

UNIVERSIDAD NACIONAL DE INGENIERIA

FACULTAD DE CIENCIAS

ESCUELA PROFESIONAL DE CIENCIA DE LA COMPUTACIÓN



Asistente académico con RAG

Curso: Inteligencia Artificial

Profesor: Marcos Antonio Alania Vicente

Integrantes	Código
Iman Noriega Melissa	20224041G
Trujillo Serva Luis Andre	20220428D
Orrego Torrejon Diego A.	20204161G
Méndez Gonzalo Miguel	20220264A
Pineda García Diego	20222117F

Fecha de entrega: 10/09/2025

Índice

1. Descripción inicial del problema	3
2. Objetivos preliminares	3
2.1. Objetivo general	3
2.2. Objetivos específicos	3
3. Identificación de técnicas de IA a utilizar	4

1. Descripción inicial del problema

En el ámbito académico, los estudiantes suelen enfrentarse a la necesidad de consultar grandes volúmenes de material (apuntes, artículos, manuales o presentaciones) para resolver dudas puntuales durante su proceso de aprendizaje. Este proceso de búsqueda suele ser lento e ineficiente, especialmente en momentos de alta demanda, como preparación para exámenes o desarrollo de trabajos.

Un **asistente académico con RAG (Retrieval-Augmented Generation)** permitiría responder preguntas frecuentes en lenguaje natural, recuperando información relevante desde documentos académicos y generando respuestas fundamentadas. De esta manera, se mejora la accesibilidad a la información y se promueve un aprendizaje autónomo, dinámico y apoyado por Inteligencia Artificial.

2. Objetivos preliminares

2.1. Objetivo general

Desarrollar un asistente académico basado en RAG que responda preguntas de estudiantes a partir de materiales de estudio predefinidos, combinando recuperación de información con generación de lenguaje natural.

2.2. Objetivos específicos

1. Implementar un sistema de procesamiento de lenguaje natural (NLP) para interpretar consultas de los estudiantes en lenguaje natural.
2. Integrar un mecanismo de recuperación semántica (buscador vectorial) que permita localizar fragmentos relevantes en documentos académicos.
3. Desarrollar un modelo de generación aumentada (RAG) que combine la información recuperada con un modelo de lenguaje para producir respuestas claras y contextualizadas.
4. Evaluar el rendimiento del asistente mediante métricas de recuperación (precisión, recall) y calidad de generación (coherencia, relevancia).

5. Analizar las implicaciones éticas del sistema, considerando riesgos como la generación de respuestas incorrectas y la necesidad de transparencia en las limitaciones del asistente.

3. Identificación de técnicas de IA a utilizar

- **Procesamiento de Lenguaje Natural (NLP):** Para interpretar las consultas y realizar preprocesamiento de texto (tokenización, embeddings).
- **Búsqueda vectorial y embeddings semánticos:** Para representar textos y realizar recuperación de información en una base de conocimientos.
- **Generación Aumentada por Recuperación (RAG):** Combinar la información recuperada con un modelo generativo para construir respuestas fundamentadas en los materiales académicos.
- **Redes Neuronales (transformers):** Para la generación de embeddings y la construcción de respuestas en lenguaje natural.
- **Evaluación con métricas de recuperación y generación:** Precisión, recall, F1-score, y métricas de calidad de texto como BLEU o ROUGE.