Scheme Research Application Report

Overview

At Haqdarshak, our dedicated research team meticulously sifts through various information portals for government schemes, extracting pertinent details based on standardized criteria. The goal of this assignment is to develop an automated Scheme Research Tool that will:

- Take the URL of a scheme article as input.
- Create an accurate and relevant summary.
- Enable users to ask questions based on the content of the article.
- The tool provides a summary covering four key criteria: Scheme Benefits, Scheme Application Process, Eligibility, and Documents required.

Features

1. Load URLs or upload text files containing URLs to fetch article content.
2. Process article content through LangChain's UnstructuredURL Loader.
3. Construct an embedding vector using OpenAI's embeddings and leverage FAISS, a powerful similarity search library, to enable swift and effective retrieval of relevant information.
4. Interact with the LLM's (ChatGPT) by inputting queries and receiving answers along with source URLs and their summaries.

Solution Structure

The solution consists of the following files:

1. main.py: The main Streamlit application Python script.
2. requirements.txt: A list of required packages for the project.
3. faiss_store_openai.pkl: A pickle file to store the FAISS index.
4. .config: Configuration file for storing your OpenAI API key.

Running the Application

Install the Required Packages within command prompt by running **pip install -r requirements.txt command.**

- Run the Streamlit Application using **streamlit run main.py** command.
- Once API loads, paste the URL in the box provided at left hand side bar.
- Click “Process URL” button then summary of the document will be shown on the web page.
- Type your question in the “Ask Question” box then click the button.

Conclusion

This Scheme Research Application automates the extraction and summarization of government scheme information from various URLs. It utilizes OpenAI's embedding models and FAISS for efficient similarity search, enabling users to interactively query the indexed data.