**Simplifying Insights Through Visual Representation - Knowledge Graph**

Global data has surged due to technological advancements and digital transformation, with unstructured data making up **80-90%** of the total. Includes text, images, video, and audio, is difficult to analyze using traditional methods.

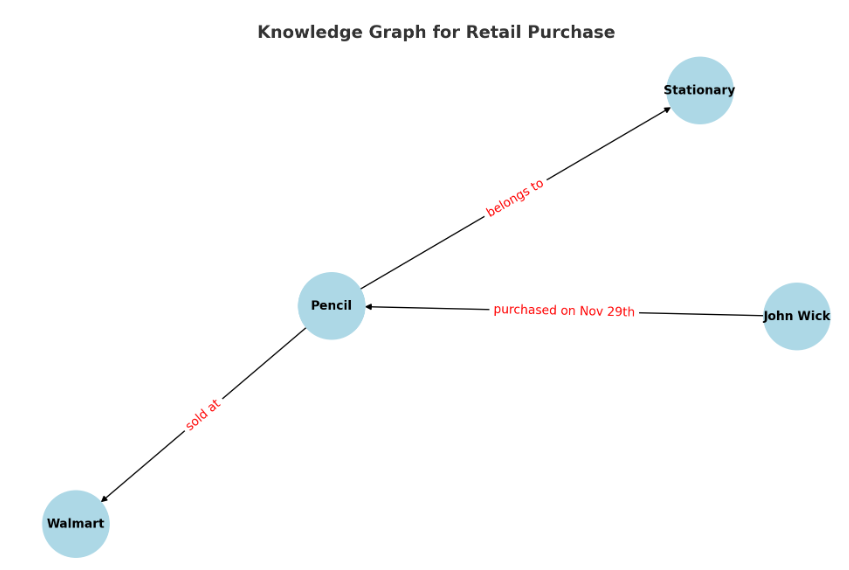
* Human mind process images faster than text, making visual representation essential for understanding complex data.
* Extracting insights from large unstructured text data is challenging without advanced tools.
* **Knowledge Graphs** simplify data interpretation by converting it into visually intuitive graphs, enhancing analysis and decision-making.

What is Knowledge Graph?

A knowledge graph is like a giant web of interconnected information. Illustrates, how different nodes (entities like people, places, or ideas) are related(edge) to each other and underlying patterns.

**Retail purchase on Black Friday:**

* **Nodes -** A customer - “John Wick”, the product - “Pencil”, the store “Walmart”, and the category “Stationary”
* **Edges -** Connect the customer to the product (indicating “purchased”)
* **Labels -** might specify “purchased on Nov 29th” for the purchase date.



**Applications:**

* Search Engines:

Enhance search results by providing contextually relevant information and direct answers (e.g., Google Knowledge Panel).

* Recommendation Systems:

Improve recommendations by understanding user preferences and relationships between entities (e.g., e-commerce, streaming platforms).

* Data Integration:

Knowledge graphs unify diverse datasets by linking entities and relationships, enabling seamless data access and improved contextual understanding.

* Question Answering Systems:

Knowledge graphs power question-answering systems by leveraging structured relationships to provide precise, context- aware answers to user queries.

Comparison between Traditional Approach vs LLM Approach

|  |  |  |
| --- | --- | --- |
| **Features** | **Traditional Knowledge Graphs** | **LLM-Based Knowledge Graphs** |
| **Core Methodology** | Built using fixed rules and relationships defined by experts. | Extracts knowledge directly from large amounts of text data. |
| **Flexibility** | Hard to update; changes need manual work. | Adapts automatically to new information. |
| **Knowledge Extraction** | Uses structured data like databases. | Can extract knowledge from unstructured data like plain text. |
| **Representation** | Stores knowledge as clear connections (nodes and edges). | Stores knowledge in hidden patterns inside the model. |
| **Querying** | Uses languages like SPARQL or Cypher to find information. | Allows natural language questions but less structured results. |
| **Scalability** | Needs a lot of effort to handle large data sets. | Handles large data easily with pre-trained models. |
| **Accuracy** | Accurate when data is clean and well-organized. | Sometimes makes errors or guesses information. |
| **Development Time** | Takes a long time to design and prepare. | Faster to start using pre-trained models. |
| **Data Dependency** | Needs well-organized and clean data. | Works with both clean and messy data. |
| **Use Cases** | Used in fraud detection, supply chains, and recommendations. | Used in chatbots, text summaries, and exploring data. |

Business case study - Flow

* Wikipedia - Download text on "Allan Turing"
* Using Python library, divide the text into chunks or way which is more suitable to pass text to create KG.
* Use LLM - Gemini to extract entities/nodes and the connections/edge with Turing.
* Use Neo4j instance to generate some labels based on the text.
* Visualize KG and Cypher query

