Knowledge Graph Creation and Graph RAG Pipeline Implementation

Execution Steps:

Step 1: Dataset Collection

- I used the Wikipedia API to extract data on the topic of "The French Revolution." The data was retrieved as a string and used for further processing.

Step 2: Text to Graph Conversion

- To handle memory efficiency and ensure the model performs well, the data was split sentence-wise and processed in batches.

Step 3: Entity Extraction

- The Babelscape/rebel-large model from Hugging Face was used for entity extraction from the text. The data, after splitting, was fed into the model one by one, which extracted entities in the following syntax:

[{'head': 'French Revolution', 'type': 'has part', 'tail': 'Estates General of 1789'}]

Step 4: Graph Creation

- The NetworkX library was used for graph creation using the extracted entities. The 'head' and 'tail' represent nodes, indicating the direction, while 'type' represents the relationship.
- To visualize the graph, the Pyvis Network library was employed. The created graph was also saved in an HTML file for future reference.

Step 5: Graph-RAG Pipeline Implementation

- The Transformers library was used for the Question-Answering model. The model automatically runs on a GPU if available; otherwise, it defaults to the CPU.

- **Query Processing:**

    1. **Entity Extraction**: The Spacy library was used to extract important entities and nouns from the question.

2. **Subgraph Retrieval**: The extracted entities were checked against the graph nodes to find similar entities. The connected relations were then used to form a subgraph based on the question.
3. **Textual Representation**: The subgraph was converted into a textual format to provide context for the next step.
4. **Question Answering**: The Transformers model takes the question and the textual representation of the subgraph as inputs to generate the most appropriate response.
5. **Response Enhancement**: The Qwen/Qwen2.5-0.5B-Instruct model was used to enhance the answer, making it more descriptive and informative.

Results:

The main graph details include a directed graph with 624 nodes and 585 edges.

Example Queries and Responses

1. **Query**: When did the French Revolution start?
    - **Answer**: 20 April 1792
    - **Enhanced Answer**: The French Revolution began on April 17, 1789.
2. **Query**: Who led the coup of 18 Brumaire in November 1799?
    - **Answer**: Napoleon Bonaparte
    - **Enhanced Answer**: The person who led the coup of 18 Brumaire was Napoleon Bonaparte. He seized power in France in November 1799 and became its dictator for three years until his death in 1821.
3. **Query**: What was the name of the assembly that replaced the monarchy in September 1792?
    - **Answer**: National Assembly
    - **Enhanced Answer**: The National Assembly replaced the monarchy in September 1792.
4. **Query**: What was the name of the radical club that played a significant role in the Revolution?
    - **Answer**: Jacobin Club
    - **Enhanced Answer**: The Jacobin Club played a crucial role in the revolution, marking the beginning of the French Revolution.
5. **Query**: What was the main cause of the French Revolution?
    - **Answer**: acute food shortages

- **Enhanced Answer**: The main cause of the French Revolution was acute food shortages. This period saw widespread famine, resulting in widespread poverty and hardship among the French population. The lack of sufficient food supply led to significant social unrest and political upheaval.

Project Files

The project folder contains the following files:

- Graph RAG Model Jupyter Notebook
- Graph RAG Model Python Script
- requirements.txt