Part 2: Interactive QA Bot Interface for Financial Data

1. Introduction

The Interactive PDF Query System allows users to upload PDF documents and ask questions based on the extracted content. This system uses advanced Natural Language Processing (NLP) techniques, including sentence embeddings and question-answering models, to extract relevant information from the document and respond to user queries. The goal is to provide an intuitive platform for querying PDFs, such as financial reports or academic papers

2. How to Use the System

Step 1: Upload Your PDF Document

- Navigate to the **"Upload a PDF"** section at the top of the page.
- Click on the **"Browse files"** button to select a PDF file from your local machine.
- Once the file is selected, it will automatically be uploaded to the platform. The system will then proceed to extract the text from the PDF document.

Step 2: Query the Document

- After the text extraction is complete, a new field labeled **"Ask a question about the PDF"** will appear.
- Enter any question related to the content of the uploaded PDF in the provided text input box.
- For example, if you uploaded a financial report, you can ask questions like:
    - "What is the total revenue in 2020?"
    - "What is the net profit for Q4?"

Step 3: Interpret the Response

- The system will analyze your query, retrieve the most relevant sections of the document, and use a pre-trained question-answering model (DistilBERT) to generate an answer.
- The answer will be displayed on the page below the query input box.

Step 4: Interact Continuously

- You can continue interacting with the system by entering more questions. The system will always use the most relevant context from the document to generate responses.

3. System Features

- **PDF Upload**: Upload any PDF document for querying.
- **Text Extraction**: The system extracts text from the PDF using **pdfplumber**, allowing it to handle diverse layouts.
- **Chunking**: The extracted text is split into manageable chunks to ensure that context is preserved while querying.
- **Embedding with FAISS**: The text chunks are embedded using **Hugging Face SentenceTransformer** and stored in **FAISS** for efficient similarity search.
- **Question Answering**: **DistilBERT** is used to generate responses based on the user's query and the context retrieved from the PDF.

Example Interactions

Example 1: Financial Query

- **Uploaded PDF**: An annual financial report of a company.
- **User Question**:
  "What is the total revenue in 2020?"
- **System Response**:
  "The total revenue in 2020 was $5 million."
- Explanation: The system extracts relevant sections from the report that contain the total revenue figure and generates the response accordingly.

Example 2: Financial Query

- **Uploaded PDF**: A quarterly financial report.
- **User Question**:
  "What is the net profit for Q4?"
- **System Response**:
  "The net profit for Q4 is $1.2 million."
- Explanation: The system retrieves and analyzes the Q4 section of the document, then generates an accurate response.

Example 3: General Query

- **Uploaded PDF**: A research paper on machine learning.

- **User Question**:
  "What is the conclusion of the paper?"
- **System Response**:
  "The conclusion highlights the importance of using deep learning techniques for improving predictive accuracy in various applications."
- Explanation: The system searches for the conclusion section of the paper and provides the answer accordingly.

Output received: