Hackathon Final Round: Chat with Document

Qadeer Ahmad
Hope To Skill

Tuesday 16th January, 2024

1 Introduction

Welcome to "Chat with Document" – an innovative chatbot designed to revolutionize the way users interact with textual documents. This intelligent assistant specializes in processing docx and pdf files, offering a seamless experience for reading, querying, and extracting information. By harnessing the power of natural language processing, the chatbot facilitates effortless communication, allowing users to ask questions about their documents and receive comprehensive responses. With a focus on document comprehension and accessibility.

2 Objective

The objective of the "Chat with Document" chatbot is to provide users with a versatile and efficient tool for interacting with text-based documents, specifically in docx and pdf formats. The chatbot aims to offer a user-friendly interface for querying and extracting information from documents, enabling natural language interactions. Key functionalities include the ability to read and comprehend document content, answer user queries about the document, and highlight relevant sections to enhance document comprehension. By combining natural language processing with document analysis, the chatbot aims to streamline information retrieval, making it an invaluable resource for users seeking quick and precise insights from their textual documents.

3 Methodology

I used Langchaing with LLM from OpenAI Chatgpt 3.5. I used streamlit as web framework.

4 Results

I found some results

- Split by character and recursive split by character

  - Recursive split by character worked better than split by character. It did not exceed the context window of LLM because it splits text on basis of multiple character

- Embedding Models

  - **BAAI/bge-small-en-v1.5** is light weight and better than intfloat/e5-large-v2.

- Vector Store

  - **Qdrant** is easy to setup and fast.

5 Features

This chatbot has some amazing features.

- This can talk to Docx and PDF files.

- This has memory and retrieval capability.

- This chatbot gives you authentic information. No any fake resources.

6 Conclusion

This porject is exam based project but I have learn a lot of things in this project. Thanks to Sir Irfan and his hardworking team.

7 References

Xeven Solution | Hope To Skill | Hackathon Final Round | Batch 1