SANGYAN

The Research Cell
School Of Applied Sciences
HSNC University, Worli

Project Overview

- **Project Title: A ChatBot for the School of Applied Sciences**
- Date: 01/07/2024- 01/12/2024
- Prepared by: Shreay Kacheria, Vansh Jain, Hussain Ujjainvala, Vinay Gajula, Dakshit Makwana, Dimple Anjara, Moaviya Sayyed, Dhruv Sarkhedi, Kruti Madkikar

Introduction

This project involves creating an interactive document-based Q&A application named JEERA-BOT using Streamlit, LlamaIndex, and HuggingFace embeddings. The application enables users to query a knowledge base related to the School of Applied Sciences and receive intelligent, context-aware responses.

Objectives

- Simplify Knowledge Access: Allow users to query school-related documents interactively.
- Efficient Document Parsing: Automate document ingestion and indexing for fast retrieval.
- Resource Optimization: Use lightweight models to balance performance and computational costs.

Scope

- Load and process files (PDF, Excel) from a local directory.
- Embed and index documents for semantic search.
- Query documents using a streamlined interface powered by Streamlit.

SANGYAN

The Research Cell
School Of Applied Sciences
HSNC University, Worli

Deliverables

- JEERA-BOT Application:

Fully functional Streamlit web app with an intuitive interface.

- Query Engine:

Intelligent response system powered by LlamaIndex and LLMs.

- Document Indexing Pipeline:

Automated loading, parsing, and vector indexing of documents.

Timeline

- Week 1-4: Setting up the team, planning out the project structure, assessing the resources and softwares needed
- Week 5-8: Setup environment, configure LlamaIndex, and integrate document parsing.
- Week 9-12: Collect the data and integrate them to text files
- Week 13-15: Develop Streamlit interface and test user query responses.
- Week 16-18: Optimize embedding model and ensure caching for efficiency.
- Week 19-21: Conduct final tests and deploy the app.

Stakeholders

- Developers: Responsible for app functionality and system integration.
- End Users: Students and faculty from the School of Applied Sciences.

Appendix (Tools and Resources)

- Language Model: Groq-powered LlamaIndex.
- Embedding Model: HuggingFace sentence-transformers/all-MiniLM-L6-v2.
- Framework: Streamlit for interactive UI.
- Code Language: Python
- Environment: GitHub, Codespace