**Research PDF Graph Extraction Website**

**Introduction:**

This Flask application allows users to upload multiple PDF files containing research data. It automatically extracts relevant graphs (stress-strain graphs) and their captions, identifies and classifies alloy information.

**Technology Stack:**

* **Frontend**:
  + HTML/CSS
  + Bootstrap 5
  + JavaScript (Fetch API)
* **Backend**:
  + Python 3.x
  + Flask Framework
  + Flask-WTF (forms & CSRF)
* **Data Processing & OCR**:
  + PyMuPDF (fitz)
  + Tesseract OCR (pytesseract)
  + OpenCV, PIL (Image processing)

**Setup and Installation:**

1.Clone the repository:

git clone <[**https://github.com/YashviBhagat/graph\_extraction.git**](https://github.com/YashviBhagat/graph_extraction.git) >

cd <project-folder>

**2.** Install dependencies:

pip install flask,

pip install flask-wtf ,

pip install PyMuPDF ,

pip install pytesseract,

pip install Pillow,

pip install opencv-python,

pip install numpy

3. Run the application: in your terminal write this command to run website

python app.py

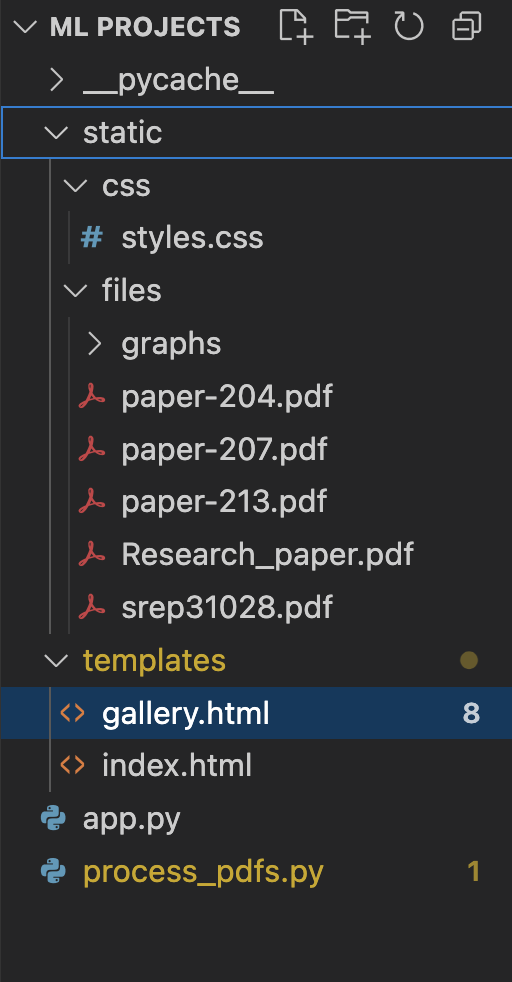
4. Access via browser:

http://localhost:5000

Detailed Functionality:

* **PDF Upload**:
  + Multiple PDF uploads via web form.
  + Extracts images and filters them for stress-strain graphs.
* **Gallery View**:
  + Interactive gallery displaying graphs grouped by their original PDFs.
  + Captions and alloy details displayed alongside images.
* **Graph Management**:
  + Select and delete individual graphs.
  + Delete entire PDFs along with their associated data.
  + Extract sub-graphs from combined graphs.

Project structure:



* App.py:  
  Main Flask application entry-point, handling web routing, form submission, file upload, processing workflows, and AJAX API endpoints.
* process\_pdfs.py:

Data processing and OCR utilities specifically for extracting graphs, detecting relevant axis labels (stress-strain), and extracting captions from uploaded PDFs.

* index.html:

A user-friendly homepage allowing researchers to upload multiple PDF files containing research graphs. Provides immediate feedback on upload status.

* Gallery.html:

Interactive gallery displaying extracted stress-strain graphs, their captions, and alloy details organized by uploaded PDF. Supports graph selection, deletion, and subgraph extraction functionalities.