

Automated Invoice Processing System

Overview:

This project is designed to help businesses process thousands of invoices received via email. It automates the extraction of key invoice details (like invoice number, date, amount, and vendor) from PDF files and logs them into an Excel sheet or database for easy tracking and reporting.

Key Features:

- Automatically reads PDF invoices from a folder or email inbox.
- Extracts important fields using text recognition and regex.
- Saves the data to Excel using pandas and openpyxl.
- Optional SQLite database support.
- Tkinter-based simple GUI for uploading and tracking invoices.
- Validation for duplicates and missing data.
- Monthly and vendor-wise summary report generation.

Tech Stack:

- Python
- PyPDF2 or pdfplumber (for reading PDF text)
- re (for extracting data using regular expressions)
- pandas and openpyxl (for Excel handling)
- SQLite (for optional database support)
- Tkinter (for GUI)

Development Phases:

Week 1:

- Setup folder watcher using the OS module.

- Extract text from PDFs using pdfplumber.

- Apply regex to parse invoice details.

Week 2:

- Save the extracted data into Excel using pandas.

- Optional: Store in SQLite database with schema.

Week 3:

- Build UI with tkinter.

- Add input validation (duplicates, missing fields).

Week 4:

- Finalize and clean code.

- Add error handling and logging.

- Generate reports (monthly/vendor-wise).

- Write README and user documentation.

How to Use:

1. Run the script to start the GUI.

2. Click "Select PDF" to upload your invoice file.

3. Click "Save to Excel" to store the invoice data.

4. Click "Generate Summary" to create a monthly report.

Output:

- invoice_collection.xlsx: Contains all uploaded invoice data.

- generate_summary.xlsx: Monthly summary of invoice amounts.

Enjoy your automated invoice tracking!