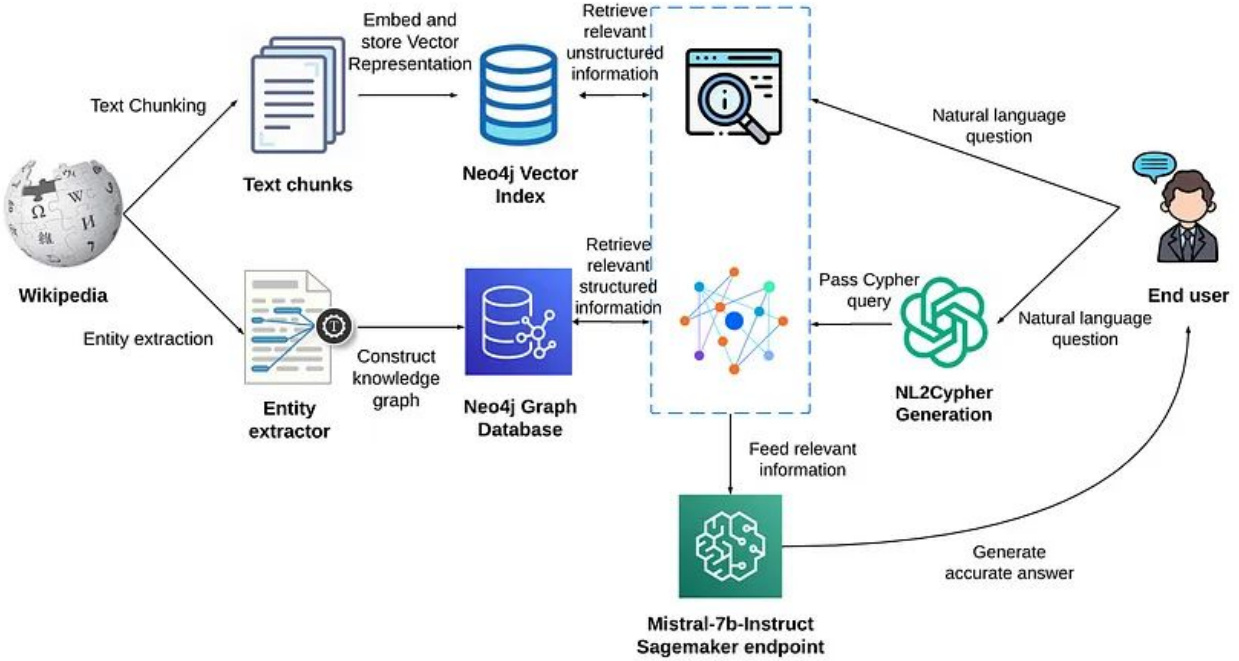
- Saurav Joshi

Enhanced QA Integrating Unstructured and Graph Knowledge Using Neo4j and LangChain



Introduction

- Masters in Data Science - University of Southern California - 2022 - Present
- ML Research - Information Sciences Institute (ISI) - 2022 - Present
- Google Summer of Code Mentor - DBpedia - 2023 - Present
- Google Summer of Code Contributor - DBpedia - 2022 - 2023
- Bachelors in Computer Science - University of Mumbai - 2018 - 2022

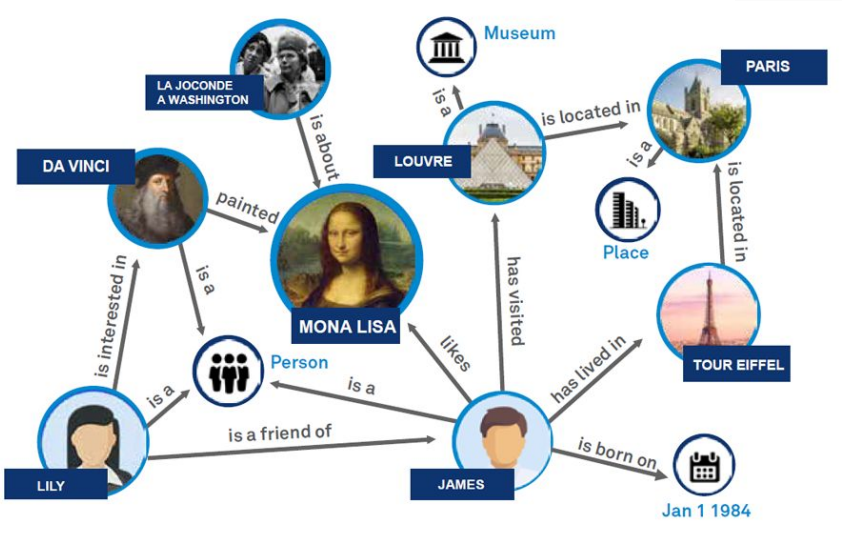
Google Summer of Code

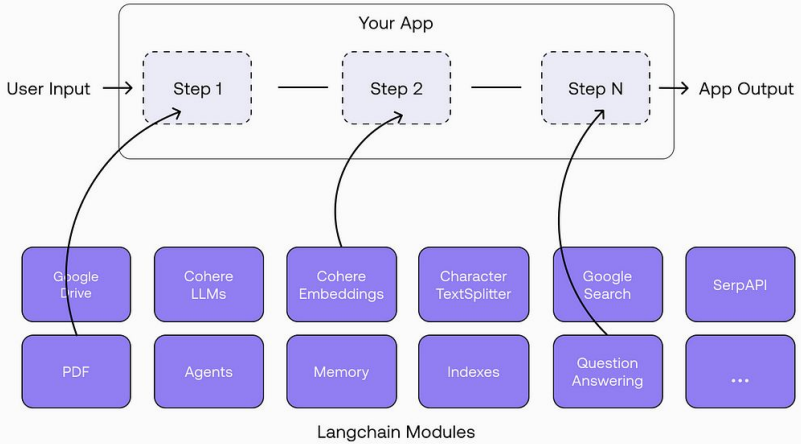
Homepage: https://sauravjoshi23.github.io/ Email: syjoshi@usc.edu

Agenda

- Presentation - 20 mins
- Code Walkthrough - 20 mins
  - Retrieval Augmented Generation using Neo4j and LangChain - 15 mins
  - Knowledge Graph Construction using Neo4j and LangChain - 5 mins
- Q&A - 10 mins

Neo4j x LangChain

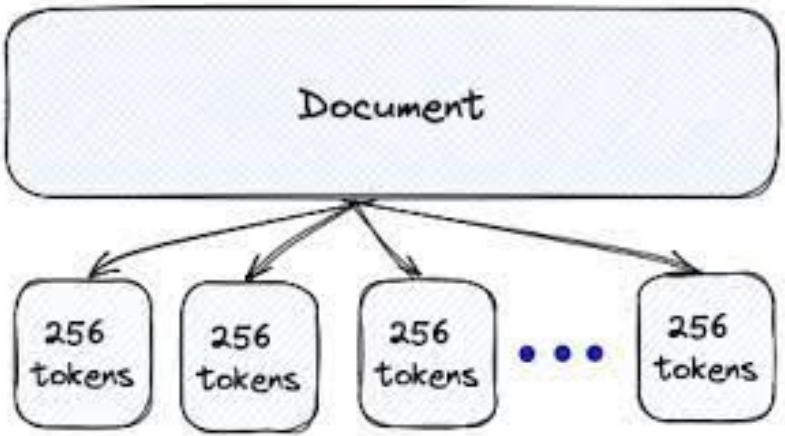


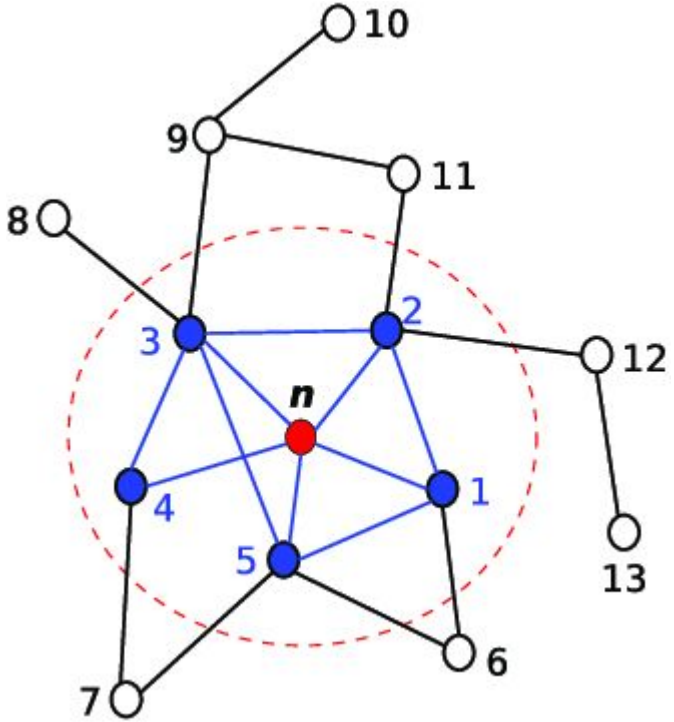


References:

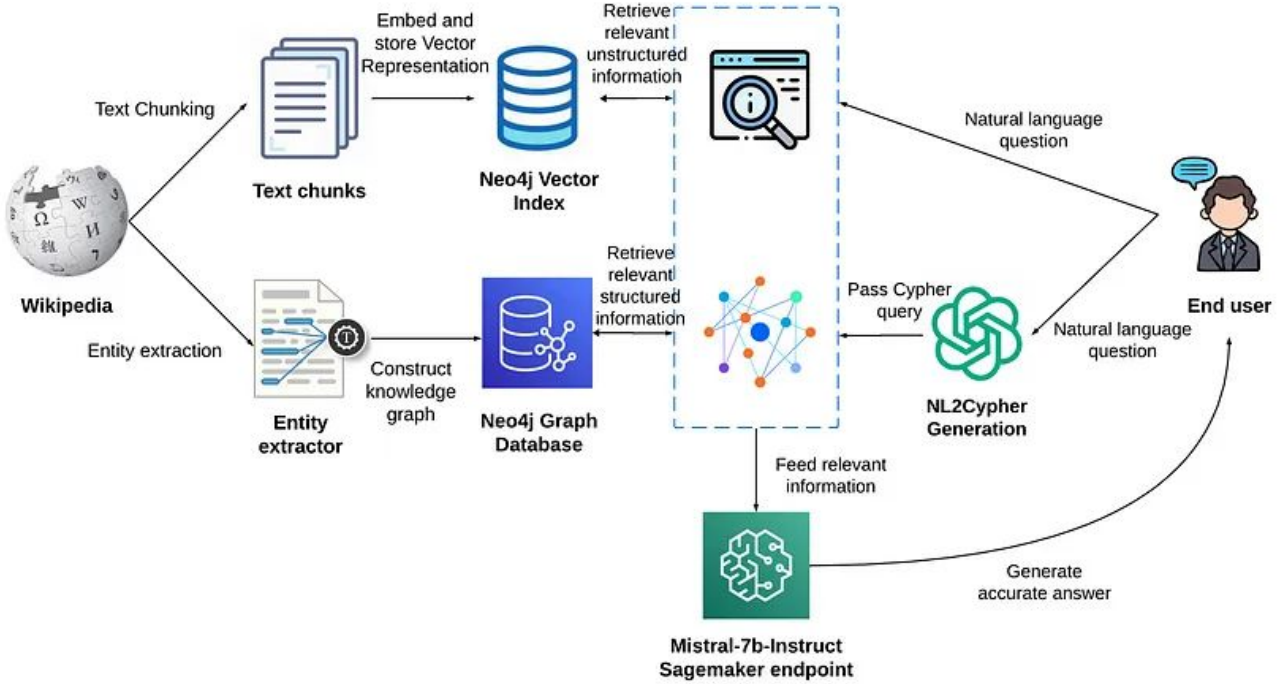
1. https://yashuseth.wordpress.com/2019/10/08/introduction-question-answering-knowledge-graphs-kgqa/
2. https://twitter.com/cohere/status/1639266554507407360

Benefits of KG vs Vector Search





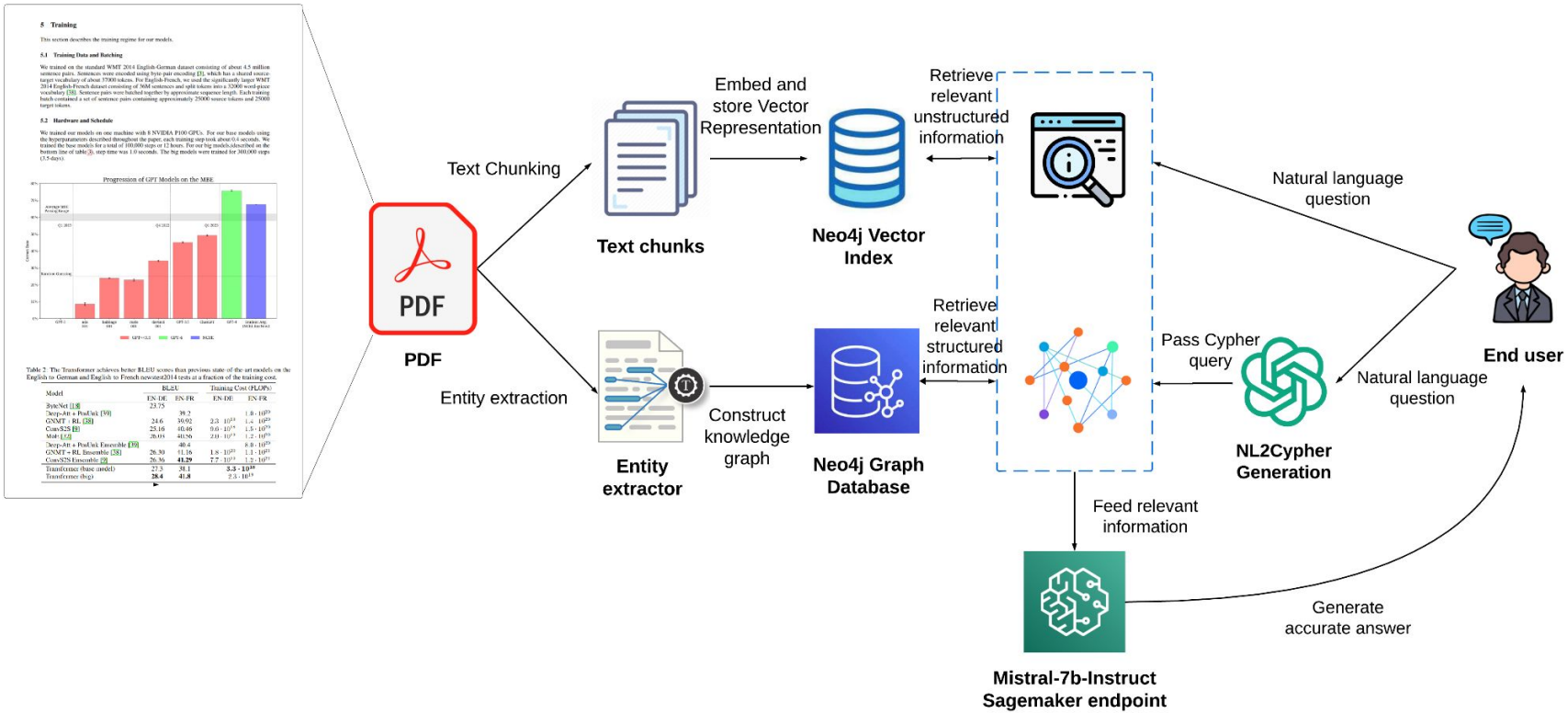
RAG Workflow



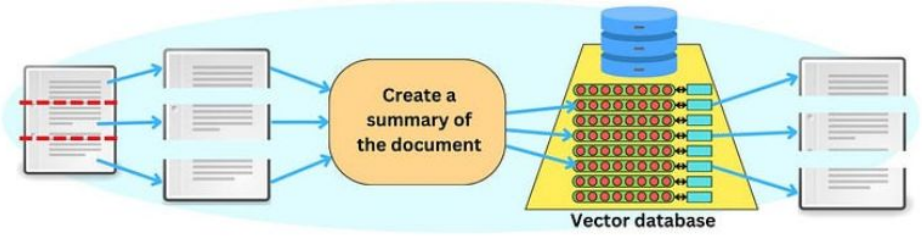
References:

1. https://medium.com/neo4j/enhanced-qa-integrating-unstructured-and-graph-knowledge-using-neo4j-and-langchain-6abf6fc24c27

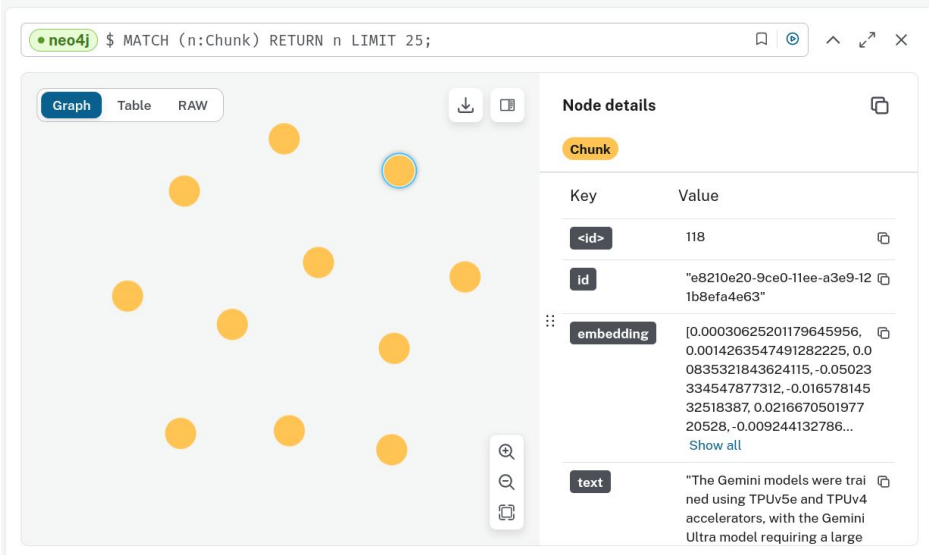
Modified RAG Workflow - Google Gemini PDF



Vector Search - Parent Child Retriever



Summarize text, tables, images(child) and index them in Neo4j Vector Index to represent specific concepts and store the actual documents(parent) in memory to represent context retention.

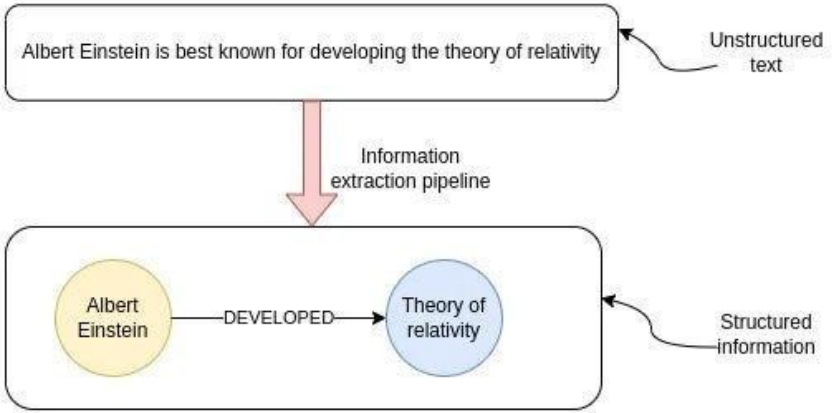


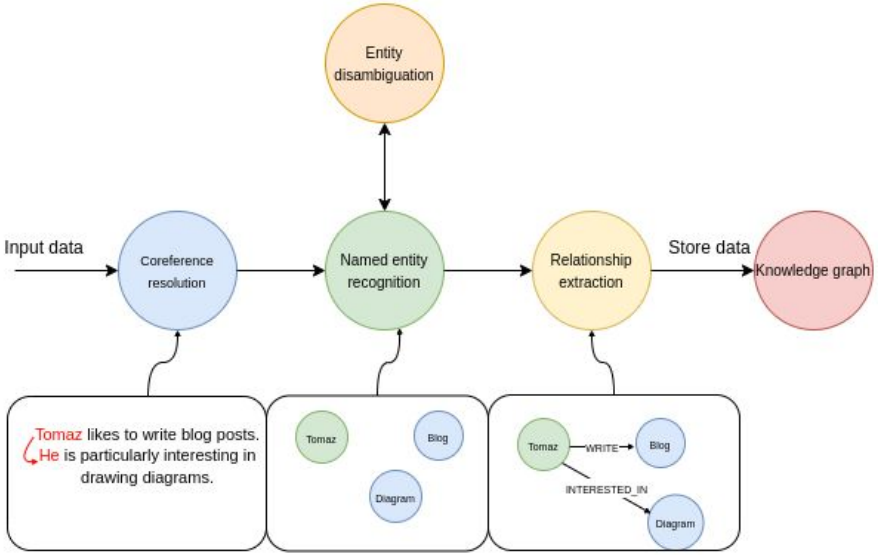
The summaries are indexed as nodes in a Neo4j Vector Index with the following properties - id, embedding, and summary text.

References:

1. https://TheAiEdge.io

Knowledge Graph Construction





References:

1. https://bratanic-tomaz.medium.com/constructing-knowledge-graphs-from-text-using-openai-functions-096a6d010c17

Knowledge Graph Construction - Image Integration

SYSTEM_MESSAGE = f"""# Knowledge Graph Instructions for GPT-4
1. Overview
You are a top-tier algorithm designed for extracting information in structured formats to build a knowledge graph.
2. Labeling Nodes
- **Consistency**: Ensure you use basic or elementary types for node labels.
3. Handling Numerical Data and Dates
- Numerical data, like age or other related information, should be incorporated as attributes or properties of the respective nodes.
4. Coreference Resolution
- **Maintain Entity Consistency**: When extracting entities, it's vital to ensure consistency.
5. Strict Compliance
Adhere to the rules strictly. Non-compliance will result in termination."""

SYSTEM_MESSAGE = f"""# Knowledge Graph Instructions for GPT-4
1. Overview
You are a top-tier algorithm designed for extracting information in structured formats to build a knowledge graph.
2. Labeling Nodes
- **Consistency**: Ensure you use basic or elementary types for node labels.
3. Identifying and Processing Tables
- **Table Detection: Identify tables by the keyword "Table" in text document.**
- **Entity and Relationship Extraction: From tables, extract entities and their relationships. Consider rows, columns, and headers for contextual understanding.**
4. Handling Image URI/Links
- **Mandatory Image URI in Each Node: Each node in the document must include an 'ImageURI' attribute. This applies to all nodes, regardless of their type or content.**
5. Handling Numerical Data and Dates
- Numerical data, like age or other related information, should be incorporated as attributes or properties of the respective nodes.
6. Coreference Resolution
- **Maintain Entity Consistency**: When extracting entities, it's vital to ensure consistency.
7. Strict Compliance
Adhere to the rules strictly. Non-compliance will result in termination."""

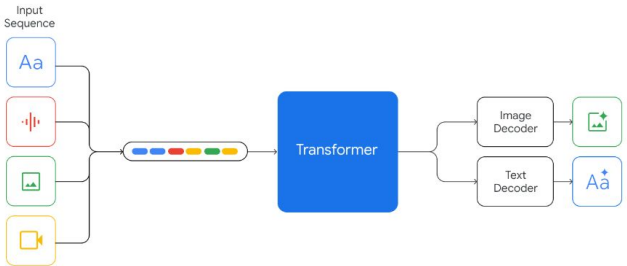
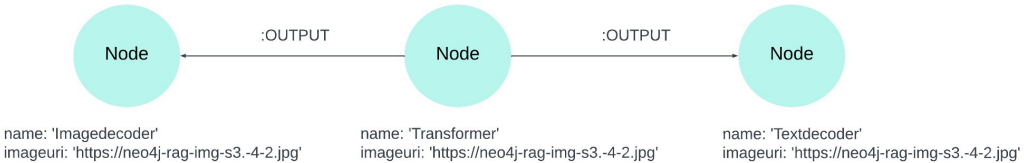
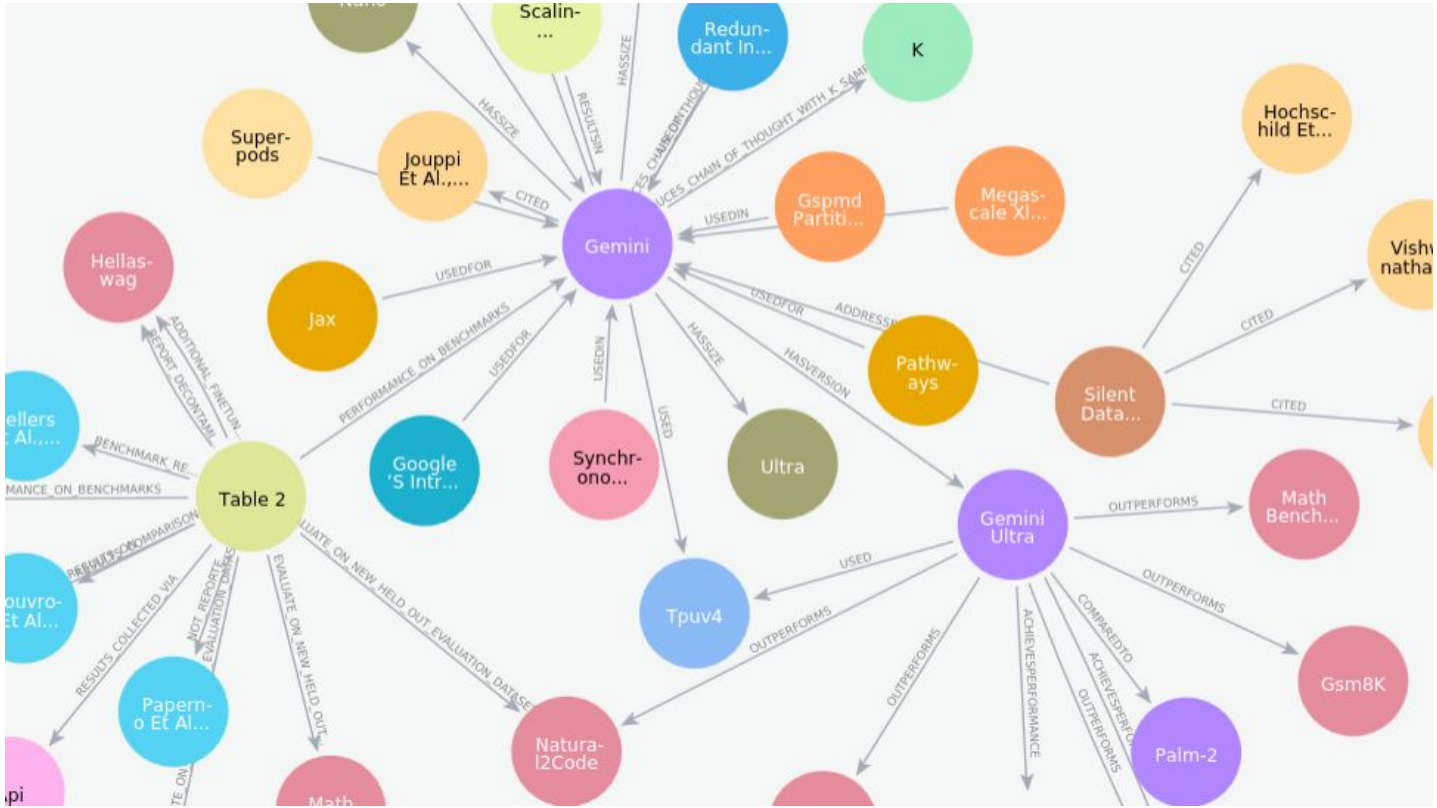


Image from PDF

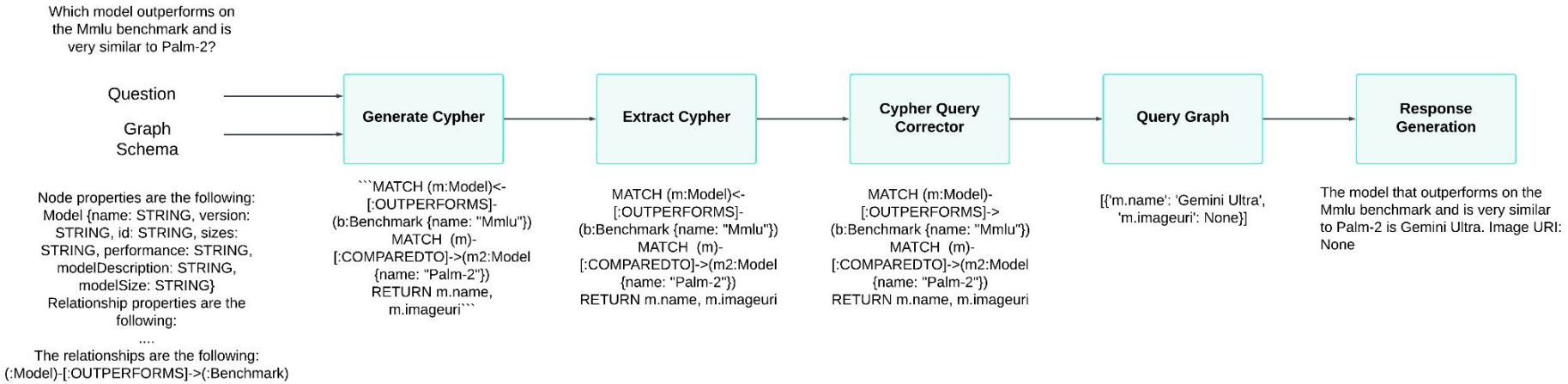


Graph Representation of an Image from the Image Text Summary. ImageURI are stored as properties for associated entities for context retention

Knowledge Graph Construction



Graph Search - GraphCypherQAChain



Code Walkthrough

- Integrated QA Neo4j Langchain semi structured data: https://github.com/sauravjoshi23/towards-agi/blob/main/retrieval%20augmented%20generation/integrated-qa-neo4j-langchain-semi-structured-data/main.ipynb
- Graph Construction: https://github.com/sauravjoshi23/towards-agi/blob/main/retrieval%20augmented%20generation/integrated-qa-neo4j-langchain-semi-structured-data/graph-construction.ipynb

Q&A