Langchain and large documents

LangChain

Langchain overview

- Highly capable Python module
- https://python.langchain.com/v0.1/docs/get_started/introduction/
- Supports different functionalities, including RAG
- Developed for chained processing of (very) large documents
- This lesson: Summarize
  - https://python.langchain.com/v0.1/docs/use_cases/summarization
- Map / Reduce-based chaining

Langchain overview

- Manage large text documents:
  - breaking them down into manageable chunks for efficient processing
- **Text Chunking**: Dividing large documents into smaller, coherent sections.
- **Context Management**: Maintaining context across different chunks to ensure coherent analysis and generation.
- **LLM support**: Langchain supports many other Large Language Models

Integration with OpenAI API

- By integrating Langchain with the OpenAI API, you can perform:
- **Summarization**: Generating concise summaries of lengthy documents.
- **Question Answering**: Extracting relevant information based on specific queries.
- **Content Generation**: Producing detailed content or expanding on sections of a document.

Example Use Case

To process a large document using Langchain and the OpenAI API:

1. **Chunk the Document**: Use Langchain to divide the document into smaller sections.
2. **Process Chunks with OpenAI**: Send each chunk to the OpenAI API for analysis or generation (Map)
3. **Combine Results**: Aggregate the results from the API to form a coherent output (Reduce).