RAG in Depth Chat History

Contents

- Goals.
- Dependencies and necessary modules.
- .env file.
- LangSmith tracking since day 1.
- Initial RAG app.
- Contextualizing the question.
- Chain with chat history.
- Advanced technique.

Goals

- Allow the user to have a back-and-forth conversation, meaning the application needs some sort of "memory" of past questions and answers.
- LangSmith tracking to visualize the operations.

Dependencies and necessary modules

- New virtual environment.
- Install Jupyter Labs.
- LLM Model: OpenAl Chat Model.
- Embeddings Model: OpenAl.
- Vector Database: Chroma.
- Necessary modules:
 - Langchain, Langchain-openai, Langchain-community, Langchainhub
 - Chromadb
 - BS4 (Beautiful Soup, for web scraping).

.env file

- OpenAl API Key.
- LangSmith credentials.
- LangSmith project name.

LangSmith tracking

The LangSmith tracking starts with the first invoke.

Initial RAG App

Same as in Quickstart.

Contextualizing the question

Sub-chain that takes historical messages and the latest user question.

Chain with chat history

Only run the chain when the chat history is not empty.

Advanced technique

Link to the documentation page.