Lab-6 Meeting Notes

Team Details

Team Name: DPS

Name	USC ID
Dolly Rijwani	9040345530
Prathmesh Lonkar	3661203370
Shreyan Yoge	9655139013

Meeting Report Day 1

Date: 10/15/2024

Overview:

- Reviewed the assignment's objectives.
- Discussed the procedure for pdf text extraction and the tools needed including Python, Streamlit, and OpenAI APIs.

System Design:

- Planned the chatbot model structure, which involves reading textbook PDFs, converting them into embeddings, and storing these in a vector database.
- Decided to use Streamlit for the web interface to facilitate user interaction with the chatbot.

Technical Setup:

- Discussed the initial setup of the development environment and the necessary libraries as detailed in the Drive folder Readme.md.

Action Items:

- Tasked with downloading necessary files and setting up the initial project structure.
- Scheduled next meeting to begin the implementation of text extraction from PDFs.

Meeting Report Day 2

Date: 10/16/2024

Progress Update:

- Confirmed setup completion and successful environment configuration for all team members.
- Shared findings on different embedding models and discussed their applicability.

Data Handling:

- Implemented the function for extracting text from PDFs using the PyPDF2 library.
- Discussed the method for splitting the extracted text into manageable chunks using LangChain's CharacterTextSplitter.

Action Items:

- Begin embedding the text chunks using the OpenAI API.
- Plan to integrate these embeddings into a vector database.

Meeting Report Day 3

Date: 10/17/2024

Progress Update:

- Reviewed the successful creation of text chunks.
- Demonstrated initial embeddings of text chunks and discussed optimization possibilities.

Database Integration:

- Discussed the setup of the vector database for storing embeddings.
- Implemented initial functionality for querying the vector database with embeddings generated from user questions.

Action Items:

- Finalize the retrieval logic for finding relevant text chunks based on query embeddings.
- Start building the conversation chain using LangChain's tools.

Meeting Report Day 4

Date: 10/18/2024

Integration Review:

- Tested integration of the retrieval system with the conversational model to ensure that user queries return relevant document sections.
- Discussed the handling of conversation context and history to improve response accuracy and relevance.

Interface Development:

- Implemented the basic layout for the web interface using Streamlit, allowing PDF uploads and user interaction through a chat window.

Action Items:

- Enhance the chat interface to better display conversations and manage user interactions.
- Prepare for comprehensive testing of the chatbot's functionality.

Meeting Report Day 5

Date: 10/19/2024

Final Testing:

- Testing the entire system.
- Debugged issues related to PDF processing and response generation.

Documentation Completion:

- Compiled and finalized all code documentation, including setup and execution instructions.
- Reviewed the entire project documentation for clarity and completeness.

Project Closure:

- Ensured all assignment requirements were met, including the chatbot's ability to process PDF text, handle user queries, and generate accurate responses.
- Prepared for project submission, including a demonstration video outlining the system's workflow and features.

.