

**VICERRECTORADO DE DOCENCIA**  
**PLAN DE TRABAJO DE INTEGRACIÓN CURRICULAR**

I. INFORMACIÓN BÁSICA	
<b>Unidad Académica:</b>	Facultad de Ingeniería de Sistemas
<b>Carrera:</b>	Ingeniería en Ciencias de la Computación
<b>Proyecto:</b>	Proyecto Centinela V2
<b>Componente:</b>	Implementación de un Sistema RAG (Retrieval-Augmented Generation) en el Buscador de Centinela
<b>Línea de investigación:</b>	Inteligencia artificial aplicada a sistemas de informacion – Tópico: Recuperacion de informacion asistida por GPTs
<b>Nombres y apellidos del estudiante:</b>	Alejandro Sebastian Chavez Vega
<b>Nombres y apellidos del Profesor:</b>	Dra. Gabriela Suntaxi
II. DESCRIPCIÓN DEL COMPONENTE	
<p>Con la popularizacion de los LLM, los sistemas de recuperacion de informacion se adapataron a busquedas con lenguaje natural y tener un contexto semantico de lo que requiere el usuario. Lo que principalmente se ajusta a lo que necesita centinela. Centinela es una plataforma que busca fortalecer la investigacion científica en Ecuador. Su objetivo principal es mejorar el acceso a la información científica y fomentar la colaboración entre investigadores. Incorporar RAG en Centinela transforma la plataforma en un agente capaz de interpretar el contexto del usuario y sus intereses para generar respuestas coherentes y utiles. A diferencia con el buscador tradicional actual que simplemente devuelve una lista de articulos, RAG primero recupera los documentos mas relevantes y luego genera un respuesta en lenguaje natural, basada en los articulos. Esto permite que el investigador tenga una sintesis clara incluso si no se usa terminos tecnicos exactos, el sistema extrae un resumen de los resultados obtenidos y los autores, todo a partir de los articulos disponibles en la base de datos.</p>	

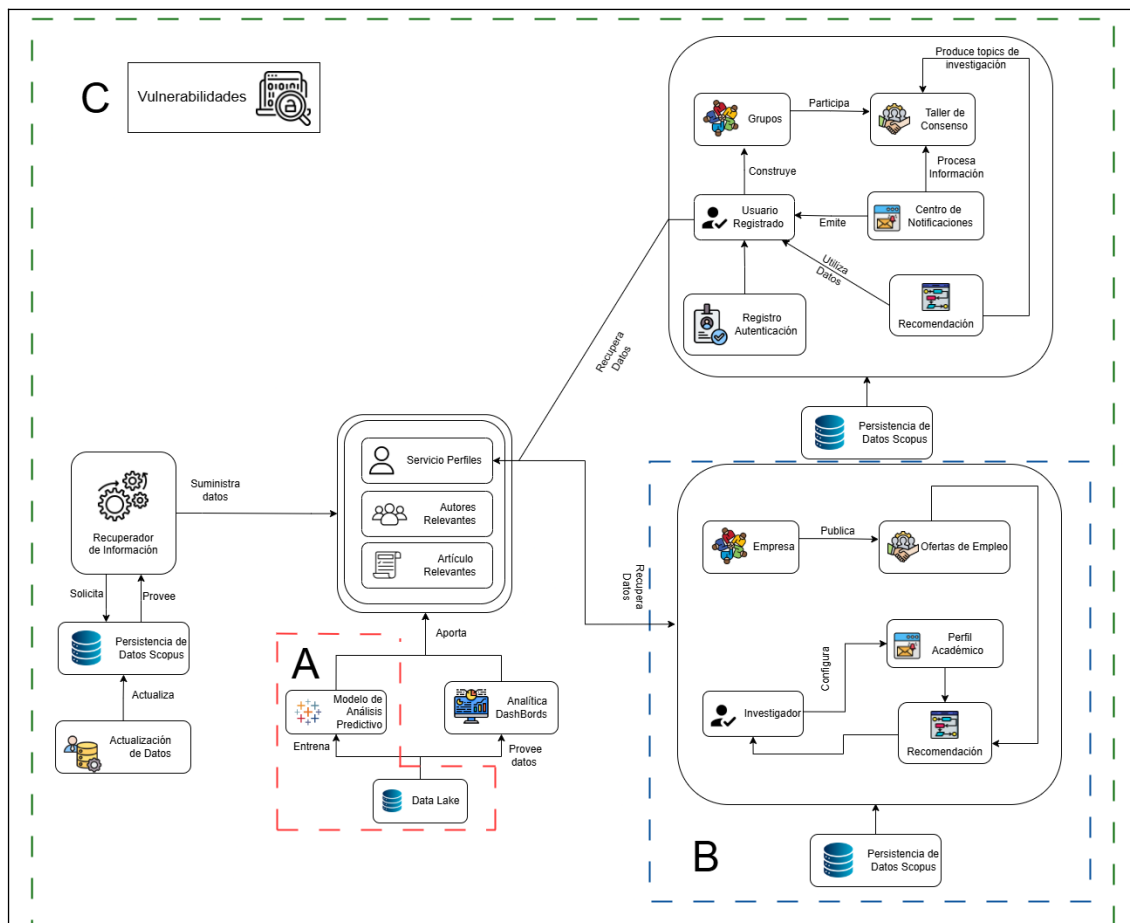


Figura 1. Arquitectura del sistema Centinela.

Como se observa en la figura, la plataforma se divide en tres módulos principales:

- **Recuperador de Información:** Este módulo gestiona la persistencia de datos provenientes de Scopus, permitiendo búsquedas actualizadas por perfiles, autores y artículos relevantes. Aquí se propone integrar el componente RAG como una capa adicional sobre el motor de búsqueda actual. Este componente sería responsable de interpretar consultas en lenguaje natural, recuperar artículos relevantes y generar respuestas comprensibles basadas en el contenido de los documentos. Esta capa puede alojarse dentro del backend del Recuperador, conectándose tanto al índice como a los modelos de generación.
- **Servicio de Analítica Descriptiva:** A través de este servicio, Centinela ofrece información para la toma de decisiones a través de visualizaciones de datos. Este módulo se complementa con un 'Data Lake' que proporciona los recursos necesarios para la analítica y la generación de dashboards interactivos.
- **Interacción Académica y Consenso:** Este módulo promueve la colaboración entre investigadores mediante la creación de grupos, votaciones y talleres de consenso. El sistema RAG puede extenderse también como asistente dentro de estos espacios, sugiriendo artículos relevantes para debates, generando resúmenes temáticos o respondiendo preguntas específicas durante las sesiones de discusión académica.

### III. OBJETIVOS

#### Objetivo General:

Desarrollar e implementar un sistema RAG que mejore el desempeño del buscador de la plataforma Centinela, permitiendo recuperar información científica relevante y generar respuestas automáticas de valor para el usuario.

#### Objetivos Específicos:

- Realizar una revisión sistemática de la literatura sobre metodologías y/o frameworks para la implementación de RAG.
- Diseñar e implementar la arquitectura técnica del sistema RAG utilizando modelos de recuperación y generación de texto.

- Evaluar el sistema RAG desarrollado en comparación con el buscador tradicional de Centinela mediante métricas estándar y pruebas de usuario.

#### IV. ALCANCE DEL COMPONENTE

Este componente del proyecto se enfocará al desarrollo e integración de un sistema RAG en el módulo de búsqueda de Centinela en los artículos relevantes. Para llevar a cabo, se usará umbrellar slr, es un tipo de revisión sistemática que sintetiza evidencia secundaria, es decir, revisiones sistemáticas y metaanálisis ya publicados, y no estudios primarios, y prisma aporta un marco estandarizado y ampliamente validado para el desarrollo del protocolo. Este enfoque combinado utiliza 2 fases, en la primera se actúa principalmente donde permite definir de forma clara y anticipada todos los componentes del protocolo necesarios para ejecutar la revisión y la segunda fase se ejecuta la búsqueda definida, se seleccionan las revisiones sistemáticas relevantes, se evalúa su calidad metodológica con las herramientas del JBI.

Por otra parte, se usa DSR para crear una solución al problema que enfrenta Centinela y, al mismo tiempo, generar conocimiento útil y aplicable sobre cómo diseñar estas soluciones.

Las etapas que se aplicarán en el presente trabajo son las siguientes:

- Definir los objetivos para la solución
- Diseño y desarrollo del artefacto
- Demostración del uso del artefacto para resolver el problema
- Evaluación del desempeño del artefacto
- Comunicación de los resultados

#### V. TAREAS ESPECÍFICAS

Con base en las actividades establecidas en el proyecto y asignadas al componente, se plantea las tareas que permiten cumplir con los objetivos del proyecto, y que, posteriormente, permiten la evaluación de su cumplimiento a través de los resultados esperados.

Semana referencial / Etapas	Fecha inicio-fin (si aplica)	Tareas específicas	Resultado esperado (si aplica)
1		Familiarización con Centinela V2: arquitectura y estado actual	
2		Despliegue de versión de pruebas del sistema	
3		Revisión de literatura	Documento de tipo SLR
4		Levantamiento de requisitos específicos del componente	Detalle de requisitos
5		Investigación de herramientas y modelos aplicables a la arquitectura RAG.	
6		Desarrollo del módulo de recuperación semántica (Retriever).	
7		Implementación del módulo de enriquecimiento y expansión de consultas (Augmentation).	
8		Integración de un modelo generativo de texto (Generator).	
9		Diseño y ejecución de pruebas de evaluación con métricas	
10		Difusión de proyecto y resultados	Reporte detallado documentando las recomendaciones específicas para mejorar la seguridad del sistema.

\*Declaración de uso de IA: Se ha utilizado ChatGPT para revisión ortográfica y edición en cuanto a descripción de contenidos.

#### VI. BIBLIOGRAFÍA

- [1] Peffers, K., Tuunanen, T., Rothenberger, M. A., y Chatterjee, S. A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3):45–77, 2008.
- [2] L. Recalde, G. Suntaxi, D. Martinez-Mosquera, R. Masabanda, and D. Cabrera, "Centinela: An Intelligent System Based on an Integrated Architecture for Supporting Scholars," in *Proc. Conf. Information and Communication Technologies of Ecuador*, Cham, Switzerland: Springer Nature Switzerland, Oct. 2024, pp. 160–177.
- [3] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—a systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, 2009. Papatheodorou, S. Umbrella reviews: what they are and why we need them. *European Journal of Epidemiology*, 34(6):543–546, 2019.

#### VII. FIRMAS DE RESPONSABILIDAD

Desarrollado por:	Alejandro Sebastian Chavez Vega
Aprobado por:	Dra. Gabriela Suntaxi