Project Report: AI Question-Answering System

By, Sreshta Praveen
Course: MSc Data Science
SRN: R23DG056

Project Overview

The AI Question-Answering System is an interactive web-based application built using Streamlit and the Groq LLM API. The project is designed to fetch real-time webpage content and answer user queries related to it. For this implementation, we used the **Reva University** website as the reference source. The tool leverages LangChain, conversational memory, and dynamic prompt creation to provide accurate responses to user questions.

Project Components

This project integrates the following tools and technologies:

- **Streamlit**: Builds an intuitive and interactive user interface.
- **LangChain**: Implements conversational capabilities with memory to handle context-aware queries.
- **Groq API**: Utilizes the Groq LLM to generate responses based on the input prompt and webpage content.
- **BeautifulSoup**: Extracts and cleans webpage content for accurate processing.
- **Dotenv**: Secures sensitive credentials using environment variables.
- **Python Libraries**: Core libraries such as requests and os handle HTTP requests and environment management.

Code Explanation

Step 1: Import Required Libraries
The required Python libraries include `Streamlit` for the interface, `os` for environment variable handling, and `requests` for fetching webpage content.

Step 2: Load Environment Variables
Environment variables are loaded using `dotenv` to secure sensitive API keys.

python

Copy code

```
from dotenv import load_dotenv

load_dotenv()

groq_api_key = os.environ['GROQ_API_KEY']
```

Step 3: Fetch Webpage Content
The `fetch_webpage_content` function fetches and cleans content from the Reva University website.

python

Copy code

```
def fetch_webpage_content(url):

    response = requests.get(url)

    if response.status_code == 200:

        soup = BeautifulSoup(response.content, 'html.parser')

        paragraphs = soup.find_all('p')

        content = " ".join([para.get_text(strip=True) for para in
paragraphs])

        return content

    else:

        return "Unable to fetch webpage content. Please check the
URL."
```

Step 4: Initialize Conversational Memory
LangChain's `ConversationBufferWindowMemory` is used to handle the conversational context.

Step 5: Set Up the Question-Answering Interface
The Streamlit application accepts a user query, fetches webpage content, and returns AI-generated responses based on the Groq LLM.

python

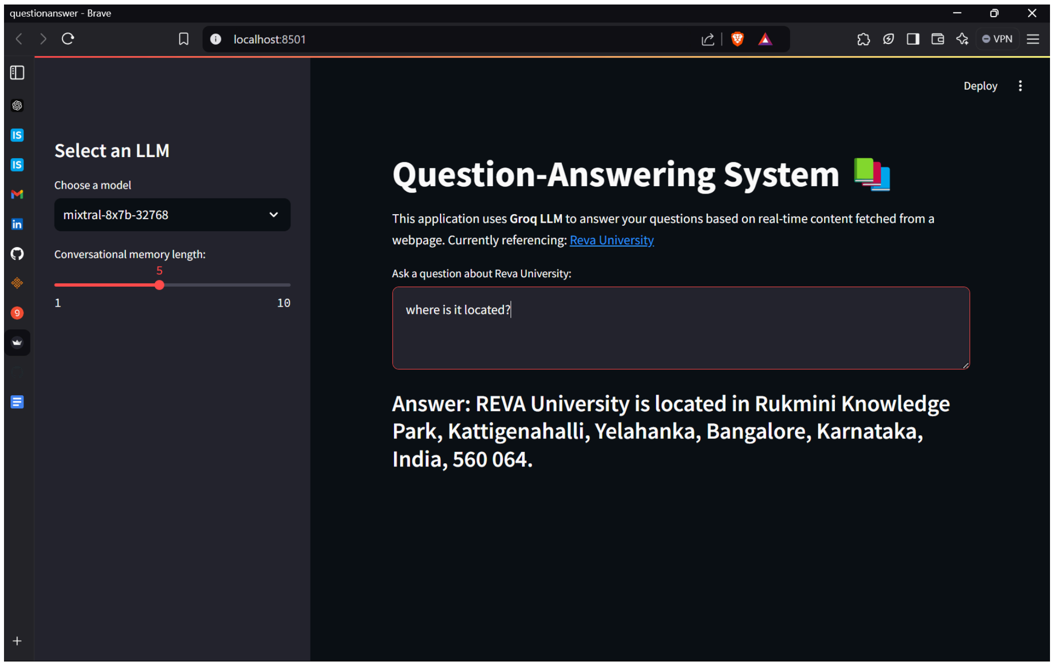
Copy code

```
if user_question:

    prompt = prompt_template.format(context=webpage_content,
question=user_question)

    response = conversation(prompt)

    st.write("### Answer:", response["response"])
```

User Interaction Example

Below are examples of user queries and corresponding system responses. (Screenshots can be added as needed.)

- **Input**: "Where is it located?"
  Output:



Choose a model: mixtral-8x7b-32768

Conversational memory length: 5 (1–10)

Question-Answering System 📚

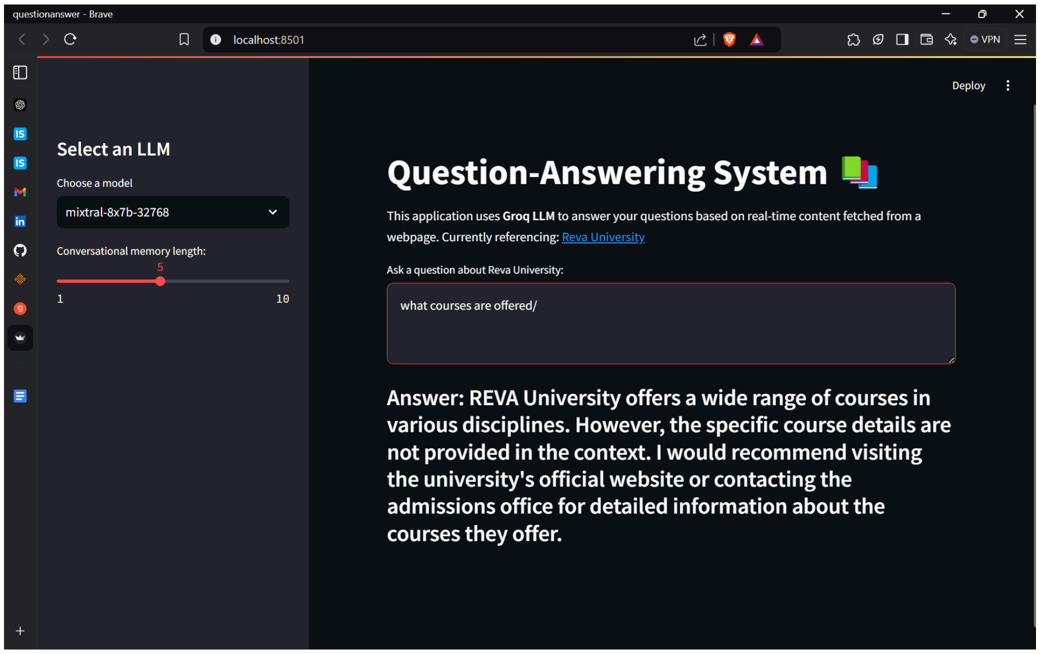
This application uses **Groq LLM** to answer your questions based on real-time content fetched from a webpage. Currently referencing: Reva University

Ask a question about Reva University:

where is it located?

Answer: REVA University is located in Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bangalore, Karnataka, India, 560 064.

- **Input**: "What courses are offered?"
  Output:



Select an LLM

Question-Answering System

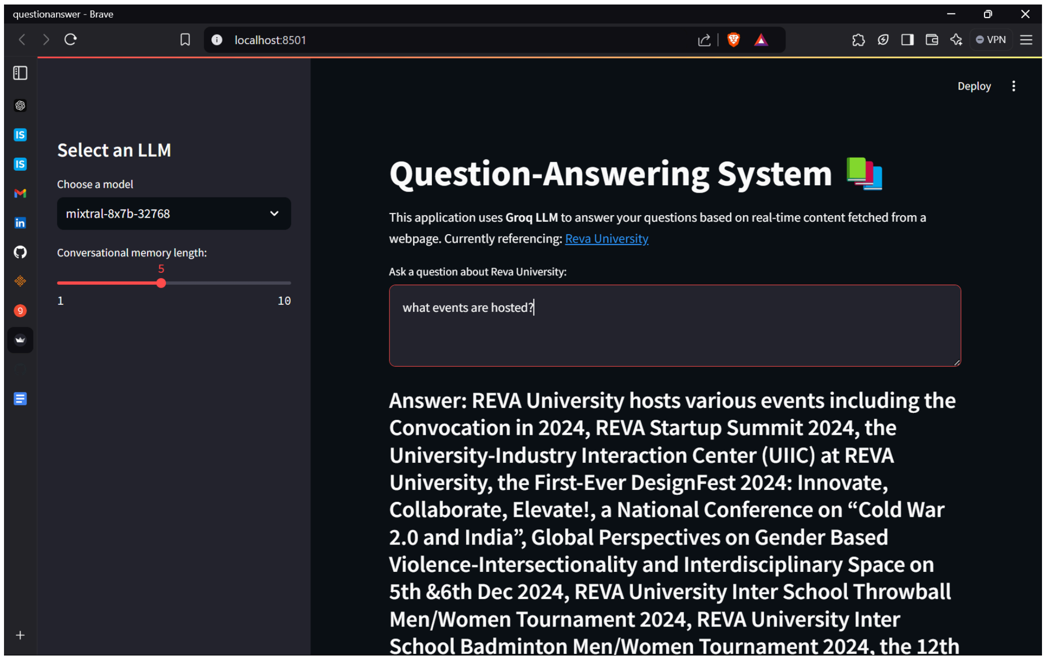
This application uses **Groq LLM** to answer your questions based on real-time content fetched from a webpage. Currently referencing: Reva University

Ask a question about Reva University:

what courses are offered/

Answer: REVA University offers a wide range of courses in various disciplines. However, the specific course details are not provided in the context. I would recommend visiting the university's official website or contacting the admissions office for detailed information about the courses they offer.

- **Input**: "What events does the university host?"
  Output:



Select an LLM

Question-Answering System

This application uses **Groq LLM** to answer your questions based on real-time content fetched from a webpage. Currently referencing: Reva University

Ask a question about Reva University:

what events are hosted?

Answer: REVA University hosts various events including the Convocation in 2024, REVA Startup Summit 2024, the University-Industry Interaction Center (UIIC) at REVA University, the First-Ever DesignFest 2024: Innovate, Collaborate, Elevate!, a National Conference on "Cold War 2.0 and India", Global Perspectives on Gender Based Violence-Intersectionality and Interdisciplinary Space on 5th &6th Dec 2024, REVA University Inter School Throwball Men/Women Tournament 2024, REVA University Inter School Badminton Men/Women Tournament 2024, the 12th

---

Future Enhancements

1. **API Expansion**: Integrate additional AI APIs for enhanced response quality.

2. **Dynamic URL Support**: Allow users to query any webpage, not just Reva University.
3. **Advanced Memory**: Incorporate context-aware features across multiple queries in a session.
4. **Custom Query Templates**: Let users select predefined templates for specific question types.
5. **User Authentication**: Save query history and personalized recommendations for registered users.

Conclusion

This AI Question-Answering System demonstrates a seamless integration of Groq LLM with LangChain and Streamlit to provide accurate, real-time responses based on live webpage content. With its modular design, the system is scalable and adaptable for various domains.

By combining efficiency, user-friendly design, and conversational capabilities, this tool is well-suited for educational and informational applications.
