



UNIVERSIDAD SAN FRANCISCO

Nombres:

Cesar Carrera 00344613

Edgar Guzmán 00344822

Juan Diego Sánchez 00344455

Byron Vinueza 00345908

Proyecto Final

Agente RAG: Funcionamiento y Pruebas

Grupo 4

Introducción

- Agente RAG (Retrieval-Augmented Generation):
 - Combina recuperación de información y generación de texto.
 - Utiliza modelos de lenguaje como GPT-4 para responder preguntas basadas en documentos.

- **Objetivo:**

- Responder preguntas utilizando información contenida en archivos PDF.

Flujo de Trabajo del Agente

1. Carga de documentos PDF:

- Los archivos PDF se cargan desde un directorio específico.
- Se procesan utilizando PyPDFLoader.

2. División del texto:

- Los documentos se dividen en fragmentos (chunks) utilizando RecursiveCharacterTextSplitter.
- Esto facilita la indexación y recuperación eficiente.

3. Creación de embeddings:

- Se generan representaciones vectoriales (embeddings) de los fragmentos usando OpenAIEmbeddings.

4. Indexación en Chroma:

- Los embeddings se almacenan en una base de datos vectorial (Chroma).
- Esto permite realizar búsquedas rápidas basadas en similitud.

5. Recuperación y generación:

- El agente utiliza un modelo de lenguaje (GPT-4) para responder preguntas.
- Recupera información relevante de la base de datos y genera respuestas.

Repartición de Actividades:

César Carrera:

- Investigación de diferentes formas de crear el algoritmo
- Investigación de librerías a usar
- Recopilación de archivos PDF
- Lectura de librerías

Byron Vinuesa:

- Creación de la lógica del código
- Uso del API Key
- Inicio de lógica para lectura de directorio de trabajo

Juan Diego Sánchez:

- Investigación acerca del uso de Open AI y su API Key
- Subida de archivos PDF a un repositorio
- Finalización de la lógica para lectura de PDFs

Edgar Guzmán:

- Recopilación de preguntas y preguntas de prueba finales
- Investigación de funcionalidad y aplicabilidad de las librerías a usar
- Creación de flujo de proceso para el funcionamiento del código