Automated Insights and Summarization Platform (AISP) - Streamlit Application Documentation

This document explains the Automated Insights and Summarization Platform (AISP), built using Streamlit to provide an interactive web interface for summarizing documents and answering questions. The app leverages transformer models and retrieval-augmented generation (RAG) to handle complex documents and provide accurate, user-friendly insights.

Overview of Key Features

1. File Upload and Document Processing: Users can upload PDF, DOCX, or TXT files for analysis. The function `load_pdf()` is responsible for extracting the document's content, which is formatted by `format_docs()` for downstream processing. This ensures the content is prepared effectively for summarization and Q&A.
2. Logo Embedding: A logo is embedded at the top-right of the interface using base64 encoding, giving the platform a professional and branded look, enhancing the user experience.
3. Streamlit Layout: The interface has three main tabs that organize the application's functionalities:
    - Tab 1 (Upload the File): Users upload the documents they wish to process.
    - Tab 2 (Document Summary): Summarizes the uploaded document, allowing users to quickly understand the key points without reading the entire text.
    - Tab 3 (Q/A): Users can ask questions about the document content, and the system will provide context-aware answers.
4. Document Summarization and Q&A: After uploading a document, a custom prompt (`custom_prompt`) is used to create a comprehensive yet positive response. The HuggingFace model (`all-MiniLM-L6-v2`) is utilized to generate embeddings, which represent the document content in a form suitable for processing. These embeddings are stored in Chroma, a vector store, to aid in efficient retrieval of relevant information. The RAG chain (`rag_chain`) combines the retriever and OllamaLLM model to generate responses tailored to user questions.
5. Summary and Question Answering:
    - Summary: In Tab 2, the platform generates a concise summary using `rag_chain.invoke()`. This helps users quickly grasp the essential insights of the document.
    - Q/A: In Tab 3, users can input specific questions, and the system provides relevant answers based on the context extracted from the document. This feature enhances the ability to interact with the content dynamically.

Running the Code

- Dependencies: Ensure all necessary libraries are installed, including Streamlit, LangChain, Chroma, HuggingFace, and others required for document processing and retrieval.
- Launching the Application: Run the script to access the application through your browser using the Streamlit URL. The interface is designed to be intuitive and interactive for seamless document processing.
- HuggingFace Token: Replace `"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"` with a valid HuggingFace token to ensure access to the required models and resources.

Summary

AISP is a user-friendly platform that helps summarize documents and answer questions based on uploaded files. By utilizing transformer models and a RAG architecture, AISP provides efficient, context-aware insights that support productivity and informed decision-making. With an easy-to-navigate interface, it simplifies the management of complex information, saving time and effort for users who need quick access to key document insights.