

Table Extraction from PDFs

Overview

This project is designed to extract tables from system-generated PDFs and store the extracted data in an Excel file. It provides an efficient way to process structured tabular data while ensuring accuracy in row alignment.

Features

- Extracts text-based tables from system-generated PDFs.
- Maintains structured row alignment for better data consistency.
- Saves the extracted tables into an Excel sheet for easy access and analysis.

Prerequisites

To run this project, ensure that Python is installed on your system. Install the necessary dependencies using the following command:

```
pip install pdfplumber openpyxl
```

Output

The extracted tables will be saved as an Excel file at the specified output location.

Additional Notes

- This tool works best with system-generated PDFs containing well-structured tables.
- It does not support scanned images or highly irregular table formats.