# PDF Question Answering System

This project implements a question answering system for PDF documents using hybrid search (vector and BM25) and language models. It provides an API for uploading PDFs, creating vector stores, and querying the documents.

## Features

* PDF text extraction and chunking
* Hybrid search combining vector search (Chroma) and BM25
* Language model integration for generating answers (using Ollama)
* FastAPI-based API for PDF upload and querying
* Persistence of vector stores for previously uploaded PDFs

## Project Structure

pdf-qa-system/

│

├── data/

│ ├── docs/ # Directory for storing uploaded PDFs

│ ├── vector\_stores/ # Directory for Chroma vector stores

│ └── bm25\_stores/ # Directory for BM25 stores

│

├── src/

│ ├── document\_processing/

│ │ ├── \_\_init\_\_.py

│ │ ├── pdf\_extractor.py

│ │ └── vector\_store.py

│ ├── embeddings/

│ │ ├── \_\_init\_\_.py

│ │ └── embedding\_model.py

│ ├── search/

│ │ ├── \_\_init\_\_.py

│ │ └── hybrid\_search.py

│ └── utils/

│ ├── \_\_init\_\_.py

│ └── file\_utils.py

│

├── main.py # Script for running queries directly

├── main\_api.py # FastAPI application for the API

├── requirements.txt

└── README.md

## Setup

1. Ensure you have Python 3.8+ installed on your system.

2. Navigate to the project directory.

3. Create a virtual environment:

python -m venv .venv

4. Activate the virtual environment:

- On Windows:

.venv\Scripts\activate

- On macOS and Linux:

source .venv/bin/activate

5. Install the required dependencies:

pip install -r requirements.txt

6. Set up the Ollama server:

- Follow the installation instructions at: https://github.com/jmorganca/ollama

- Ensure the Ollama server is running at http://localhost:11434/ before executing the scripts

## Usage

### Running the API

1. Start the API server:

python main\_api.py

The API will be available at http://localhost:8000

2. Use the following endpoints:

- Upload a PDF:

POST /upload\_pdf

Content-Type: multipart/form-data

file: [PDF file]

- Query a single question:

POST /query

Content-Type: application/json

{

"query": "Your question here",

"pdf\_id": "PDF\_ID\_returned\_from\_upload"

}

- Query multiple questions:

POST /multiple\_queries

Content-Type: application/json

{

"queries": ["Question 1", "Question 2", ...],

"pdf\_id": "PDF\_ID\_returned\_from\_upload"

}

### Running the Original Script

To use the original script without the API:

1. Place your PDF file in the project root directory.

2. Modify the `questions` list in `main.py` to add or change the questions you want to ask about the PDF document.

3. Run the main script:

python main.py

## Troubleshooting

* If you encounter OpenMP-related errors, add the following line at the beginning of your Python scripts:

import os

os.environ["KMP\_DUPLICATE\_LIB\_OK"] = "TRUE"

* Ensure that the Ollama server is running before making queries.
* If you're having issues with specific PDFs, try re-uploading them to generate new vector stores.

## Note

This project uses the Ollama library, which requires separate installation and setup of the Ollama server. Make sure to complete the Ollama setup before running the scripts or API.