

RAG in Depth

Streaming

Contents

- Goals.
- Dependencies and necessary modules.
- .env file.
- LangSmith tracking since day 1.
- Initial RAG app.
- Streaming.

Goals

- Streaming refers to a continuous flow of text being delivered in the response, rather than waiting for the entire response to be completed and displayed all at once.

Dependencies and necessary modules

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

.env file

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

LangSmith tracking

- The LangSmith tracking starts with the first invoke.

Initial RAG App

- Same as in Quickstart + Return Sources.

Streaming

- .stream