Project Title: Intelligent Document Finder with Llama Index

Overview

The Intelligent Document Finder project aims to create a seamless, user-friendly platform that allows for the uploading and automatic indexing of various document formats, including PDFs, PPTs, Word documents, and other forms of unstructured data. By leveraging Llama Index for data indexing and retrieval, the system will enable precise query handling through a front-end application, enhancing the user's ability to find specific information efficiently.

Objectives

User Data Upload:

- Enable users to upload documents of various formats (PDF, PPT, Word, etc.) to a designated Google Drive folder.
- Support for multiple file formats allows for a broad range of document types to be indexed and searched.

Automated Data Storage and Indexing:

- Automate the process of transferring uploaded documents from Google Drive to a server environment where they can be processed and indexed.
- Use Llama Index to create a searchable database of the uploaded documents, incorporating detailed metadata for each file, such as titles, paragraph numbers, page numbers, etc.

Development of a Query Interface:

- Develop a Streamlit application or a similar frontend-based system that interfaces with the indexed data.
- Allow users to perform searches with the indexed data, returning relevant document snippets along with comprehensive metadata.

Deliverables

1. **Code Repository:**
   - All project code, including Google Drive integration scripts, server-side document processing and indexing scripts, and the Streamlit application, should be submitted to a GitHub repository. And the link to that git hub will be submitted to AI Trainer.

2. **Documentation:**
   - Provide comprehensive documentation covering setup instructions, how to use the system, and an architectural overview of the project components.(a flowchart)
   - Include detailed comments within the code to explain the functionality of critical sections.

3. **Demonstration Video:**
   - Create a short video demonstrating the system's capabilities, from uploading documents to querying and retrieving information.