

Phase 1: System Prerequisites

Before you begin, ensure the following software is installed on your new system:

1. **Git:** To download the project files from your repository.
2. **Docker Desktop:** To build and run the containerized PDF processing. Make sure it is running.
3. **Python:** To run the API server and the model downloader script.
4. **pip:** Python's package installer, to install necessary libraries.

Phase 2: Project Setup

1. Get the Project Code:

Open your terminal or command prompt and clone your Git repository to your machine.

Bash

```
git clone <your_repository_url>
cd <your_project_folder_name>
```

2. Download the ML Model (The "Setup File"):

This is a crucial one-time setup step. Run the setup_model.py script to get the model needed for Round 1B.

Bash

```
# First, install the necessary Python libraries on your local machine
pip install sentence-transformers torch

# Now, run the downloader script
python setup_model.py
```

This will create a folder named `all-MiniLM-L6-v2` inside your `backend` directory.

Phase 3: Build and Run the PDF Processor (Round 1A)

1. Build the Docker Image:

run the build command. This creates the self-contained environment for processing PDFs.

Bash

```
docker build --platform linux/amd64 -t pdf-processor .
cd ..
```

2. Process Your PDFs:

- Place all the PDF files you want to analyze into the `input` folder.
- Run the Docker container to generate the JSON outlines. This command reads from your `input` folder and writes to your `output` folder.

For Windows (PowerShell):

PowerShell

```
docker run --rm -v ${PWD}/input:/app/input:ro -v ${PWD}/output:/app/output --  
network none pdf-processor
```

For Linux/macOS:

Bash

```
docker run --rm -v $(pwd)/input:/app/input:ro -v $(pwd)/output:/app/output --  
network none pdf-processor
```

After this step, your `output` folder should be filled with `.json` files.

Phase 4: Launch the Full Application

1. Start the Backend API Server:

This server powers the entire web application. Run it from your main project directory.

Bash

```
python backend/api.py
```

Wait for the terminal to show "Model loaded successfully" and "Running on <http://127.0.0.1:5001>".

2. Configure the Frontend:

- Get a free **Client ID** from the Adobe PDF Embed API website.
- Open the `frontend/index.html` file in a text editor.
- Find the line `const adobeClientId = "YOUR_ADOBE_CLIENT_ID";` and replace the placeholder with your actual key.

3. Open the Web App:

- Navigate to the `frontend` folder on your computer.
- Double-click the `index.html` file to open it in your web browser.

The application is now fully operational on your new system.