Interactive Data Exploration Using Knowledge Graphs and Queries

Srinivas Rahul Sapireddy, Nimisha, Padma Sai

University of Missouri Kansas City

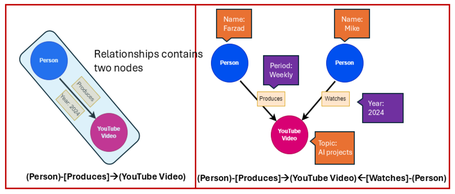
INTRODUCTION

The "Knowledge_Graph_LLM" project, focuses on constructing and utilizing a knowledge graph derived from a movie dataset. The primary goal is to create a knowledge graph that can be queried, enhanced with Retrieval-Augmented Generation (RAG), and integrated into a chatbot interface for answering user queries.

Knowledge Graph:

Knowledge graph: A database that stores information in nodes and relationships

- Node properties: **key** (e.g: name) + **value** (e.g: Farzad)
- Nodes can have labels that groups them together.
- Relationships have **direction**, **type**, and **properties**



This diagram illustrates how a knowledge graph organizes complex data, enabling efficient querying and understanding of entity relationships. In this case, it demonstrates how user interactions with content (like producing or watching videos) are stored and queried.

OBJECTIVES

- Develop a knowledge graph using a movie dataset.
- Enable complex querying capabilities over the knowledge graph.
- Integrate the knowledge graph with RAG to enhance information retrieval.
- Design and implement a chatbot that leverages the knowledge graph.

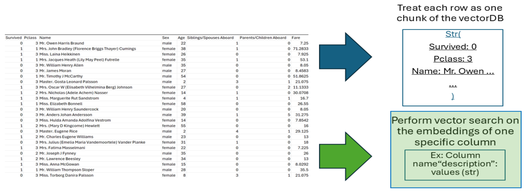
MATERIALS AND METHODS

- **Data Preparation:** Curate and preprocess a comprehensive movie dataset, ensuring it includes relevant entities such as movies, actors, genres, and directors.
- **Graph Database Setup:** Utilize a graph database to store and manage the structured data, facilitating efficient querying and relationship mapping.
- **Data Ingestion:** Populate the graph database with the prepared dataset, establishing nodes and edges to represent entities and their relationships.
- **Query Implementation:** Develop and execute queries to retrieve information from the knowledge graph, addressing complex questions about movies and related entities.
- **RAG Integration:** Enhance the system's retrieval capabilities by incorporating Retrieval-Augmented Generation, allowing for more contextually relevant responses.
- **Chatbot Development:** Design a chatbot interface that interacts with users, leveraging the knowledge graph and RAG to provide accurate and informative answers.

RESULTS

The project successfully constructed a knowledge graph from the movie dataset, enabling complex queries and information retrieval. The integration of RAG improved the system's ability to generate contextually relevant responses. The developed chatbot effectively utilized the knowledge graph to answer user queries related to movies, actors, genres, and other related information.

RAG on Tabular Data

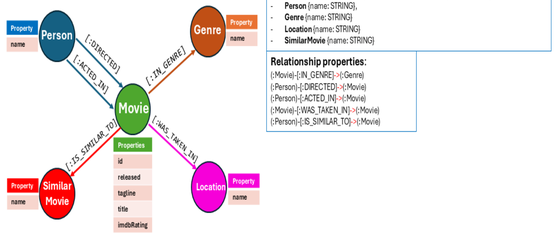


Knowledge Graph for Movie Dataset

Dataset

	movieId	released	title	actors	director	genres	imdbRating
0	1	1995-11-22	Toy Story	Jim Varney|Tim Allen|Tom Hanks|Don Rickles	John Lasseter	Adventure|Animation|Children|Comedy|Fantasy	8.3
1	2	1995-12-15	Jumanji	Robin Williams|Bradley Pierce|Kirsten Dunst|Jo...	Joe Johnston	Adventure|Children|Fantasy	6.9
2	3	1995-12-22	Grumpier Old Men	Walter Matthau|Ann-Margret|Jack Lemmon|Sophia ...	Howard Deutch	Comedy|Romance	6.6
3	4	1995-12-22	Waiting to Exhale	Whitney Houston|Lela Rochon|Angela Bassett|Lor...	Forest Whitaker	Romance|Drama|Comedy	5.6
4	5	1995-12-08	Father of the Bride Part II	Steve Martin|Kimberly Williams-Paisley|Diane K...	Charles Shyer	Comedy	5.9
5	6	1995-12-15	Heat	Al Pacino|Robert De Niro|Val Kilmer|Jon Voight	Michael Mann	Action|Crime|Thriller	8.2
6	7	1995-12-15	Sabrina	Julia Ormond|Harrison Ford|Nancy Marchand|Greg...	Sydney Pollack	Comedy|Romance	6.2

Knowledge graph



Nodes and their properties:
- **Movie** {imdbRating: FLOAT, id: STRING, released: DATE, title: STRING},
- **Person** {name: STRING},
- **Genre** {name: STRING}
- **Location** {name: STRING}
- **SimilarMovie** {name: STRING}

Relationship properties:
(:Movie)-[:IN_GENRE]->(:Genre)
(:Person)-[:DIRECTED]->(:Movie)
(:Person)-[:ACTED_IN]->(:Movie)
(:Movie)-[:WAS_TAKEN_IN]->(:Movie)
(:Person)-[:IS_SIMILAR_TO]->(:Movie)

WORKFLOW

1.User Query:
1. Users input their queries through a user-friendly interface like **Gradio**.
2. Queries can be in natural language, enabling intuitive interaction without the need for technical expertise.

2.SQL Agent Interaction (Yellow Path):
1. The SQL Agent processes structured queries designed to extract specific details from the **Knowledge Graph**.
2. It identifies and retrieves the most relevant nodes and relationships based on the user query.

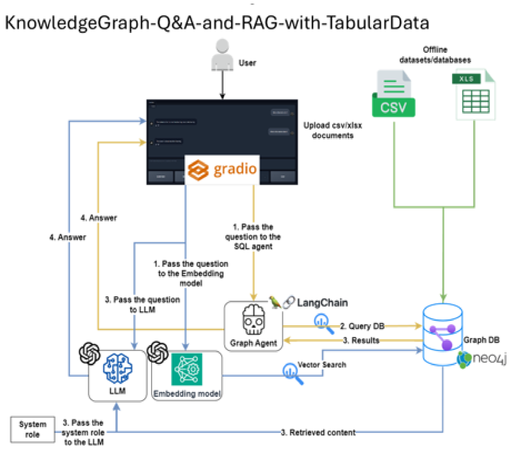
3.Embedding Model and LLM Integration (Blue Path):

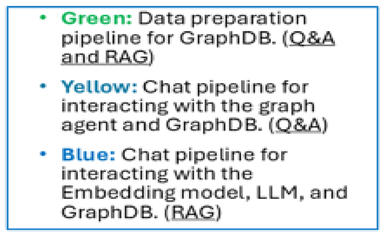
a. Embedding Model:
1. Converts the natural language query into a vector representation.
2. This vectorized query is used to search for semantically similar content within the database.

b.Content Retrieval:
1. Using **vector similarity**, the system retrieves the most relevant information from the database.
2. This includes relational data, graph-based connections, and contextual insights.

c.LLM (Large Language Model):
1. The LLM processes the query and integrates the retrieved data.
2. It formulates a **comprehensive, context-aware answer** by synthesizing the user query with the retrieved information.
3. The final response may include structured data, detailed explanations, and insights extracted from both tabular and graph-based data sources.

4.Response Delivery:
1. The system presents the user with a clear and concise answer.
2. Depending on the query, the output can include textual responses, visualizations, or a combination of both.

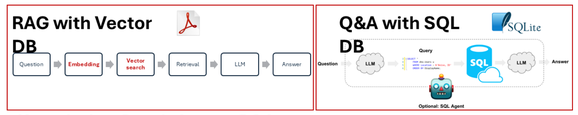




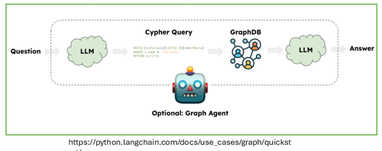
CONCLUSIONS

This project demonstrates the feasibility and effectiveness of combining knowledge graphs with Retrieval-Augmented Generation to enhance information retrieval systems. The integration into a chatbot interface showcases practical applications in providing users with accurate and contextually relevant information.

Q&A on Graph DB using Graph Agent



Knowledge Graph Agent: Q&A



https://python.langchain.com/docs/use_cases/graph/quickstart/

REFERENCES

- **Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., & Maleshkova, M.** (2021). *"Knowledge Graphs."* ACM Computing Surveys (CSUR), 54(4), 1-37.
- Comprehensive survey on knowledge graphs, covering their construction, management, and applications.
- **Paulheim, H.** (2017). *"Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods."* Semantic Web, 8(3), 489-508.
- Discusses techniques for refining and enhancing knowledge graphs.
- **Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., & Riedel, S.** (2020). *"Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks."* NeurIPS 2020.
- Introduces RAG, a method that combines retrieval and generation for knowledge-based tasks.
- **Yih, W. T., Chang, M. W., Meek, C., & Pastusiak, A.** (2014). *"Question Answering Using Knowledge Base with Search-Based Question Reformulation."* Proceedings of the 2014 NAACL-HLT Conference.
- Presents a method for enhancing QA systems with knowledge bases, relevant to knowledge-driven chatbot design.
- **Robinson, I., Webber, J., & Eifrem, E.** (2015). *"Graph Databases: New Opportunities for Connected Data."* O'Reilly Media.
- A practical guide on using graph databases, crucial for managing large-scale knowledge graphs.
- **Wood, P. T.** (2012). *"Query Languages for Graph Databases."* ACM SIGMOD Record, 41(1), 50-60.
- Survey of query languages for graph databases, useful for implementing querying capabilities.

ACKNOWLEDGES AND CONTACTS

1. Srinivas Rahul Sapireddy
   Email: ssdx5@umsystem.edu
1. Nimisha
2. Padma Sai