en los sprints anteriores. Así mismo, se logró evidenciar que la reutilización de código de la

versión anterior de ResNet fue una buena decisión, ya que permitió acelerar el desarrollo de las funcionalidades. También el uso de Arquitectura Hexagonal permitió diagramar de manera clara y concisa la implementación de las funcionalidades.

## 2.6. Sprint 5

### 2.6.1. Introducción

El quinto Sprint tiene como objetivo, desarrollar el Componente A de la arquitectura propuesta en Figura 1.1, este componente se encarga de la automatización del proceso de extracción y limpieza de datos de Scopus para posteriormente almacenarlos en la base de datos Neo4j. Sumando así a Centinela la capacidad de obtener información actualizada desde Scopus.

### 2.6.2. Objetivos

- Implementar la automatización para la extracción de datos de Scopus.
- Implementar la automatización para la carga de datos de Scopus en Neo4j.
- Implementar un sistema de monitoreo para la automatización de la extracción, limpieza y carga de datos de Scopus.

### 2.6.3. Planificación

En este Sprint se abordarán las historias de usuario que se presentan en la Tabla 2.5.

Los criterios de aceptación para las historias de usuario HU-SE-07 y HU-SE-08 han sido definidos y se pueden consultar en las siguientes figuras: Figura 2.44 y Figura 2.45.

Acceptance Criteria
<b>Automatización del Proceso:</b> <ul style="list-style-type: none"><li>• El sistema debe ser capaz de realizar llamadas periódicas a la API de Elsevier para obtener la información de los artículos más recientes.</li><li>• Debe existir un mecanismo para detectar y evitar la duplicación de artículos ya almacenados en la base de datos.</li><li>• La solución debe integrarse con el sistema existente sin interrupciones.</li></ul>
<b>Procesamiento de Datos:</b> <ol style="list-style-type: none"><li>1. La información obtenida debe ser procesada y transformada antes de ser almacenada.</li><li>2. El procesamiento debe incluir validaciones para asegurar la integridad y la calidad de los datos.</li></ol>

Figura 2.44: Criterios de Aceptación HU-SE-07

Identificador	Historia de Usuario	Tareas
HU-SE-07	Como administrador de Centinela, deseo poder actualizar la información de artículos de manera automatizada, para que se reduzca el esfuerzo manual y se garantice que la base de datos esté siempre actualizada con la información más reciente.	<ul style="list-style-type: none"> <li>■ Configurar Acceso a la API de Scopus</li> <li>■ Procesar y limpiar los datos de Scopus</li> <li>■ Almacenar los datos en la base de datos Neo4j</li> <li>■ Crear un servicio de búsqueda de artículos</li> <li>■ Crear un panel admin del sistema</li> </ul>
HU-SE-08	Como administrador de Centinela, deseo poder monitorear el estado del proceso automatizado a través de un sistema de logs, para que pueda revisar y entender lo que está ocurriendo durante la carga y el procesamiento de los datos.	<ul style="list-style-type: none"> <li>■ Configurar un sistema de logging</li> <li>■ Almacenar los logs</li> <li>■ Desarrollar la interfaz para visualizar los logs</li> </ul>

Tabla 2.5: Historias de Usuario del Sprint 5

#### Acceptance Criteria



- El sistema debe generar logs detallados durante la carga y el procesamiento de los datos obtenidos de Scopus.
- Los logs deben registrar eventos clave, incluyendo errores, advertencias, y pasos completados exitosamente.
- El administrador debe poder acceder y revisar los logs de manera sencilla a través de una interfaz.

Figura 2.45: Criterios de Aceptación HU-SE-08

A continuación, en la Figura 2.46 se muestran las historias de usuario colocadas en Azure Boards para el Sprint 5.



Title	State
<ul style="list-style-type: none"> <li>Como administrador de Centinela, deseo poder actualizar la información de artículos de manera automatizada, para que se reduzca el esfuerzo manual y se garantice que la base de datos esté si...</li> </ul>	New
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Crear un panel admin en el sistema</li> </ul> </li> </ul>	To Do
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Procesar y limpiar los datos obtenidos:</li> </ul> </li> </ul>	To Do
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Configurar Acceso a la API de Scopus</li> </ul> </li> </ul>	To Do
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Desarrollar el endpoint para carga de datos</li> </ul> </li> </ul>	To Do
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Almacenar los datos en la base de datos Neo4j</li> </ul> </li> </ul>	To Do
<ul style="list-style-type: none"> <li>Como administrador de Centinela, deseo poder monitorear el estado del proceso automatizado a través de un sistema de logs, para que pueda revisar y entender lo que está ocurriendo durante ...</li> </ul>	New
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Configurar el sistema de logging</li> </ul> </li> </ul>	To Do
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Integrar el sistema de logging con la automatización de la carga de datos</li> </ul> </li> </ul>	To Do
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Almacenar los logs</li> </ul> </li> </ul>	To Do
<ul style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Desarrollar una interfaz para visualizar los logs</li> </ul> </li> </ul>	To Do

Figura 2.46: Historias de Usuario en Azure Boards para el Sprint 5

## 2.6.4. Implementación

Para el desarrollo de este Sprint se ha utilizado la API de Elsevier Scopus [9] para obtener la información de los artículos. Específicamente dos API's:

- **Search API:** Para obtener la información de los Artículos.
- **Author Retrieval API:** Para obtener la información detallada de los Autores.

El flujo de la automatización tanto para extraer la información de los Artículos así como información detallada sobre un Autor, se puede observar en la Figura 2.47.



Figura 2.47: Flujo de la Automatización

El flujo empezará con *Data Collector* que será el encargado de hacer las peticiones a la API de Elsevier, en este método se analizarán las posibles respuestas, pero solo si la respuesta tiene un código 200 se procederá a la siguiente fase, que es *Data Cleaning* este no es un método, si no un conjunto de reglas definidas. Por ejemplo en la Sección 2.1 se mencionó que los modelos van a tener un campo único que permitirá identificarlos, el mismo servirá en este caso para evitar duplicados. Finalmente el proceso de *Data Storage* se encargará de almacenar la información en la base de datos.

Las reglas de limpieza de datos que se han definido son las siguientes:

- **Topic:** El *Topic* debe ser único, por lo que en su campo *name* se ha definido como único.

- **Artículo:** El *Artículo* debe ser único, por lo que sus campos *scopus\_id* y *doi*<sup>4</sup> se han definido como únicos.
- **Autor:** El *Autor* debe ser único, por lo que se ha definido su *scopus\_id* como identificador único.
- **Afiliación:** La *Afiliación* debe ser única, por lo que se ha definido su *scopus\_id* como identificador único. Sin embargo, esto no fue suficiente ya que existían muchos datos duplicados con el nombre, por lo que el nombre de la *Afiliación* también pasó a ser único.

Para esta Sección, cabe mencionar que se reutilizaron ciertos métodos de Resnet, los mismos que permitirán configurar el cliente de la API de Elsevier.

Para la tarea HU-SE-07: Configurar Acceso a la API de Scopus, se ha configurado el cliente de la API de Elsevier, debido a que este cliente utiliza información sensible, se ha optado por utilizar variables de entorno para almacenar las credenciales de acceso.

En la Figura 2.48 se muestra la estructura del archivo que contendrá las variables de entorno.



```

.env
NEO4J_HOST=VALUE
NEO4J_PASSWORD=VALUE
NEO4J_USERNAME=VALUE
NEO4J_PORT=VALUE
X_ELS_APIKEY=VALUE
X_ELS_INSTTOKEN=VALUE
X_ELS_AUTHTOKEN=VALUE
DEBUG=VALUE
MONGO_DB_NAME=VALUE
MONGO_DB_USERNAME=VALUE
MONGO_DB_PASSWORD=VALUE
MONGO_DB_HOST=VALUE
MONGO_DB_PORT=VALUE
ADMIN_CENTINELA=VALUE
PASSWORD_CENTINELA=VALUE

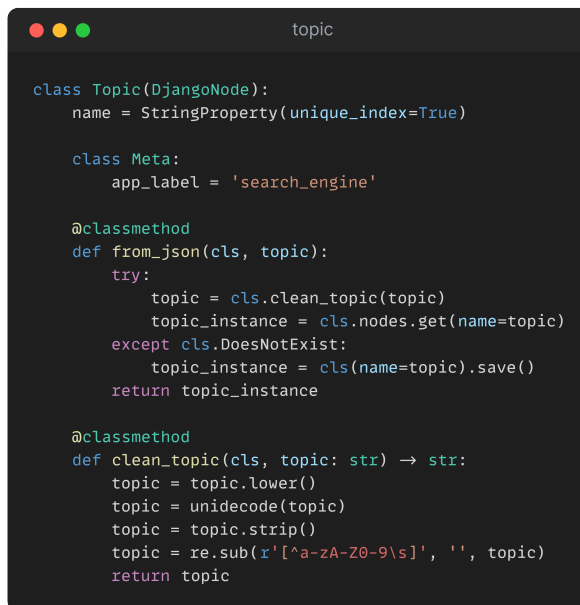
```

Figura 2.48: Estructura de las Variables de Entorno

Para la tarea HU-SE-07: Procesar y limpiar los datos de Scopus, como se menciono anteriormente, se ha definido un conjunto de reglas para limpiar los datos, al extraer información por ejemplo de un Topic, se ha definido que el mismo debe ser único, por lo que se ha definido un campo único en el modelo, con la propiedad *{unique\_index}* como verdadera. También para este caso específico se analiza si el tópico ya existe caso contrario se crea, a

<sup>4</sup>Digital Object Identifier es un identificador único y permanente para las publicaciones electrónicas

su vez tiene un método que hace la limpieza del tópico previo a almacenarlo en la base de datos como se muestra en la Figura 2.49.



```
class Topic(DjangoNode):
    name = StringProperty(unique_index=True)

    class Meta:
        app_label = 'search_engine'

    @classmethod
    def from_json(cls, topic):
        try:
            topic = cls.clean_topic(topic)
            topic_instance = cls.nodes.get(name=topic)
        except cls.DoesNotExist:
            topic_instance = cls(name=topic).save()
        return topic_instance

    @classmethod
    def clean_topic(cls, topic: str) -> str:
        topic = topic.lower()
        topic = unidecode(topic)
        topic = topic.strip()
        topic = re.sub(r'[^a-zA-Z0-9\s]', '', topic)
        return topic
```

Figura 2.49: Modelo Topic

Para la tarea HU-SE-07: Almacenar los datos en la base de datos Neo4j, vamos a utilizar como ejemplo la Figura 2.49 para mostrar como se almacena la información en la base de datos. Esta clase tiene un método `from_json` que se encarga de guardar la información en la base de datos, en este caso en Neo4j. Este método es un método de clase <sup>5</sup>, por lo que no es necesario instanciar la clase para poder utilizarlo. Toda la información que se obtiene de la API de Elsevier es en formato JSON <sup>6</sup>, por lo que se decidió utilizar este enfoque para facilitar la carga de datos en la base de datos.

Con esto se puede evidenciar como funcionan los procesos de extracción, limpieza y carga de datos que provienen desde Scopus. Este mismo procedimiento es aplicado para todos los modelos de la aplicación.

Para la tarea HU-SE-07: Crear un servicio de búsqueda de Artículos, se reutilizó la misma lógica de Resnet. Con la diferencia que en este caso al llegar la información en formato JSON, nos permite acceder a esos datos y guardarlos en la base de datos, haciendo referencia al flujo mencionado anteriormente.

En la Figura 2.50 se muestra un diagrama de secuencia de como se integra la API de Elsevier con la aplicación. La interacción empieza cuando el usuario hace una peti-

<sup>5</sup>Decorador que se utiliza para declarar un método estático dentro de una clase

<sup>6</sup>Javascript Object Notation, archivo con datos estructurados que se utiliza para comunicar aplicaciones

ción al endpoint `/api/v1/scopus-integration/` (GET) para buscar los Artículos que tengan alguna Afiliación en Ecuador. La vista (`ScopusIntegrationViewSet`) llama al caso de uso `ScopusIntegrationUseCase` que a su vez llama al repositorio `SearchAffiliationUseCase`. Este repositorio se encarga de hacer la petición a la API de Elsevier y devolver la información en formato JSON. La misma que es procesada y almacenada en la base de datos. Finalmente se tiene un `ModelCorpusObserver`, el que se encargará de eliminar los archivos Pickle tanto el Corpus como el Modelo ya que con nueva información ambos deben ser generados nuevamente.

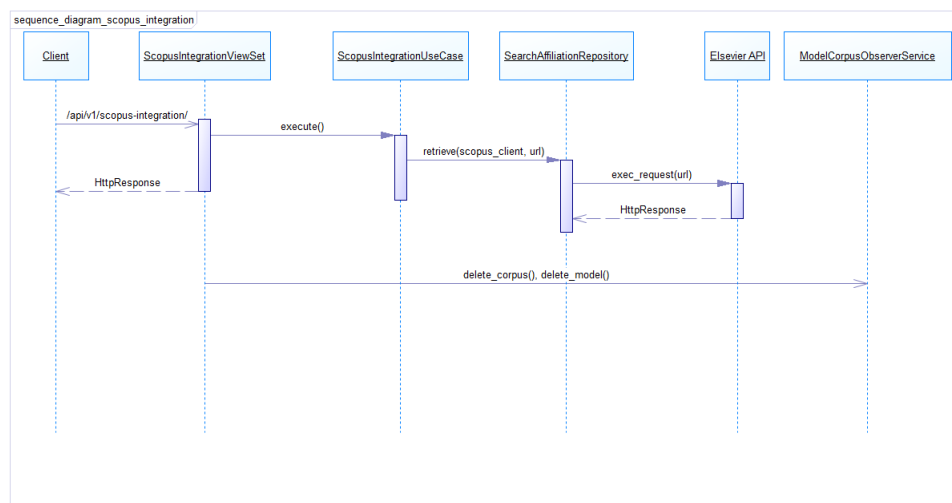


Figura 2.50: Diagrama de Secuencia de la Integración con la API de Elsevier

Entre otra información, para cada Artículo que devuelve la búsqueda en Elsevier, se obtiene la información general del Artículo, Afiliaciones, Autores, entre otros como se puede ver en la Figura 2.51.

Debido a esto, se identificó un problema. Si bien la información referente a los autores nos permite armar las redes de coautoría, mismas que fueron construidas a partir de la Similitud del Coseno [10], esta información no es suficiente para los intereses de Centinela. Por lo que se decidió utilizar la API de Author Retrieval para obtener la información detallada de los autores.

Para la tarea HU-SE-07: Implementar el servicio para obtener la información detallada de los Autores, se ha utilizado la API de Author Retrieval de Elsevier. De igual manera la información que se obtiene es en formato JSON, el enfoque presentado anteriormente es aplicado en este caso. Dado que la información que se obtiene de la API de Author Retrieval es muy extensa, no se ha podido mostrar un ejemplo de la misma. Sin embargo se puede consultar en la documentación de Elsevier [9].

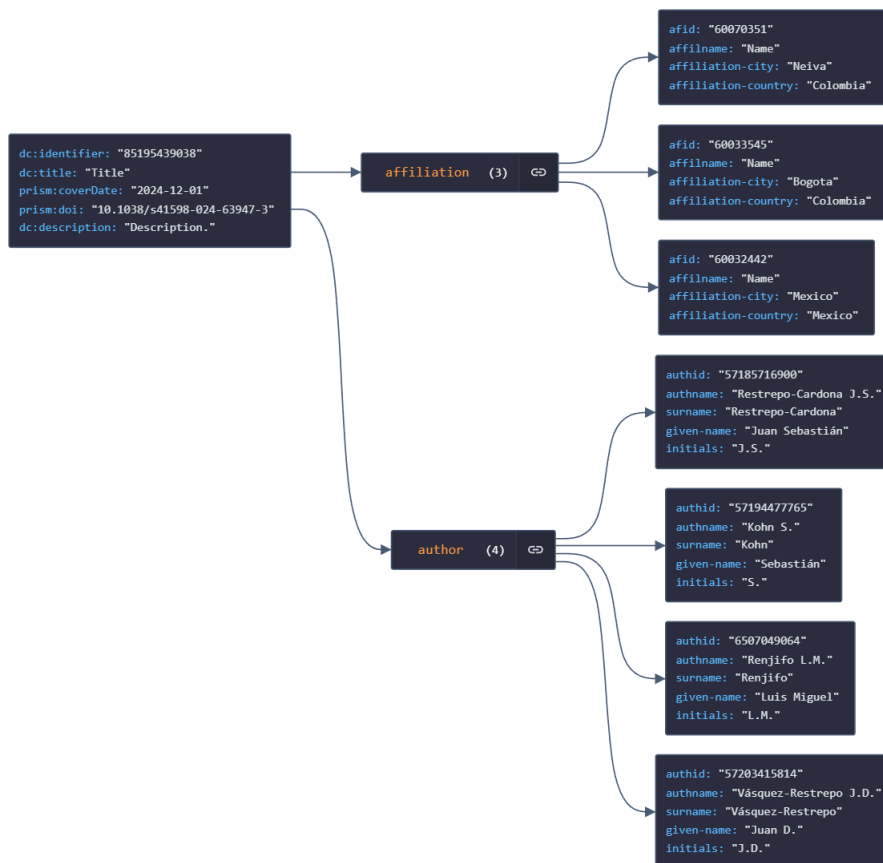


Figura 2.51: Ejemplo de JSON de un Artículo

El diagrama de secuencia de la integración con la API de Author Retrieval se muestra en la Figura 2.52.

El mismo que muestra información general sobre el estado de Centinela. Por ejemplo cuantos Artículos y Autores se han almacenado en la base de datos. Y una comparación de los mismos con la información que se obtiene de la API de Elsevier. De igual manera se tiene una sección para actualizar el corpus y el modelo.

De manera similar al diagrama de secuencia de la integración con la API de Elsevier, el flujo empieza con el usuario haciendo una petición al endpoint `/api/v1/author-retrieval/` (GET) para buscar la información detallada de un Autor. La vista `AuthorRetrievalView` llama al caso de uso `AuthorRetrievalUseCase` que a su vez llama a su respectivo repositorio `AuthorRetrievalRepository`. Este repositorio se encarga de hacer la petición a la API de Author Retrieval y devolver la información en formato JSON. La misma que es procesada y almacenada en la base de datos. Nuevamente el Corpus y el Modelo son eliminados para ser generados nuevamente.

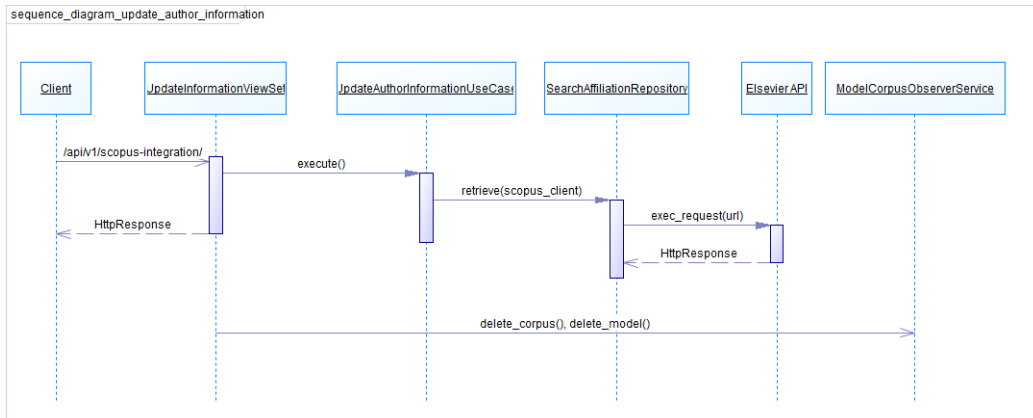


Figura 2.52: Diagrama de Secuencia de la Integración con la API de Author Retrieval

Una vez que se ha desarrollado estas funcionalidades, el siguiente paso es dar un punto de entrada. Para esto se ha desarrollado un panel de administración, que permitirá al administrador actualizar la información de los Artículos y Autores de manera automatizada.

Para la tarea HU-SE-08: Configurar un sistema de logging, se ha utilizado la librería `logging` de Python, la misma que permite almacenar los logs en un archivo. En la Figura 2.53 se muestra un ejemplo de como se configura el sistema de logging.

La configuración mostrada en la Figura 2.53 permite almacenar los logs en un archivo, con rotación diaria a medianoche y conservación de hasta 30 archivos de respaldo. Los mensajes se formatean incluyendo la fecha y hora, el nivel del registro y el mensaje en sí. El nivel del registro que se consideró fue `INFO`, debido a que es el nivel que ofrece información relevante sobre el proceso, pero se puede cambiar a `DEBUG` para obtener más información.

Finalmente para la tarea HU-SE-07: Crear un panel admin para actualizar la información de los Artículos y Autores, se ha creado un Login previo para que únicamente el administrador de Centinela pueda acceder a dicho panel. El login se muestra en la Figura 2.54.

```
settings
LOGGING = {
    "version": 1,
    "disable_existing_loggers": False,
    "handlers": {
        "file": {
            "level": "INFO",
            "class": "logging.handlers.TimedRotatingFileHandler",
            "filename": "/centinela_logs/info.log",
            "when": "midnight",
            "backupCount": 30,
            "formatter": "verbose",
            "encoding": "utf-8",
        },
    },
    "formatters": {
        "verbose": {
            "format": "{asctime} {levelname} {message}",
            "style": "{",
        },
    },
    "loggers": {
        "django": {
            "handlers": ["file"],
            "level": "DEBUG",
            "propagate": True,
        },
    },
}
```

Figura 2.53: Configuración del Sistema de Logging

Welcome to Centinela's Admin

Enter your credentials to access your account

Username

Password

Login

Figura 2.54: Login de Administrador

Dado que ese un sistema de Autenticación y Autorización está fuera de el alcance de este componente se ha optado por utilizar un sistema de autenticación básico, que consiste en un usuario y contraseña que se almacenan en variables de entorno que se muestran en la Figura 2.48.

Una vez autenticado el administrador, se le presentará el panel que se muestra en la Figura 2.55.

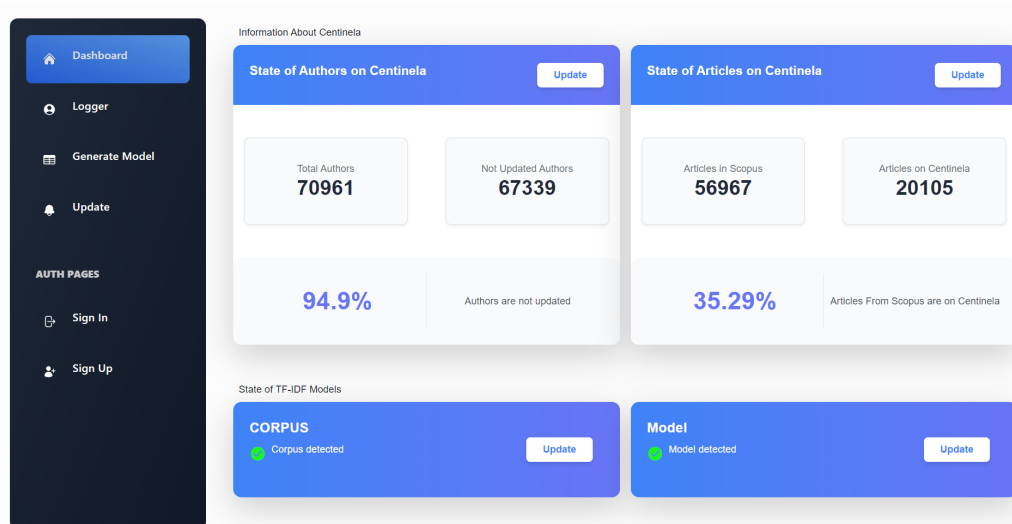


Figura 2.55: Panel de Administración

Este muestra información general sobre el estado de Centinela, como la cantidad de artículos y autores almacenados en la base de datos, y ofrece una comparación con los datos obtenidos de la API de Elsevier. También incluye una sección para actualizar tanto el corpus como el modelo.

La otra sección es el panel en donde se pueden observar los logs de la aplicación, como se muestra en la Figura 2.56, este cuenta con una interfaz que permite filtrar los logs por nivel, fecha o palabra clave.



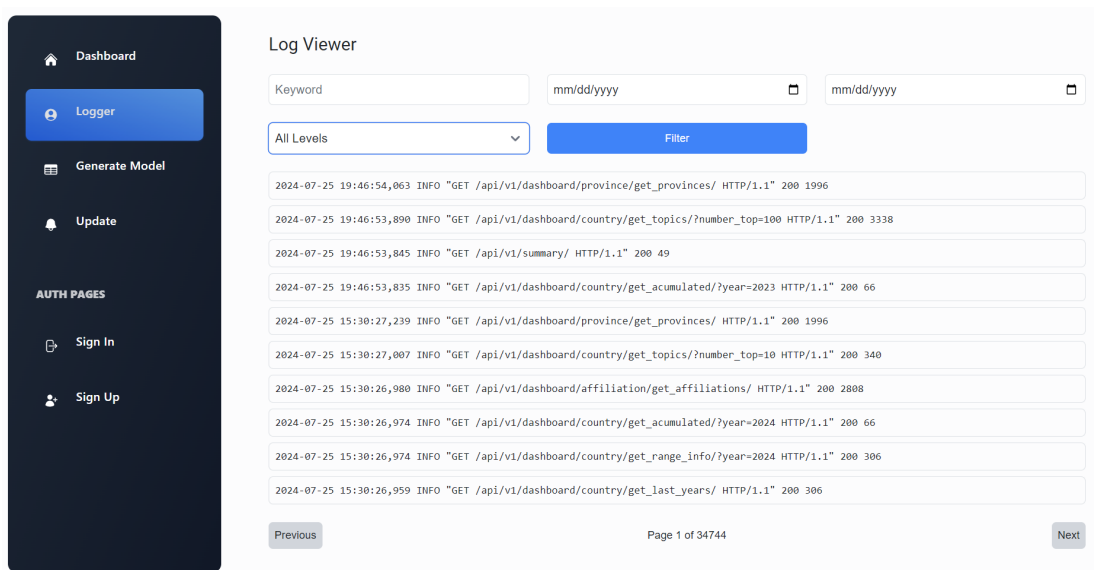


Figura 2.56: Panel de Logs

## 2.6.5. Revisión y Retrospectiva

Al finalizar el Sprint 5, se ha logrado implementar la automatización para la extracción de datos de Scopus, la automatización para la carga de datos de Scopus en Neo4j y el sistema de monitoreo para la automatización de la extracción, limpieza y carga de datos de Scopus. Se ha logrado cumplir con los objetivos planteados en la planificación del Sprint. Sin embargo la base de datos aun no cuenta con toda la información que genera las consultas a la API de Elsevier, esto debido a que la información que se obtiene de la API de Elsevier es muy extensa y se requiere de un mayor tiempo para poder procesarla y almacenarla en la base de datos. Además que las credenciales de acceso a la API tienen restricciones semanales que limitan la cantidad de información que se puede obtener.

## Capítulo 3

# RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

### 3.1. Resultados

#### 3.1.1. Capacidades de Centinela

Después de completar la implementación de seis historias de usuario, este componente ha dado lugar a una API robusta que consta de 22 endpoints. Esta API no solo suministra información crucial al motor de búsqueda de Centinela, sino que también facilita la integración con Scopus a través de la API de Elsevier. Este conjunto de API's están detalladas con la funcionalidad que hacen, el método al cual hacer referencia y el resultado y si aplica, lo que se debe enviar a través del body (Ver Figura 3.1).

List all logs

List all logs

AUTHORIZATIONS: > *cookieAuth or basicAuth or None*

QUERY PARAMETERS

end_date	string End date in YYYY-MM-DD format
keyword	string Keyword to search in logs
level	string Log level (e.g., DEBUG, INFO, ERROR)
lines_per_page	integer Lines per page
page	integer Page number
start_date	string Start date in YYYY-MM-DD format

Response samples

200

Content type  
application/json

```
{
  "lines_per_page": 10,
  "page": 1,
  "level": "INFO",
  "start_date": "",
  "end_date": "",
  "keyword": ""
}
```

Figura 3.1: Documentación detallada

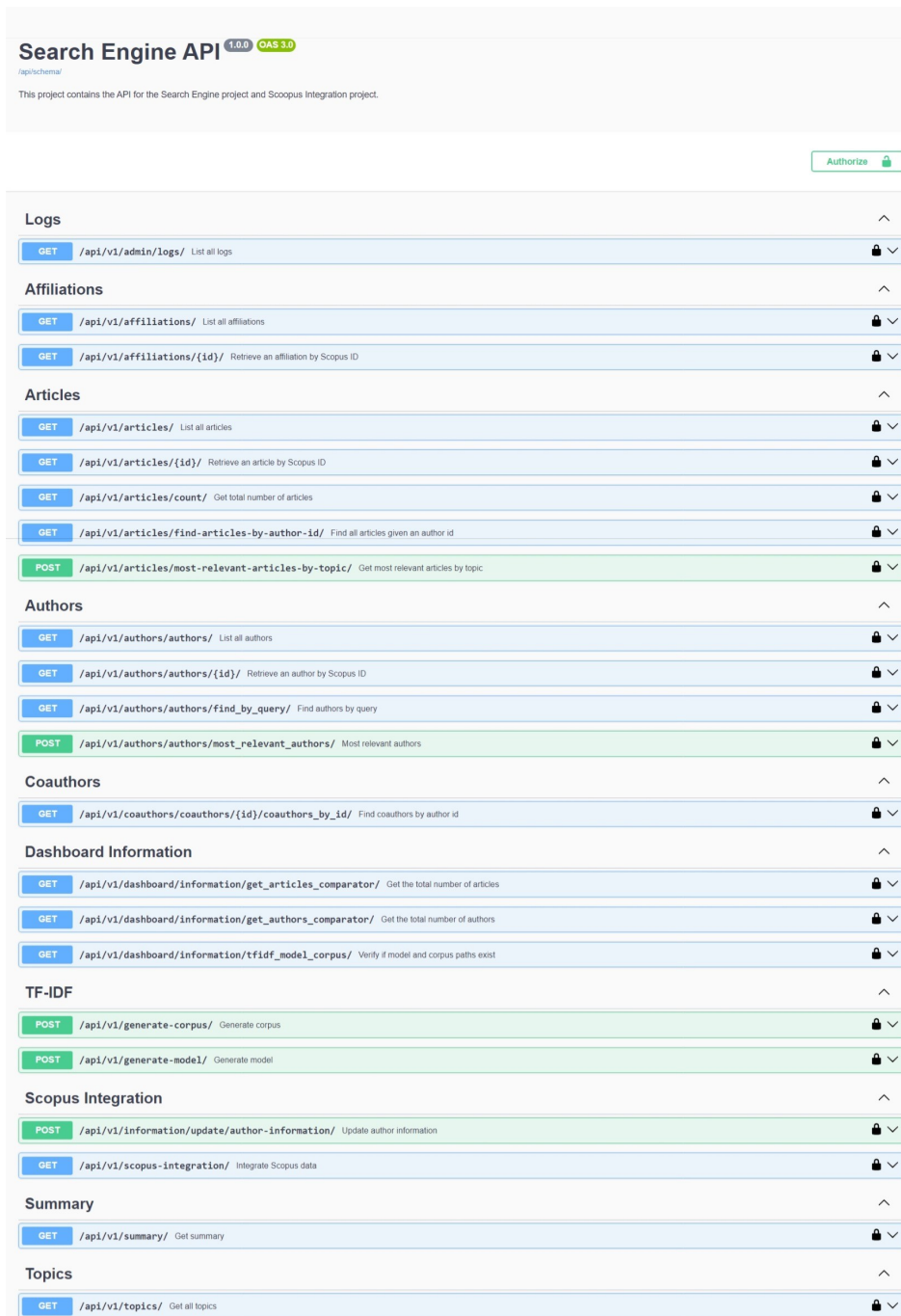


Figura 3.2: Documentación de la API generada con Swagger

### 3.1.2. Prueba con usuarios finales

Para evaluar la usabilidad de la aplicación se utilizó el cuestionario SUS <sup>1</sup>, se trata de un cuestionario de 10 preguntas como se muestra en la Tabla 3.1.2 con una escala de 5

<sup>1</sup>SUS por sus siglas en ingles System Usability Scale permite medir que tan usable es un sistema mediante un cuestionario de 10 preguntas

puntos, donde 1 es muy en desacuerdo y 5 muy de acuerdo.

Antes de la prueba, se les proporcionó a los participantes una breve introducción sobre las funciones básicas de la aplicación. Posteriormente, se les permitió explorar libremente la interfaz y sus características durante 10 minutos, con el fin de familiarizarse con su uso. A continuación, se les asignaron tareas específicas para evaluar distintos aspectos de la usabilidad, como la facilidad de navegación, la claridad de las instrucciones y la eficiencia en la búsqueda de información relevante.

En la Figura 3.3 se muestra los resultados para la pregunta 1 del cuestionario SUS.

**Creo que me gustaría utilizar este sistema con frecuencia**

14 respuestas

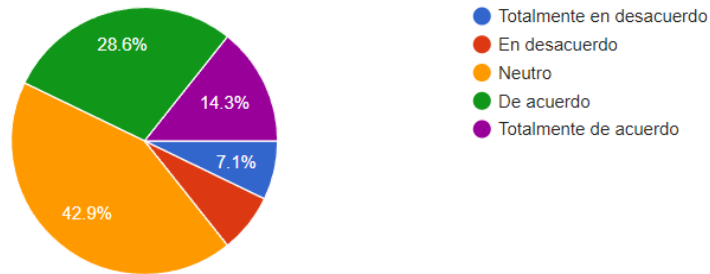


Figura 3.3: Resultados de la pregunta 1 del cuestionario SUS

Las demás preguntas del cuestionario SUS se encuentran en el Anexo 5.

Numero	Pregunta
1	Creo que me gustaría usar esta aplicación frecuentemente.
2	Encontré la aplicación innecesariamente compleja.
3	Creo que la aplicación es fácil de usar.
4	Creo que necesitaría el apoyo de un experto para usar esta aplicación.
5	Encontré que las diferentes funciones de la aplicación estaban bien integradas.
6	Creo que la aplicación es muy inconsistente.
7	Imagino que la mayoría de la gente aprendería a usar esta aplicación muy rápidamente.
8	Encontré la aplicación muy difícil de usar.
9	Me sentí muy confiado al usar la aplicación.
10	Necesité aprender muchas cosas antes de poder empezar a usar la aplicación.

Tabla 3.1: Cuestionario SUS

El cuestionario fue aplicado a 14 estudiantes de la Facultad de Ingeniería de Sistemas de la Escuela Politécnica Nacional, los resultados obtenidos se muestran en el Anexo 5.

La puntuación obtenida fue de 70.89 puntos. De acuerdo con la escala de Jeff Sauro [23], presentada en la Figura 3.4, esta puntuación indica que la usabilidad de la aplicación es aceptable. No obstante, aún hay margen para realizar mejoras.

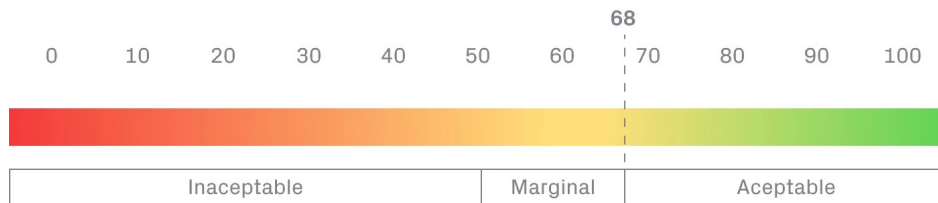


Figura 3.4: Escala de Jeff Sauro

## 3.2. Conclusiones

El análisis de ResNet y ResearchDecide permitió identificar puntos clave en cada una de estas herramientas. A partir de este análisis, se definió una estrategia de integración que mantiene la independencia de cada sistema, mientras se facilita una comunicación eficiente entre las mismas. Esta estrategia asegura que ambos sistemas puedan colaborar sin comprometer su autonomía, optimizando el flujo de información y mejorando la eficacia global del proceso.

El uso de la Arquitectura Hexagonal en la versión 1 de Centinela, bajo el enfoque de 'arquitectura limpia', permitió abordar de manera eficiente la separación de responsabilidades y la integración de componentes. Este enfoque facilitó la escalabilidad del sistema, mejoró la mantenibilidad del código y aseguró que los módulos centrales permanecieran independientes de las tecnologías externas, permitiendo adaptaciones futuras sin afectar la lógica de negocio. Aunque su implementación requirió un mayor esfuerzo en la fase de diseño, se logró una mayor eficiencia en la fase de desarrollo.

La integración automatizada vía API con Scopus ha convertido a Centinela en una herramienta más robusta y completa. Esta integración permite acceder a la información más actualizada y confiable de la base de datos de Scopus. Gracias a este proceso automatizado, un administrador puede extraer información directamente desde Scopus, asegurando que la base de datos de Centinela se mantenga siempre actualizada de manera eficiente y sin intervención manual. Además que se garantiza la calidad, integridad y disponibilidad

de los datos, lo que mejora la confiabilidad de los resultados obtenidos por los usuarios de Centinela.

### **3.3. Recomendaciones**

- Se sugiere obtener credenciales de acceso para el API Elsevier, con mayor capacidad de consultas, en especial para la obtención de información de Autores, que se hace mediante la API de Author Retrieval. Actualmente se cuenta con la limitación de 5000 consultas por semana, lo que puede ser ineficiente para la obtención de información de autores en grandes cantidades.
- Se recomienda mantener el enfoque de Arquitectura Hexagonal para futuras funcionalidades y mejoras en el sistema, ya que permite una fácil escalabilidad y mantenimiento del sistema.
- Se recomienda revisar constantemente la documentación de Neomodel, ya que al ser una herramienta relativamente nueva esta en constante actualización y puede presentar cambios en su funcionamiento y en la forma de implementar ciertas funcionalidades.
- La funcionalidad de buscar Autores dado un query (nombre, apellido, ID, etc) en ocasiones no retorna resultados esperados, se recomienda revisar la lógica de búsqueda y la forma en que se realiza la consulta a la base de datos de Neo4j.

## Capítulo 4

# REFERENCIAS BIBLIOGRÁFICAS

- [1] Primicias, «Investigación científica en Ecuador despunta,» *Primicias*, 2020, Disponible en: <https://www.primicias.ec/noticias/tecnologia/investigacion-cientifica-ecuador-despunto>.
- [2] J. Arias y L. Recalde, «Search and Visualization of Researcher Networks: Co-authorship in Ecuador,» en *Conference on Information and Communication Technologies of Ecuador*, Springer, 2023, págs. 448-463.
- [3] D. E. A. Checa, *Desarrollo de un sistema de toma de decisiones y consenso grupal aplicado a los motores de recomendación: desarrollo del módulo de manejo de usuarios, grupos y notificaciones para el sistema de toma de decisiones y consenso grupal aplicado a los motores de recomendación*, Quito, ago de 2023. dirección: <http://bibdigital.epn.edu.ec/handle/15000/25087>.
- [4] R. F. P. Molina, *Desarrollo de un sistema de toma de decisiones y consenso grupal aplicado a los motores de recomendación: desarrollo de los módulos de presentación de la recomendación y de selección de tópicos del sistema de recomendación para grupos de investigadores que soporte la toma de decisiones y el consenso*, Quito, oct. de 2023. dirección: <http://bibdigital.epn.edu.ec/handle/15000/25084>.
- [5] Elsevier, *Scopus*, Accedido: 2024-06-25, 2024. dirección: <https://www.scopus.com>.
- [6] Nukas. «La teoría de redes sociales y un ejemplo de su aplicación en la pandemia de COVID-19.» Accedido el 15 de julio de 2024. (2021), dirección: <https://naukas.com/2021/12/06/la-teoria-de-redes-sociales-y-un-ejemplo-de-su-aplicacion-en-la-pandemia-de-covid-19/#:~:text=La%20teor%C3%ADa%20de%20>

20redes%20sociales%20es%20la%20confluencia%20de%20diferentes , de%20una%20sociedad%20%5B10%5D..

- [7] M. Á. de Dios, *Scrum: qué es y cómo funciona este marco de trabajo*, Last accessed 09 June 2024, 2022. dirección: <https://www.wearemarketing.com/es/blog/metodologia-scrum-que-es-y-como-funciona.html>.
- [8] D. M. Rehkopf, *¿Qué son los sprints en la gestión de proyectos?* Last accessed 09 June 2024, 2017. dirección: <https://atlassian.com/es/agile/scrum/sprints>.
- [9] Elsevier, *Elsevier Developer Portal*, Accedido: 2024-07-25, 2024. dirección: <https://dev.elsevier.com>.
- [10] G. Everywhere, *Algoritmo de Similitud de Coseno*, Accedido: 28-06-2024, 2024. dirección: <https://www.grapheverywhere.com/algoritmo-de-similitud-de-coseno/>.
- [11] D. D. West, *Funciones y responsabilidades de scrum de metodología ágil*, Last accessed 09 June 2024, 2017. dirección: <https://www.atlassian.com/es/agile/scrum/roles#:~:text=%C2%BFcu%C3%A1les%20son%20las%20tres%20funciones,miembros%20del%20equipo%20de%20desarrollo..>
- [12] JetBrains, *PyCharm Professional para el desarrollo web*, Last accessed 08 June 2024, 2024. dirección: <https://www.jetbrains.com/es-es/pycharm/>.
- [13] D. S. Foundation, *Django Rest Framework*, Last accessed 08 June 2024, 2024. dirección: <https://www.django-rest-framework.org/>.
- [14] Microsoft, *Visual Studio Code Update v1.91*, Accedido: 2024-06-25, 2024. dirección: [https://code.visualstudio.com/updates/v1\\_91](https://code.visualstudio.com/updates/v1_91).
- [15] Neomodel, *Neomodel Documentation*, Accedido: 2024-06-25, 2024. dirección: <https://neomodel.readthedocs.io/en/latest/>.
- [16] Angular, *Angular Build Guide*, Accedido: 2024-06-25, 2024. dirección: <https://v17.angular.io/guide/build>.
- [17] IBM, *IBM App Connect Documentation: APIs and Swagger*, Accedido: 2024-06-25, 2024. dirección: <https://www.ibm.com/docs/es/app-connect/11.0.0?topic=apis-swagger>.
- [18] RedHat, *¿Qué Docker y cómo funciona?* Last accessed 09 June 2024, 2023. dirección: <https://www.redhat.com/es/topics/containers/what-is-docker>.



- [19] S. (SAP), *SAP PowerDesigner*, Herramienta de modelado de datos y diseño de bases de datos, 2024. dirección: <https://www.powerdesigner.biz/ES/>.
- [20] M. Learn, *¿Qué es Azure DevOps?* Last accessed 09 June 2024, 2024. dirección: <https://learn.microsoft.com/es-es/azure/devops/user-guide/what-is-azure-devops?view=azure-devops>.
- [21] E. Novoseltseva, *¿Qué Es La Arquitectura Hexagonal? Definición Y Ejemplos*, Last accessed 09 June 2024, 2020. dirección: <https://apiumhub.com/es/tech-blog-barcelona/arquitectura-hexagonal/>.
- [22] M. D. Network, *Mobile First - Glosario de MDN Web Docs*, Last accessed 09 June 2024, 2023. dirección: [https://developer.mozilla.org/es/docs/Glossary/Mobile\\_First](https://developer.mozilla.org/es/docs/Glossary/Mobile_First).
- [23] Teacup Lab. «¿Qué es la escala SUS y cómo usarla para medir la usabilidad?» Ac-  
cedido: 28-06-2024. (2024), dirección: <https://www.teacuplab.com/es/blog/que-es-la-escala-sus-y-como-usarla-para-medir-la-usabilidad/>.

## Capítulo 5

# ANEXOS

**Anexo 1:** Repositorio de Centinela en Angular: Puedes encontrar el repositorio en **GitHub**.

**Anexo 2:** Repositorio del servicio de motor de búsqueda en Django: Disponible en **GitHub**.

**Anexo 3:** Prototipo de la aplicación en Figma: Accede al prototipo en **Figma**.

**Anexo 4:** Evaluación SUS: Resultados de la evaluación en **OneDrive**.

**Anexo 5:** Diagramas de secuencia: Visualiza los diagramas de secuencia en **OneDrive**.