App Name: **PDF Reader**

Question:

Write a python code -

a) Install and import langchain library.

b) Use embeddings from OPENAI to embed the attached document

* Create at least 3 different docs by splitting one document.

c) Connect to OPENAI gpt3 api and do the following things,

* Ask a question.
* It should give back the relevant document through the stored vector database ( embedded document )
* It will answer the asked question through GPT3.

**Requirements:**

* Langchain
* OpenAI
* Streamlit
* PyPDF2
* Python-dotenv
* Tiktoken
* Faiss-cpu

To install above requirements, run below code in terminal:

pip install langchain openai streamlit PyPDF2 python-dotenv tiktoken faiss-cpu

Enter your api key in “.env” file

Enter below code in terminal to run the app:

streamlit run app.py

After running the above code, you can use website in below URL:

<http://localhost:8501/>

**Working of the application:**

The application reads the PDF and splits the text into smaller chunks that can be then fed into a LLM. It uses OpenAI embeddings to create vector representations of the chunks. The application then finds the chunks that are semantically similar to the question that the user asked and feeds those chunks to the LLM to generate a response.

Code: 

app.py:

from dotenv import load\_dotenv

import streamlit as st

from PyPDF2 import PdfReader

from langchain.text\_splitter import CharacterTextSplitter

from langchain.embeddings.openai import OpenAIEmbeddings

from langchain.vectorstores import FAISS

from langchain.chains.question\_answering import load\_qa\_chain

from langchain.llms import OpenAI

def main():

    load\_dotenv()

    st.set\_page\_config(page\_title="PDF reader")

    st.header("PDF reader")

*# To upload file*

    pdf = st.file\_uploader("Upload your PDF", type="pdf")

*#To extract the text*

    if pdf is not None:

      pdf\_reader = PdfReader(pdf)

      text = ""

      for page in pdf\_reader.pages:

        text += page.extract\_text()

*#To split into chunks*

      text\_splitter = CharacterTextSplitter(

        separator="\n",

        chunk\_size=1000,

        chunk\_overlap=200,

        length\_function=len

      )

      chunks = text\_splitter.split\_text(text)

*#To create embeddings*

      embeddings = OpenAIEmbeddings()

      knowledge\_base = FAISS.from\_texts(chunks, embeddings)

*#To show user input*

      user\_question = st.text\_input("Ask a question about your PDF:")

      if user\_question:

        docs = knowledge\_base.similarity\_search(user\_question)

        llm = OpenAI()

        chain = load\_qa\_chain(llm, chain\_type="stuff")

        response = chain.run(input\_documents=docs, question=user\_question)

        st.write(response)

if \_\_name\_\_ == '\_\_main\_\_':

    main()

.env:

*# Enter your api key below*

OPENAI\_API\_KEY=""

Abhilash D R

+91 82172 94310

abhi01horse@gmail.com