

VICERRECTORADO DE DOCENCIA

PLAN DE TRABAJO DE INTEGRACIÓN CURRICULAR

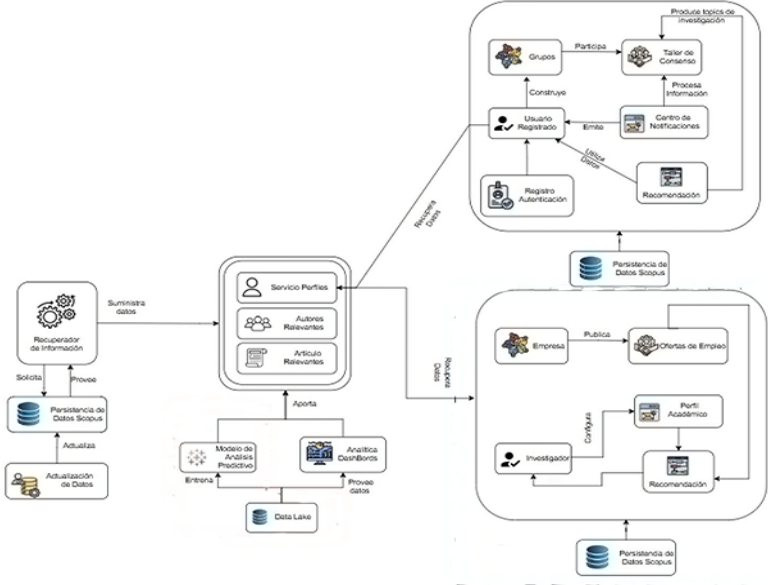
I. INFORMACIÓN BÁSICA	
Unidad Académica:	Facultad de Ingeniería de Sistemas
Carrera:	Ingeniería en ciencias de la computación
Proyecto:	Proyecto Centinela Avanzado V2: Búsqueda Optimizada y Fortalecimiento en Colaboración Académica
Componente:	Gestión y Personalización de Feeds Académicos
Línea de investigación:	Ciencias de la Computación e Inteligencia Artificial - Tópico: Inteligencia artificial Ingeniería de Software - Tópico: Motores de Búsqueda y Recuperación de Información
Nombres y apellidos del estudiante:	Alejandro Sebastian Chavez Vega
Nombres y apellidos del Profesor:	Ing. Gabriela Suntaxi, Ph.D.
II. DESCRIPCIÓN DEL COMPONENTE	
<p>Con la popularización de los LLM, los sistemas de recuperación de información se adaptaron a búsquedas con lenguaje natural y a tener un contexto semántico de lo que requiere el usuario. Esto se ajusta principalmente a lo que necesita Centinela [1]. Centinela es una plataforma que busca fortalecer la investigación científica en Ecuador. Su objetivo principal es mejorar el acceso a la información científica y fomentar la colaboración entre investigadores. Que es rag Incorporar RAG en Centinela transformará la plataforma en un agente capaz de interpretar el contexto del usuario y sus intereses, para generar respuestas coherentes y útiles. A diferencia del buscador tradicional actual, que simplemente devuelve una lista de artículos, RAG primero recuperará los documentos más relevantes y luego generará una respuesta en lenguaje natural, basada en los artículos académicos almacenados en las bases de datos del sistema. Esto permitirá que el investigador tenga una síntesis clara, incluso si no se usan términos técnicos exactos. El sistema extraerá un resumen de los resultados obtenidos y los autores, todo a partir de los artículos disponibles.</p> 	

Figura 1. Arquitectura del sistema Centinela.

Como se observa en la Figura 1, la plataforma centinela 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 mediante 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 sistemas 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 centrará en el desarrollo e integración de un sistema RAG (Retrieval-Augmented Generation) en el módulo de búsqueda de Centinela, con el objetivo de recuperar artículos relevantes de manera precisa y contextualizada.

Para ello, se empleará la metodología Umbrella Systematic Literature Review (Umbrella SLR) [2], un tipo de revisión sistemática que no aborda directamente estudios primarios, sino que se enfoca en la síntesis de evidencia secundaria, es decir, revisiones sistemáticas y metaanálisis ya publicados. Este enfoque permite construir una visión consolidada, rigurosa y de alto nivel sobre un área específica del conocimiento, lo que facilita la identificación de patrones comunes, vacíos temáticos y recomendaciones fundamentadas en múltiples fuentes confiables.

Como estrategia de búsqueda principal dentro de Umbrella SLR, se utilizará el enfoque Back-and-Forward Citation Propagating [3][4], una metodología que permite la construcción dinámica de conocimiento a través de la exploración de la red de citaciones entre los estudios. Este enfoque contempla dos procesos complementarios:

- **Propagación hacia atrás (backward citation):** consiste en revisar las referencias citadas por cada revisión seleccionada, con el fin de identificar trabajos previos relevantes que podrían no haber sido detectados en la búsqueda inicial.
- **Propagación hacia adelante (forward citation):** permite localizar investigaciones más recientes que hayan citado los trabajos clave, enriqueciendo y actualizando el corpus de evidencia.

El desarrollo de esta revisión se organizará en dos fases principales:

- Fase de planificación, en la que se establecen de forma clara y anticipada los elementos que componen el protocolo de revisión, como los criterios de selección, las fuentes de información y las herramientas de evaluación.
- Fase de ejecución, en la que se lleva a cabo la búsqueda de literatura, la selección de revisiones sistemáticas relevantes y la evaluación de su calidad metodológica utilizando los instrumentos del Joanna Briggs Institute (JBI).

Al mismo tiempo, se aplicará la metodología Design Science Research (DSR) [5] para generar conocimiento práctico y generalizable sobre el desarrollo de este tipo de sistemas. Las etapas de la metodología DSR que se abordarán en este trabajo son:

- Definición de los objetivos de 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 y Deployment de versión de pruebas del sistema	Versión de pruebas (funcional) del sistema
2		Revisión de literatura	
3		Revisión de literatura + Investigación de herramientas y modelos aplicables a la arquitectura RAG.	Resumen de la revisión de la literatura
4		Levantamiento de requisitos específicos del componente	Detalle de requisitos
5		Desarrollo del módulo de recuperación semántica (Retriever).	
6		Desarrollo del módulo de recuperación semántica (Retriever).	
7		Implementación del módulo de enriquecimiento y expansión de consultas (Augmentation).	

8		Implementación del módulo de enriquecimiento y expansión de consultas (Augmentation).	
9		Integración de un modelo generativo de texto (Generator).	
10		Integración de un modelo generativo de texto (Generator).	
11		Componente integrado en Centinela	
12		Difusión de proyecto y resultados	Reporte detallado documentando las recomendaciones específicas para mejorar la calidad de la búsqueda.
13		Revisión del Trabajo de Integración Curricular por parte de los profesores asignados	
14		Revisión del Trabajo de Integración Curricular por parte de los profesores asignados	
15		Revisión del Trabajo de Integración Curricular por parte de los profesores asignados	
16		Ajustes finales y presentación del Trabajo de Integración Curricular.	Trabajo de Integración Curricular

NOTAS:

- El estudiante, en la fecha establecida en el Calendario Académico (aproximadamente cuatro semanas antes del cierre del SAI), remite el documento final y/o el producto al Director del TIC.
- El Director, en la fecha establecida en el Calendario Académico, remite el documento final y/o el producto para revisión, al Decano o Director de la ESFOT.
- Los revisores remiten la calificación al Decano o Director de la ESFOT con copia al Director del TIC, en la fecha establecida en el Calendario Académico. Los revisores no podrán solicitar correcciones.

VI. BIBLIOGRAFÍA

[1]L. Recalde, G. Suntaxi, D. Martinez-Mosquera, R. Masabanda, & D. Cabrera. (2024). Centinela: An Intelligent System Based on an Integrated Architecture for Supporting Scholars. En Proc. Conf. Information and Communication Technologies of Ecuador (pp. 160–177). Cham, Switzerland: Springer Nature Switzerland.

[2]Papatheodorou, S. (2019). Umbrella reviews: What they are and why we need them [Commentary]. European Journal of Epidemiology, 34(6), 543–546. <https://doi.org/10.1007/s10654-019-00505-6>

[3]Aristodemou, L., & Tietze, F. (2018). Citations as a measure of technological impact: A review of forward citation-based measures. World Patent Information, 53, 39–44. <https://doi.org/10.1016/j.wpi.2018.05.001>

[4]Hu, X., Rousseau, R., & Chen, J. (2011). On the definition of forward and backward citation generations. Journal of Informetrics, 5(1), 27–36. <https://doi.org/10.1016/j.joi.2010.07.004>

[5]Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2008). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77.

VII. FIRMAS DE RESPONSABILIDAD	
Desarrollado por:	Alejandro Sebastian Chavez Vega
Aprobado por:	Ing. Gabriela Suntaxi, Ph.D.