



ChatPDF Handling PDF extraction, text processing, and user interaction through a web interface.

CODE

Chhunna Noeun^a, Asst. Prof.Dr.Sucha Smanchatb^b, Mr. Tongpool Heeptaisongc^c

^aGinkgo Soft Co., Ltd. 140/30 2nd floor Soi Ratchadapisek 29 (Ratchapracha 4), Ratchadapisek Rd., Chatuchak, Chatuchak, Bangkok 10900

^bKing Mongkut's University of Technology North Bangkok, 1518 Pracharad Road, Wong Sawang, Bang Sue, Bangkok 10800, Thailand

Abstract

For the purpose of to answer user questions, this project creates a chatbot that retrieves and examines financial data from PDF files. In addition to handling a variety of financial papers, including quarterly reports, it offers data on important financial metrics, like the ratio of non-performing loans to total loans, total assets, liabilities, loans, deposits, ROA, and NIM. The chatbot makes financial data more approachable and user-friendly by using Natural Language Processing (NLP) to provide comprehensive responses based on categories such as Financial Position, Performance Indicators, Asset Quality Ratios, and Operating Results.

Introduction

Fast access to reliable financial data is essential for making well-informed decisions in the fast-paced financial world of today. There is a lot of important information in financial reports, like quarterly results and performance summaries, but it can be difficult and time-consuming to sort through these complicated records. The creation of automated technologies that can extract and analyze financial data is becoming more and more important in order to handle this.

The goal of this project is to develop a chatbot application that can automatically extract, analyze, and respond to user inquiries about financial data from PDF files, hence expediting access to financial insights. The chatbot processes a variety of financial documents, such as quarterly statements and bank reports, using Natural Language Processing (NLP) techniques.

Methodology

The development of the chatbot application follows a structured approach, focusing on extracting and processing financial data from PDF reports to provide accurate responses to user queries.

1.Data Collection and Text Extraction

Financial reports, such as quarterly statements, are sourced in PDF format. These documents are processed using PDF parsing libraries (e.g., PyPDF2 or pdfplumber) to extract raw text for further analysis.

2.Data Processing and Cleaning

The extracted text undergoes cleaning and processing using regex patterns and text analysis techniques. This step ensures that key financial terms and figures (e.g., assets, liabilities, NPL, ROA, and NIM) are accurately identified and ready for query handling.

3.Natural Language Processing (NLP) and Query Handling

NLP techniques are applied to interpret user queries, enabling the chatbot to understand questions related to financial data. The chatbot uses predefined logic to extract relevant information from the processed text and answer queries about key metrics like total assets, liabilities, and performance indicators.

4.Response Generation and Testing

The chatbot generates structured responses in natural language for each query. If any data points are missing, the chatbot provides clear explanations. The system is rigorously tested with various financial reports to ensure its accuracy and reliability across different queries and document types.

Technology Stack

In building this chatbot application, a robust technology stack was employed to ensure efficient text extraction, natural language processing, and seamless user interaction. The stack combines powerful libraries and frameworks for text analysis, data handling, and response generation, enabling the chatbot to process financial reports in PDF format and deliver accurate, timely insights to users.

- Backend Framework: Flask (for handling web requests and API development)
- PDF Parsing Libraries: PyPDF2, pdfplumber (for extracting text from PDF financial reports)
- Natural Language Processing (NLP): SpaCy, NLTK (for understanding user queries and extracting relevant financial data)
- Text Processing: Regular Expressions (Regex) (for identifying financial data such as assets, liabilities, loans, deposits, etc.)
- Database: SQLite or any other database (for storing previous queries or processed data, if necessary)
- Frontend: HTML, CSS, JavaScript (for user interaction with the chatbot).

References

- Natural Language Processing in Finance. (n.d.). By Edward Kim and Bryan Routledge NLP in financial documents. [Natural Language Processing in Finance](#)
- Sentiment Analysis and Text Mining in Finance(n.d). By Bing Liu for data extraction techniques from financial documents <https://weather.com/weather/today/l/13.86,100.55?par=google>
- Parsing PDFs for Text Extraction: Tools and Techniques(n.d) <https://stackoverflow.com/questions/937808/how-to-extract-data-from-a-pdf-file-while-keeping-track-of-its-structure>
- Automated Data Extraction from PDFs in Python(n.d) .By DataCamp <https://www.datacamp.com/tutorial/reading-and-editing-pdfs-and-word-documents-from-python>

Chat with PDF

Drag and drop PDF files here or click to upload

Upload

New Folder

Q: Summary

Answer for 1Q24_KBank_unreviewed-E-pages-2.pdf: Financial Overview 1. Financial Position (as of March 31, 2024) Total Assets: 4,318,809 million Baht (up 0.82% from December 31, 2023) Total Liabilities: 3,703,156 million Baht (up 0.45%) Loans: 3 million Baht (down 1.07%) Deposits: 2,702,266 million Baht (up 0.10%) 2. Performance Indicators (Q1 2024) Net Interest Margin (NIM): Not found (down from 3.83% in Q4 2023) Return on Average Assets (ROA): Not found (up from 0.88%) Return on Average Equity (ROE): Not found (up from 7.32%) Cost to Income Ratio: 41.30% (down from 48.15%) 3. Asset Quality Ratios Non-Performing Loans (NPL) to Total Loans: 3.19% (unchanged) Coverage Ratio: 150.35% (down from 152.23%) Loans to Deposits Ratio: Not found (down from 92.25%) Capital Adequacy Ratio: Not found (down from 19.41%) Tier 1 Capital Ratio: Not found (down from 17.44%) 4. Operating Results (Q1 2024) Net Profit: Not found million Baht (up 43.65% from Q4 2023) Basic Earnings per Share: Not found Baht (up 39.65%) Total Operating Income: 50,152 million Baht (down 1.46%) Total Other Operating Expenses: 20,713 million Baht (down 15.48%) Operating Profit Before Expected Credit Loss: Not found million Baht (up 11.57%) Summary of Revenues and Expenses (2023) Total Revenues: Not found million Baht Total Expenses: Not found million Baht

Q: What is the percentage of NPL to total loans?

Answer for 1Q24_KBank_unreviewed-E-pages-2.pdf: The percentage of NPL to total loans is 3.19%.

Type your question here.

Ask

Figure 1:User interaction through a web interface.

Results

User-friendliness and functionality are properly balanced in the "Chat with PDF" application's user interface. While communicating with a chatbot, users can upload and arrange PDF files with ease. The user experience is improved overall by the well-considered design features, which include the responsive chat input, easy folder management, and a clear file upload area. Consequently, the program functions as a useful instrument for efficiently handling and analyzing financial records, rendering it usable and effective for users with different degrees of technical proficiency.

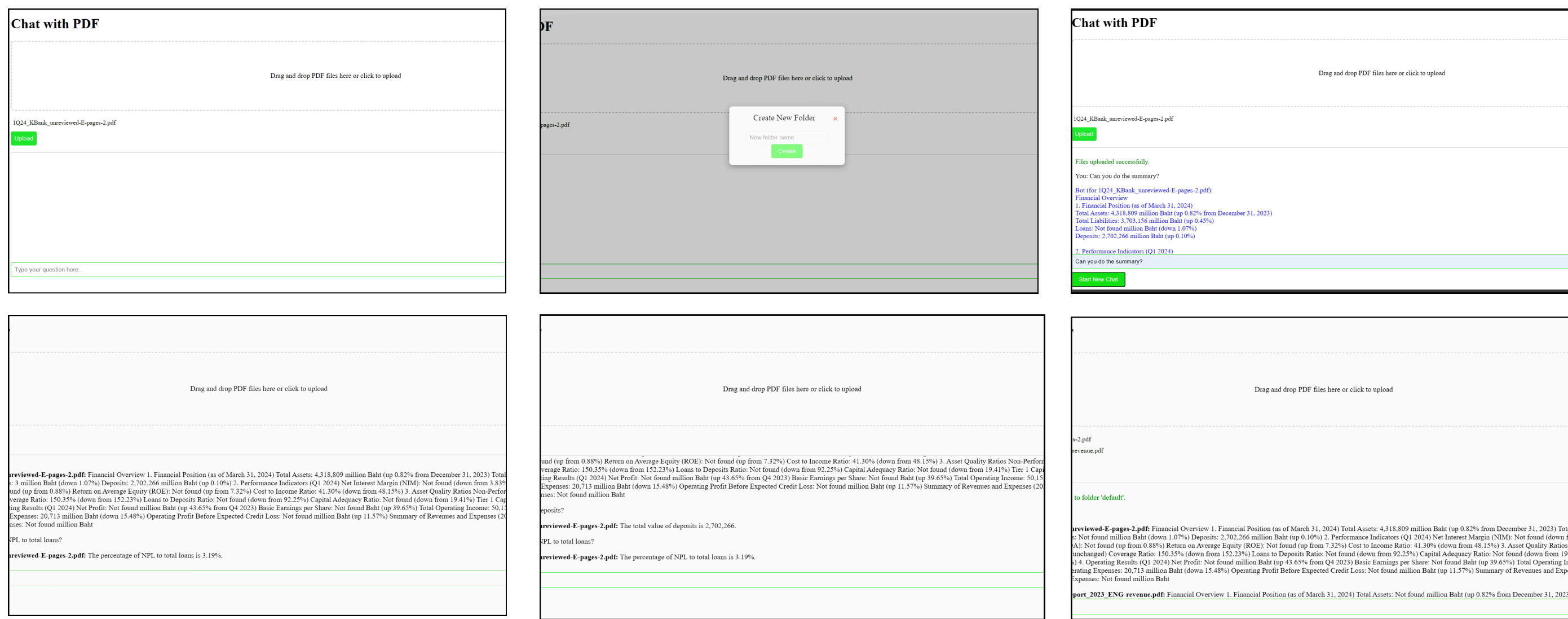


Figure 2: Workflow of User interaction in financial overview

Conclusion

The ChatPDF project successfully integrates PDF processing, chatbot functionality, and financial data analysis, resulting in a powerful tool for users looking to extract and interact with financial report information efficiently. The project has potential for further enhancement, including additional functionalities, improved data visualization, and expanding the range of documents supported.

Acknowledgements

This co-operative program For the priceless experience I obtained throughout my cooperative experience, I would want to sincerely thank Ginkgo Soft Company. I experienced the great privilege of working with a talented team that promoted a collaborative and learning surroundings, and I was able to gain practical abilities in business analysis. In addition to enhancing my professional skills, this experience has increased my enthusiasm future . I sincerely appreciate all of the help and inspiration I got while I was here.