Multi-PDF Chatbot

Ramarao Kadiyala - Mentor

Reethu RG Thota

S R Monish Raj

Shreyas Rajiv

Sruthika Sivakumar

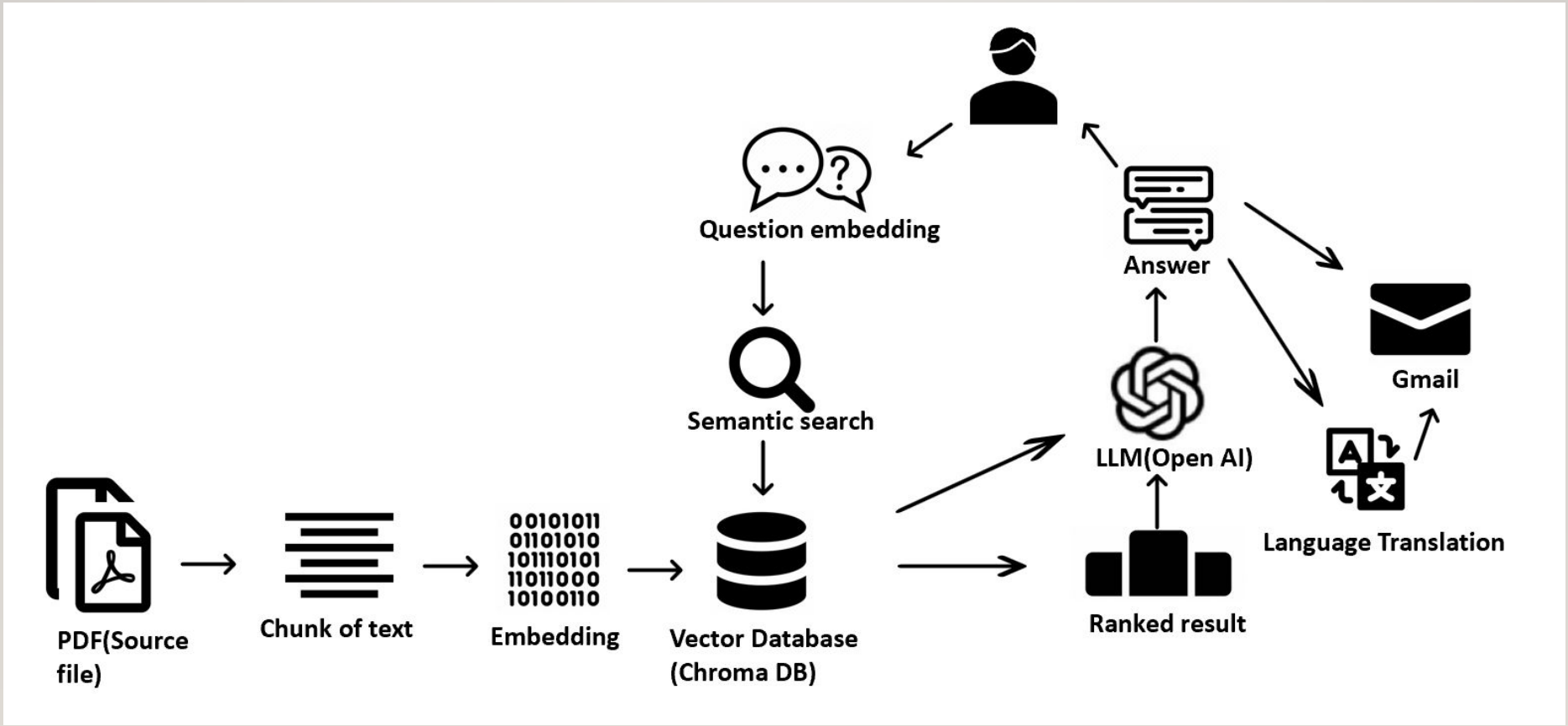
PROJECT INTRODUCTION

- This project is a comprehensive Python application that combines text extraction, document analysis, language translation, and email automation.
- The system is designed to facilitate efficient handling of multiple PDF documents, enabling users to upload files, extract text by posing targeted questions, and receive precise language translations along with email capabilities.
- This application is valuable for professionals, researchers, and students seeking to expedite document analysis, perform language translation, and sharing information within their respective domains.

PROJECT FEATURES

- **Efficient Document Understanding**
  The application streamlines document comprehension through advanced text extraction, vector representations, and a responsive chatbot, enabling users to interactively explore and gain insights from complex PDF files.
- **Multilingual Access and Collaboration**
  Overcoming language barriers, the tool facilitates precise language translations within the document, promoting collaboration among users with different language preferences and enhancing accessibility to a wider audience.
- **Workflow Optimization with Email Automation**
  The integration of GMail API automates the email process, allowing users to seamlessly share valuable document excerpts or insights with colleagues. This feature optimizes communication workflows and enhances overall productivity.

PROJECT ARCHITECTURE



TOOLS AND TECHNOLOGIES USED

- Streamlit - Frontend user interface
- Langchain -
  - CharacterTextSplitter
  - OpenAIEmbeddings
  - Load_qa_chain
  - PromptTemplate
- OpenAI LLM - Used for Retrieval-Augmented Generation (RAG)
- Chroma DB - Vector Database
- GMail API - Integration and authentication with Google Cloud Platform (GCP)
- PyPDF2 - Retrieve text and metadata from PDF

DEMO AND CONCLUSION