**Querying over documents using Retrieval chain**

In this solution, I have developed a robust information retrieval system using Large Language Models (LLMs) from OpenAI, complemented by the LangChain library for enhanced functionality. The process begins by importing necessary libraries, including LangChain modules and OpenAI, and loading documents from a specified directory. To optimize document handling, I implemented a text-splitting mechanism. The core functionality resides in the **find\_answer** function, where user queries trigger a retrieval chain query process. The LangChain library plays a crucial role in vectorizing and embedding the documents using SentenceTransformerEmbeddings, thereby creating a Chroma vector database for efficient storage and retrieval. The OpenAI API seamlessly integrates into the ChatOpenAI model, which is specifically tailored for question-answering. A predefined system prompt sets the context for the chatbot's responses, ensuring both relevance and accuracy. The solution demonstrates its prowess through sample queries and corresponding answers, showcasing its ability to fetch and present precise information from the loaded documents. This comprehensive approach, combining OpenAI and LangChain, ensures a reliable and efficient information retrieval system with practical applications in various domains.

Query on the documents…..

**Code:**

#Need to import all these:

import openai

import os

from langchain.chat\_models import ChatOpenAI

from langchain.embeddings.openai import OpenAIEmbeddings

from langchain.vectorstores import Chroma

from langchain.chains import RetrievalQA

from langchain.llms import OpenAI

from langchain.text\_splitter import RecursiveCharacterTextSplitter

from langchain.embeddings import SentenceTransformerEmbeddings

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

from langchain.document\_loaders import DirectoryLoader

from langchain.document\_loaders import UnstructuredExcelLoader

directory ='C:\\Users\\msgupta9\\PycharmProjects\\pythonProject6\\test2'

# Function to load documents

def load\_docs(directory):

loader = DirectoryLoader(directory)

documents = loader.load()

return documents

documents = load\_docs(directory)

len(documents)

print(documents)

def split\_docs(documents, chunk\_size=1000, chunk\_overlap=20):

text\_splitter = RecursiveCharacterTextSplitter(chunk\_size=chunk\_size, chunk\_overlap=chunk\_overlap)

docs = text\_splitter.split\_documents(documents)

return docs

# Function to give answers of all the queries

def find\_answer(user\_question):

documents = load\_docs(get\_me\_data\_path)

print(f"In this directory we have {len(documents)} files")

docs = split\_docs(documents)

# print(docs)

embeddings = SentenceTransformerEmbeddings(model\_name="all-MiniLM-L6-v2")

persist\_directory = "chroma\_db\_data"

print("Adding to chromadb")

vectordb = Chroma.from\_documents(

documents=docs, embedding=embeddings, persist\_directory=persist\_directory

)

print("Added data to chromadb")

vectordb.persist()

os.environ["OPENAI\_API\_KEY"] = “”

llm = ChatOpenAI(model\_name="gpt-4-32k", temperature=0,

openai\_api\_key="”

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

system\_prompt = """You are an assistant chatbot. Your purpose is to help management consultants at Deloitte. You are an expert at developing business case models. Specifically, you work on Application Modernization business cases, which fall under Tech Cost and Performance Management. Help the user think through tasks and questions.

Your answers must be based on the context provided, which will contain values from a database and/or a table, and documents that contain industry benchmarks. The user will ask you to find specific benchmarks or data points such as “What is the typical labor rate for roles needed for application modernization?” To support the business case.

Your response should provide a clear and concise data point and/or assumptions. Ensure that the data point and/or assumptions are relevant and accurate. Make sure to provide additional context around the data point and/or benchmarks and assumptions with a description.

Be truthful in your answer, if a sufficient answer isn’t found to a user’s chat, respond “I don’t know”. Make sure you refer to the given context to respond to the user."""

query = user\_question

matching\_docs = vectordb.similarity\_search(query)

answer = chain.run(input\_documents=matching\_docs, question=query, prompt=system\_prompt)

print(answer)

find\_answer('What was the revenue for Q1FY23?')

Answer:

The revenue for Q1FY23 was Rs 570.3 crores.

find\_answer("What was the revenue from operations difference between Q1FY23 and Q4FY22?")

The revenue attrition at 21.0% in Q4FY23 in comparison with 23.3% in Q3FY23.

find\_answer("What was the net profit difference between Q1FY23 and Q1FY22?")

The net profit for Q1FY23 was 135.0 crores and for Q1FY22 it was 84.4 crores. Therefore, the net profit difference between Q1FY23 and Q1FY22 is 50.6 crores.

find\_answer('What was the operating EBITDA margin for the full year FY23?')

The operating EBITDA margin for the full year FY23 was 17.8%.

find\_answer('What was the Q4 FY23 Revenue?')

The Q4 FY23 Revenue was ₹ 709 Cr.

find\_answer('Who is the Auditor of Mastek?')

Arun Agarwal is the Auditor of Mastek.

find\_answer('How many new clients did the company add in Q1FY23?')

According to the context,33 Clients add in Q1FY23.

find\_answer('How many employees were based offshore in India as of 30th June, 2022? ')

As of 30th June, 2022, there were 4,283 employees based offshore in India.

find\_answer(' How many large frameworks has Mastek been shortlisted for in the UK Public Sector?')

Fortune India ‘Next 500’ 2022 - Fortune has listed Mastek in its prestigious ‘Next 500’ list,