Documentation for QA Bot on Profit & Loss (P&L) Data

Part 2: Interactive QA Bot Interface for Financial Data

Problem Statement:

Develop an interactive interface for the QA bot that allows users to upload PDF documents containing P&L data and query financial information in real time.

Approach:

1. Frontend Development:

- **Framework:**
    - Built a web-based interface using **Streamlit** to provide an intuitive and interactive user experience.
- **Features:**
    - **Uploading Documents:** Users can upload multiple PDF files containing financial data.
    - **Query Input:** Users can input specific financial questions to interact with the bot.
    - **Response Display:** Responses are displayed alongside relevant extracted financial data for clarity.

---

2. Backend Integration:

- **Document Processing:**
    - Extracted tables and text from uploaded PDF files using **pdfplumber**.
    - Enhanced the extracted documents by appending structured tables as metadata to improve indexing and retrieval.
- **Indexing and Embedding:**
    - Leveraged **llama_index** to create a robust indexing mechanism for financial data.
    - Configured the **HuggingFaceEmbedding** model (**BAAI/bge-small-en-v1.5**) for embedding P&L data points, ensuring efficient and accurate retrieval of relevant information.
- **LLM Integration:**
    - Integrated the **Gemini-1.5-flash** language model.
    - Applied a detailed system prompt to guarantee that responses are precise, well-structured, and contextually relevant to financial queries.

---

3. Performance Optimization:

- **Asynchronous Operations:**
    - Utilized asynchronous I/O for efficient handling of large datasets and simultaneous user queries.

- **Granular Parsing:**
    - Implemented window-based sentence parsing with **SentenceWindowNodeParser** for creating fine-grained nodes. This ensures high accuracy in data retrieval and query matching.
- **Metadata Post-Processing:**
    - Used **MetadataReplacementPostProcessor** to dynamically replace metadata in responses for added context and clarity.

Deployment Instructions:

1. System Requirements:

- **Python Version:** Python 3.8 or later.
- **Libraries:**
    - Streamlit
    - pdfplumber
    - llama_index
    - HuggingFace Transformers
    - Nest Asyncio

2. Installation:

1. Clone the repository from your source control system.

bash

CopyEdit

git clone <repository-url>

2. Navigate to the project directory:

bash

CopyEdit

cd financial-qa-bot

3. Install required dependencies:

bash

CopyEdit

pip install -r requirements.txt

3. Configuration:

- Update the settings in initialize_llm and initialize_embeddings to specify the desired models and API keys.
- Set logging levels and formats as needed for debugging or production.

4. Running the Application:

1. Start the Streamlit server:

bash

CopyEdit

streamlit run app.py

2. Open the application in your browser at http://localhost:8501.

5. Access the Deployed Application:

- The application is also deployed at https://queryassistant.streamlit.app/.

6. Usage:

- Upload one or more PDF files containing P&L data.
- Enter financial queries (e.g., *"What is net profit in segment reporting?*").
- View responses alongside extracted financial insights.

Usage Instructions:

1. Uploading Documents:

- Click on the "Upload PDF files" button to select one or more PDF files.
- The system will process the files and prepare the data for query handling.

2. Entering Queries:

- Use the text input field to type your financial query.
- Example: *"What are the operating expenses for Q1?"*

3. Viewing Results:

- Responses will be displayed in the application along with relevant excerpts from the uploaded documents.

Example Queries:

1. *"What is the net profit in segment reporting?"*

   Response: Based on the provided text, the net profit is not directly stated within the segment reporting section. However, the segment operating income is shown, with a total of ₹8,784 crore for the period ending March 31, 2024, and ₹8,998 crore for the period ending March 31, 2023. The overall net profit for the company (not broken down by segment) is given as ₹7,975 crore for the period ending March 31, 2024, and ₹6,134 crore for the period ending March 31, 2023.

2. *"List the revenue details for the last quarter."*

   Response: The provided text gives revenue details for the year ended March 31, 2024 and 2023, but not specifically for the last *quarter* ended March 31, 2024. The document states that this quarterly information is available in Note 2.23, but that note is not included in this context.

3. *"Summarize operating costs for the fiscal year."*

Response: For the fiscal year ended March 31, 2024, total operating expenses were ₹14,510 crore. This comprised ₹6,973 crore in selling and marketing expenses and ₹7,537 crore in general and administration expenses.

Key Features:

- **Scalable Interface:** Handles multiple file uploads and concurrent queries seamlessly.
- **Accurate Retrieval:** Enhanced document processing ensures precise answers to complex financial queries.
- **Optimized Performance:** Asynchronous processing and node parsing enhance speed and accuracy.
- **User-Friendly Design:** Simple and intuitive layout for ease of use by financial analysts and other stakeholders.

Conclusion:

This QA bot provides an effective and efficient solution for analyzing P&L data through an interactive web-based interface. By leveraging state-of-the-art NLP models and embedding techniques, it ensures reliable, precise, and actionable insights for financial professionals.