

LAB 6 PART 1

Table of Contents

| | |
|--------------------------------|----------|
| Meeting Notes for Day 1 | 2 |
| Meeting Notes for Day 2 | 3 |
| Meeting Notes for Day 3 | 4 |
| Meeting Notes for Day 4 | 5 |
| Meeting Notes for Day 5 | 6 |

Team Details

Team Name: SSK

| Name | USC ID |
|-------------------------|------------|
| Shreyansh Baredia | 3739756887 |
| Soma Meghana Prathipati | 4226812081 |
| Kayvan Shah | 1106650685 |

Meeting Notes for Day 1

Date: 03/05/24

Agenda:

- Review of Assignment Requirements
- Discussion on Tools and Libraries
- Assignment Task Delegation
- Setting Up Weekly Meeting Schedule

Minutes:

- Reviewed the assignment requirements, focusing on the tasks related to PDF text extraction, text chunking, embeddings, and the use of OpenAI for conversation chains.
- Discussed the tools and libraries required for the assignment, emphasizing the use of Python scripts and tools in the Linux environment.
- Delegated specific tasks to each team member based on the assignment components.
- Agreed upon meeting virtually each day of the week to discuss progress and next steps.
- Acknowledged the presence of a Readme.md file in the Drive folder with setup instructions and a list of required libraries for Parts 1 and 2 of Lab 6.
- Confirmed the use of the skeleton file App_p1.py for the lab and the need to add actual code into the file.

Action Items:

- Familiarize yourself with the setup instructions in the Readme.md file and ensure all required libraries are installed.
- Begin drafting Python scripts for PDF text extraction, text chunking, embedding creation, and conversation chain setup.
- Collaborate on understanding the parameters of the CharacterTextSplitter function provided by langChain for efficient text chunking.
- Start working on the function to create word embeddings using OpenAI Embeddings and store them in a vector database.
- Explore the ConversationMemoryBuffer function provided by langchain for creating a conversation chain based on an LLM model and the vector store.
- Develop a driver function that integrates the above functions and interacts with the user, taking questions, and providing responses.

Meeting Notes for Day 2

Date: 03/06/2024

Agenda:

- Progress Update
- Challenges Faced
- Discussion on Python Scripts
- Coordination on Database Setup

Minutes:

- Each team member provided updates on their progress with PDF text extraction, text chunking, and embedding creation.
- Discussed challenges faced during the implementation of Python scripts, including any errors or difficulties encountered.
- Collaborated on resolving challenges, shared insights on efficient script development, and discussed potential optimizations.
- Coordinated on the setup of the vector database, ensuring integration with the embedding creation script.

Action Items:

- Continue refining Python scripts, addressing any challenges faced.
- Collaborate on improving script efficiency and functionality.
- Verify the successful integration of the vector database with the embedding creation script.
- Research and implement techniques for optimizing text chunking.

Meeting Notes for Day 3

Date: 03/07/2024:

Agenda:

- Progress Review
- Data Processing Updates
- Vector Datastore Function Discussion
- Conversation Chain Implementation

Minutes:

- Reviewed the progress made on PDF text extraction, text chunking, and embedding creation.
- Discussed updates on data processing techniques, especially efficient text chunking using the CharacterTextSplitter.
- Discussed the implementation of the vector datastore function, ensuring proper creation and storage of word embeddings.
- Collaborated on implementing the conversation chain using the ConversationMemoryBuffer function for understanding user questions and generating responses.

Action Items:

- Finalize the vector datastore function, ensuring successful integration with the embedding creation script.
- Optimize the text chunking process based on insights gained during the discussion.
- Begin implementing the conversation chain, testing it with sample queries to ensure proper functionality.

Meeting Notes for Day 4

Date: 03/08/2024:

Agenda:

- Database Setup Progress
- Text Chunking Optimization
- Discussion on Conversation Chain Testing
- Assignment Documentation Review

Minutes:

- Discussed the progress made on the database setup, addressing any issues encountered and ensuring seamless integration with the project.
- Shared insights and strategies for optimizing the text chunking process, ensuring efficiency and accuracy.
- Reviewed the conversation chain implementation, discussed testing methodologies, and identified potential improvements.
- Discussed the requirements for assignment documentation, including setup steps, code documentation, and any additional notes or insights gained during the process.

Action Items:

- Complete the database setup, ensuring its smooth integration with the entire project pipeline.
- Continue optimizing text chunking based on the discussed strategies.
- Conduct thorough testing of the conversation chain, addressing any issues or improvements.
- Start compiling assignment documentation, including setup steps, code documentation, and insights gained during the project.

Meeting Notes for Day 5

Date: 03/09/2024: Final Refinement and Documentation

Agenda:

- Final System Integration
- User Interaction Testing
- Assignment Documentation Finalization
- Project Presentation Preparation

Minutes:

- Discussed the final integration of all project components, including PDF text extraction, text chunking, embedding creation, vector database setup, and the conversation chain.
- Conducted user interaction testing, simulating various user queries to ensure the chatbot's responsiveness and accuracy in generating relevant responses.
- Finalized the assignment documentation, ensuring it included comprehensive setup steps, code documentation, and any additional notes or insights gained during the project.
- Initiated preparations for the project presentation, outlining key points, functionalities, and the overall structure to be covered.

Action Items:

- Perform a final review of the integrated system, addressing any last-minute issues or improvements.
- Continue user interaction testing, ensuring a smooth and intuitive experience for users.
- Complete and polish the assignment documentation, cross-referencing with the project code and addressing any gaps or inconsistencies.
- Prepare the necessary documentation and the code files for final submission.