DocuQuery: Intelligent Document Q&A

# Overview

DocuQuery is a powerful tool designed to facilitate intelligent question-answering from various document formats. Whether you have a PDF, TXT file, or even a URL link, DocuQuery allows you to upload the content and seamlessly ask questions related to the document's context.

# Features

## Document Upload

- Supported Formats: PDF, TXT files, or URL links.  
- Easy Access: Upload documents directly through the Streamlit user interface.

## Question-Answering

- Advanced NLP: Utilizes advanced natural language processing to provide accurate answers based on the document's content.

## OpenAI Integration

- Llama 3: Employs Llama 3, an open-source model with OpenAI embeddings, for enhanced language understanding.

## Faiss Database

- Efficient Retrieval: Implements Faiss for efficient storage and retrieval of document embeddings.

## Streamlit UI

- User-Friendly Interface: Powered by Streamlit for smooth navigation and interaction.

## Tracking and Monitoring

- LangSmith Integration: Integration with LangSmith for comprehensive tracking and monitoring functionalities.

## LangChain Integration

- Enhanced Processing: Utilizes LangChain document loader, text splitter chains, etc., for improved document processing.

# Installation

## Prerequisites

Ensure you have Python installed on your system.

## Steps

1. Clone the Repository:  
 ```bash  
 git clone https://github.com/your\_username/DocuQuery-Intelligent-Document-Q-A.git  
 ```  
2. Navigate to the Project Directory:  
 ```bash  
 cd DocuQuery  
 ```  
3. Install the Required Dependencies:  
 ```bash  
 pip install -r requirements.txt  
 ```

# Usage

## Running the Application

1. Start the Streamlit App:  
 streamlit run app.py  
2. Access the Application:  
 - Open the provided URL in your browser.  
3. Upload Your Document:  
 - Upload a PDF, TXT file, or provide a URL link.  
4. Ask Questions:  
 - Ask questions related to the content of the document.  
5. Receive Answers:  
 - Receive intelligent answers generated by DocuQuery.

# Detailed Components

## 1. Document Upload

DocuQuery allows users to upload various document formats (PDF, TXT, or URL) through a simple and intuitive interface. The documents are then processed and prepared for question-answering.

## 2. Question-Answering

Leveraging advanced natural language processing techniques, DocuQuery can understand and answer questions based on the uploaded document's content. The integration of Llama 3 and OpenAI embeddings ensures high accuracy and contextual understanding.

## 3. OpenAI Integration

By employing Llama 3 with OpenAI embeddings, DocuQuery enhances its language understanding capabilities, making it more effective in comprehending and responding to user queries.

## 4. Faiss Database

Faiss, a library for efficient similarity search and clustering of dense vectors, is utilized to store and retrieve document embeddings efficiently. This ensures fast and accurate retrieval of information.

## 5. Streamlit UI

The application features a user-friendly interface powered by Streamlit. This interface allows for easy document uploads, question submissions, and answer displays, providing a smooth user experience.

## 6. Tracking and Monitoring

DocuQuery integrates with LangSmith for robust tracking and monitoring functionalities. This helps in keeping track of document processing and query handling, ensuring reliability and efficiency.

## 7. LangChain Integration

LangChain is employed for enhanced document processing. It includes features like document loader and text splitter chains, which facilitate better handling and processing of uploaded documents.

# Conclusion

DocuQuery is a comprehensive tool designed to streamline the process of intelligent question-answering from various document formats. With its advanced features and integrations, it provides an efficient and user-friendly solution for accessing information from documents.  
  
For more details and to get started, visit the [DocuQuery GitHub repository](https://github.com/Vansh3503/DocuQuery-Intelligent-Document-Q-A).