.Development of a Generative AI Chat Interface for Processing PDFs and Wikipedia Content

Content

Introduction

1. This project involves creating a web-based application that leverages Generative AI to enable users to interact through chat with information extracted from PDF documents or Wikipedia URLs.
2. The goal is to provide a seamless interface where users can input either a PDF file or a website URL from Wikipedia, and the system processes this input to generate a chat-based interaction model, allowing users to ask questions and receive responses based on the content of the PDF or Wikipedia page.

Objectives

- To develop a web application that accepts PDF files and Wikipedia URLs as input.
- To extract and process the content from these inputs using text extraction techniques for
- PDFs and web scraping for Wikipedia.
- To utilize Generative AI to create a chat interface that dynamically generates responses
- based on the processed content.
- To ensure the chat interface is intuitive, user-friendly, and capable of handling a variety of
- queries related to the input content.

Functional Requirements

Web Interface:

A clean and user-friendly web interface that allows users to upload PDF files or enter Wikipedia URLs.
A chat window where users can type in their queries and receive responses.
PDF Processing:

- Implement a PDF parser to extract text content from uploaded PDF files.
- Preprocess the text to make it suitable for the Generative AI model (e.g., removing headers and footers, correcting typos).

Wikipedia Scraping:

- Develop a web scraper to extract content from the provided Wikipedia URL.
- The scraper should be capable of handling different Wikipedia page structures and extracting the main content effectively.
- Extracted content should be preprocessed to remove irrelevant sections like references and citations for the chat model.

Generative AI Chat Model:

- Integrate a Generative AI model to process extracted text and generate chat responses.
- The model should be trained or fine-tuned to understand and generate human-like responses based on the context of the input content.
- Implement caching or other optimization techniques to improve response time and efficiency.

Agentic AI :

- Extend the system to support intelligent Q&A over uploaded PDF documents or web links, enabling dynamic interactions with the content.
- At the end of each Q&A interaction, the system should automatically fetch and append a relevant, high-quality link from the internet that allows users to explore the topic further.
- The external reference should be contextually aligned with the user's question and response, and clearly labeled to guide deeper exploration.

● You can use Langflow to create Agent flow