

SauLikh Assignment: Create an API to Process and Store Data from Media Files

Objective

Your task is to develop a Django-based API that efficiently processes two types of media files a **PDF document** and an **image** and stores the extracted data in a structured database.

Requirements

1. Process the provided **PDF file** and extract all readable text.
2. Process the provided **image file** and extract any visible text within it.
3. Store the extracted content in a structured format within a database.
4. Build a Django-based API that accepts media files, extracts their content, and stores it efficiently.
5. The focus is on fast processing and accurate data storage. API endpoints for data retrieval are optional but can be included if necessary.

Expected Outcome

- A working API that accepts PDF and image files, processes them, and stores extracted text.
- A properly structured and labeled database containing the extracted content.
- Efficient handling of different file formats while ensuring data integrity.

Submission Guidelines

- Share the code repository (GitHub) with clear instructions on how to run the solution.
- Provide documentation on how the data is processed and stored.
- Mention any challenges faced and how you approached solving them.
- **Deadline:** Submit the assignment within **48 hours** of receiving this document.

Resources

📁 SauLikh Assignment