

TUIA NLP - TP2 2023

Read time: 2 minutes

TUIA NLP 2023 TRABAJO PRÁCTICO 2

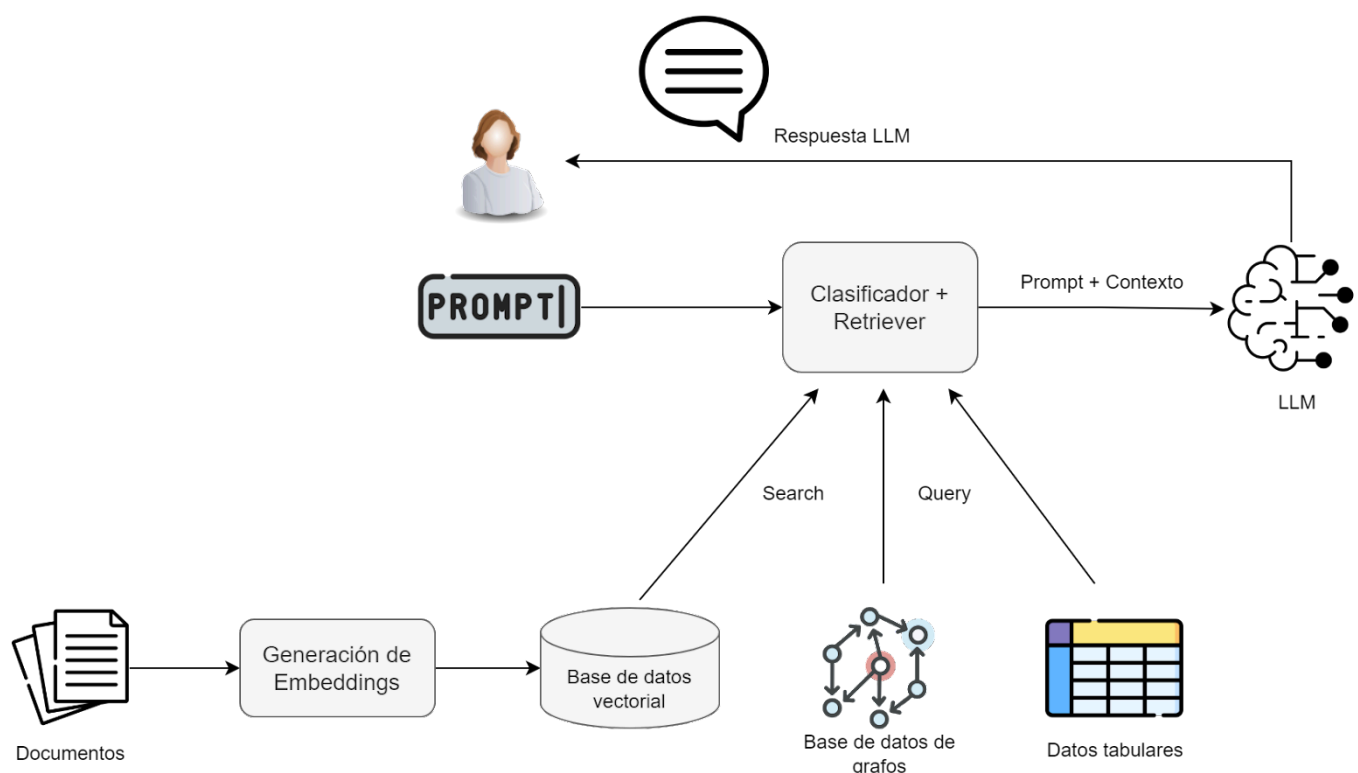
Pautas generales:

- El trabajo deberá ser realizado individualmente.
- Deberá informar cuál es la url del repositorio con el que van a trabajar y las definiciones de en qué problemáticas quisieran solucionar con un sistema multiagente en el siguiente formulario:
https://docs.google.com/forms/d/e/1FAIpQLSdOZNGOzQ1gbf43caA4ygAbLx5tm5bU-s8RdfdtOzd_aXzhA/viewform?usp=pp_url
- Se debe entregar un informe en el cual se incluya las justificaciones y un vínculo a los archivos que permitan reproducir el proyecto. No se acepta solamente código.
- Temas deseables a cubrir en el tp
 - Recuperación de datos de bases de datos de grafos
 - Extracción de conocimiento de texto y posterior inserción en una base de datos de grafos
 - Agentes (estará cubierto en el Ejercicio 2)

Ejercicio 1 - RAG

Crear un chatbot experto en un tema a elección, usando la técnica RAG (Retrieval Augmented Generation). Como fuentes de conocimiento se utilizarán al menos las siguientes fuentes:

- Documentos de texto
- Datos numéricos en formato tabular (por ej., Dataframes, CSV, sqlite, etc.)
- Base de datos de grafos (Online o local)



El sistema debe poder llevar a cabo una conversación en lenguaje español. El usuario podrá hacer preguntas, que el chatbot intentará responder a partir de datos de algunas de sus fuentes. El asistente debe poder clasificar las preguntas, para saber qué fuentes de datos utilizar como contexto para generar una respuesta.

Requerimientos generales

- Realizar todo el proyecto en un entorno Google Colab
- El conjunto de datos debe tener al menos 100 páginas de texto y un mínimo de 3 documentos.
- Realizar split de textos usando Langchain (RecursiveTextSearch, u otros métodos disponibles). Limpiar el texto según sea conveniente.
- Realizar los embeddings que permitan vectorizar el texto y almacenarlo en una base de datos ChromaDB
- Los modelos de embeddings y LLM para generación de texto son a elección
- Para el desarrollo del “Clasificador” es posible utilizar diversas técnicas aprendidas en la materia, por ejemplo en Unidad 3 y Unidad 6

Ejercicio 2 - Agentes

Realice una investigación respecto al estado del arte de las aplicaciones actuales de agentes inteligentes usando modelos LLM libres.

Plantee una problemática a solucionar con un sistema multiagente. Defina cada uno de los agentes involucrados en la tarea. Es importante destacar con ejemplos de conversación, la interacción entre los agentes.

Realice un informe con los resultados de la investigación y con el esquema del sistema multiagente, no olvide incluir fuentes de información.

Opcional: Resolución con código de dicho escenario.