Software Requirements Specification

Problem Statement ID : 25231

Title : OmniSearch-Offline-Multimodal-RAG

1.Introduction

This project aims to develop an offline, multimodal intelligence platform that enables unified search and question answering across diverse data formats such as PDF documents, Word files, images, screenshots, and audio recordings. In many organizations, critical information is scattered across different formats, making manual searching slow and inefficient. Traditional search tools operate separately for text, images, and audio, leading to incomplete results and loss of context. The proposed system uses a Retrieval-Augmented Generation (RAG) approach combined with a local Large Language Model (LLM) to retrieve relevant information and generate accurate, grounded responses with transparent citations, all while operating completely offline.

2. Overall Description

The system consists of a local user interface, a multimodal data ingestion and processing pipeline, a shared vector database, and a local LLM-based response generator. Users can upload documents, images, and audio files into the system. The platform extracts textual content from documents, applies OCR on images, converts audio to text, and stores all extracted content as semantic embeddings in a unified index. When a user submits a natural-language query, the system retrieves the most relevant information across all modalities and generates a concise, well-grounded answer along with references to the original sources. This platform is designed to function entirely offline, ensuring data privacy and security.

3. Functional Requirements

- The system shall allow users to upload PDF and DOC/DOCX files.
- The system shall allow users to upload images and screenshots.
- The system shall extract text from PDF and DOC/DOCX files.
- The system shall perform OCR on images to extract textual information.
- The system shall generate semantic embeddings for all extracted content.
- The system shall store embeddings in a unified vector database.
- The system shall accept natural-language text queries from users.

- The system shall retrieve relevant text passages, images, and audio segments based on semantic similarity.
- The system shall generate answers using a local LLM grounded in retrieved content.
- The system shall include numbered citations for each generated answer.
- The system shall allow users to view original source documents, images, or audio segments linked to citations.

4. Non-Functional Requirements

- The system should operate entirely in offline mode.
- The system should provide fast and responsive query results.
- The system should ensure high accuracy and relevance of retrieved information.
- The system should protect all stored data and prevent external access.
- The system should be reliable and prevent loss of indexed data.
- The system should be scalable to support increasing volumes of multimodal data.
- The user interface should be simple, intuitive, and easy to use.
- The system should be modular to allow future enhancements.

5. External Interfaces

The system does not rely on cloud services during runtime.
All processing is performed locally using offline AI models and libraries.

- The user interface serves as the interaction point for querying and uploading data.
- The local file system acts as the storage interface for documents and indexes.
- Internal AI components (OCR engine, speech-to-text engine, embedding model, and LLM) interact through defined software interfaces within the system.

6. Conclusion

The OmniSearch – Offline Multimodal RAG Intelligence Platform provides a powerful and secure solution for unified information retrieval across heterogeneous data formats. By integrating multimodal semantic search with Retrieval-Augmented Generation and transparent citations, the system significantly reduces manual effort and improves information accessibility. Its offline-first design ensures data privacy, reliability, and suitability for sensitive or restricted environments, making it a robust and scalable intelligence platform.