Proposal for Multimodal Retrieval-Augmented Generation (RAG) System

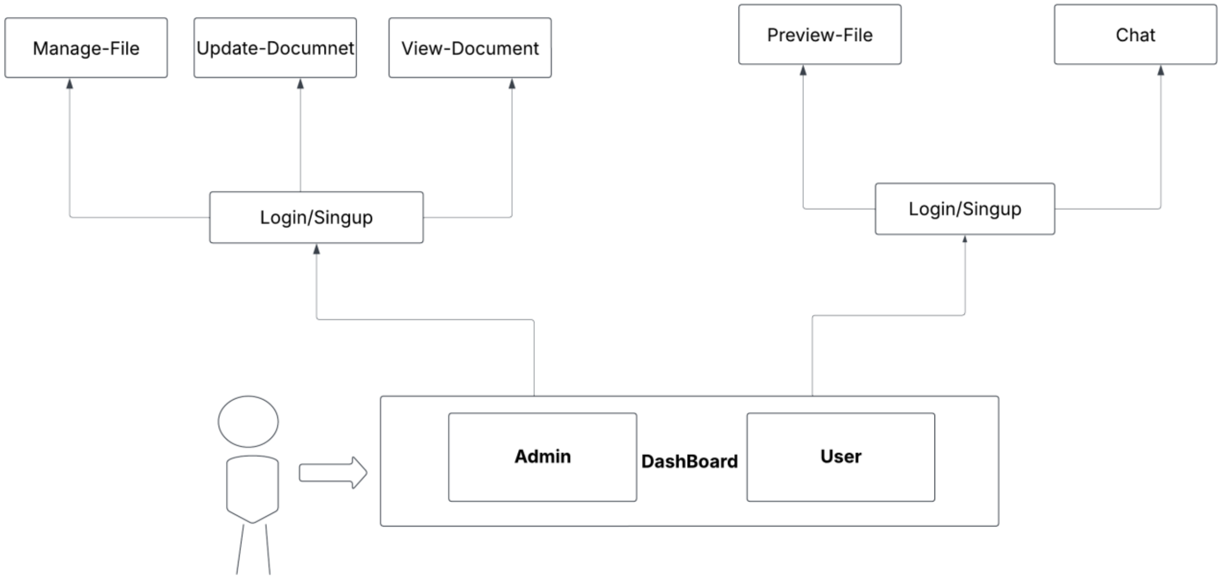
Project Overview

The proposed system is a **Multimodal Retrieval-Augmented Generation (RAG) System** that securely processes text documents and videos, enabling users to retrieve contextual responses using a chatbot. It incorporates authentication, document indexing, and intelligent retrieval mechanisms to enhance user experience and response accuracy.

Objectives

- Develop a secure authentication system with role-based access.
- Implement a file upload and indexing system for PDFs, DOCX, and videos.
- Create an intelligent chatbot for multimodal retrieval of text and video-based answers.
- Enhance user experience with interactive document previews and video segment retrieval.

Flow-Diagram



Project Scope

1. Authentication System

- User registration and login with role-based access control.
- Roles:
  - **Admin:** Upload, view, and manage files.
  - **User:** View uploaded files and interact with the chatbot.

2. Document & Video Upload Module (Admin Only)

- Upload and store PDFs, DOCX, and videos.
- Extract text from documents and transcribe video content.
- Index content for fast and efficient retrieval.

3. Uploaded Files Section (Admin & User Access)

- List all uploaded files with metadata.
- Provide a preview for PDFs and DOCX files.
- Enable in-app video playback.

4. Chatbot with Multimodal Retrieval (Admin & User Access)

- Users can input queries, and the system will retrieve relevant content.
- **Text-based Retrieval:** Highlight relevant text from PDFs/DOCX.
- **Video-based Retrieval:** Identify and display relevant video segments.
- **Hybrid Response:** If multiple sources are relevant, provide a combined response with document preview and video snippet.

Technology Stack

Frontend:

- **React.js** for interactive UI with a modern user experience.

Backend:

- **FastAPI** for handling authentication, file uploads, and chatbot interactions.
- **Database:** PostgreSQL / MongoDB for storing user data and metadata.
- **Storage:** Local or cloud storage for uploaded files.

AI Model:

- **Text Processing:** *Qwen (Alibaba)/Gemini 2.0 for text generation and retrieval.*
- ***OCR & Document Indexing:*** PyMuPDF / pdfplumber / LangChain.
- ***Video Processing:*** OpenAI Whisper / FFMPEG (For frame extraction).
- ***Embedding & Retrieval:*** Pinecone for efficient vector search.

Expected Outcomes

- Secure authentication and role-based access.
- An intelligent chatbot capable of retrieving multimodal content.
- Interactive document and video previews for a seamless user experience.
- Scalable architecture for fast and efficient content retrieval.

Deliverables

1. Authentication module (Sign-in, Sign-up, Role Management).
2. File upload & retrieval system with indexing.
3. Multimodal chatbot providing contextual responses.
4. Testing and optimization to enhance retrieval accuracy.

Milestone (15 Days)

- **Day 1-3:** Setup project structure, authentication module implementation.
- **Day 4-6:** File upload module, document indexing, and storage setup.
- **Day 7-9:** Implement chatbot for text-based retrieval.
- **Day 10-12:** Integrate video processing and retrieval.
- **Day 13-15:** UI/UX improvements, testing, and optimization.