Future. Faster. Together.

PoV

Connecting the Dots:

Knowledge Graphs Transforming Pharma Production

Background

Life sciences companies have complex information technology (IT) and operational technology (OT) landscapes with multiple suppliers and various platforms, generating terabytes of data from procurement to product delivery. Organizations are investing significantly in robust technology infrastructure encompassing various hardware/software solutions such as asset management, RFID-based asset tracking, process control and execution systems.

Despite these efforts, only a small proportion of data is utilized to enhance the throughput and quality of finished goods; challenges due to siloed data, knowledge gaps, limited data visibility, etc., lead to disruptions in production cycles. Additionally, most data about procedures and process controls is usually unstructured and can often be found in standard operating procedures (SOPs), operation manuals, and hand-written notes.

Integrating, retrieving, and exploring this data presents a significant challenge, especially in pharmaceutical manufacturing. Semantic integration of structured and unstructured data is paramount to addressing these challenges, extracting valuable insights, and making informed decisions. This integration allows machines to understand context more effectively and leverage data from various sources. In today's Gen AI era, AI models perform exceptionally well in content generation tasks such as document summarization, content creation, answering user queries, etc. Yet, these models often struggle with decision-making and reasoning, as they lack the ability to understand the underlying meaning behind words.

Human decision-making inherently considers multiple factors, making it difficult for AI systems to comprehend and deliver suggestions at par with human-level reasoning. To this end, companies today need systems that integrate structured and unstructured data with advanced reasoning capabilities to bridge the 'semantic gap' in AI systems.

Knowledge graphs play an instrumental role in addressing this issue and enhancing the performance of AI systems by building a network of connected data. They can facilitate semantic search, context-driven retrieval, exploration, and reasoning by improving the ability to comprehend key life sciences terminologies and manage large-scale data.

What are knowledge graphs and ontologies?

Knowledge graphs (KGs) are structured models that organize large-scale data by mapping entities (nodes) and relationships (edges) between them. In simpler terms, they use a graph-structured model to integrate different data sources and represent complex relationships.

Let's start with an example. The KG of Reliefex, a pseudo brand used to relieve pain and headaches, may consist of nodes like 'oral solids,' 'tablet,' and 'acetaminophen,' connected by edges representing relationships such as 'subclass of' and 'has raw material.' Hence, the statement - Reliefex has acetaminophen - is represented by connecting the nodes (Reliefex and Acetaminophen) with the edge (has raw material).

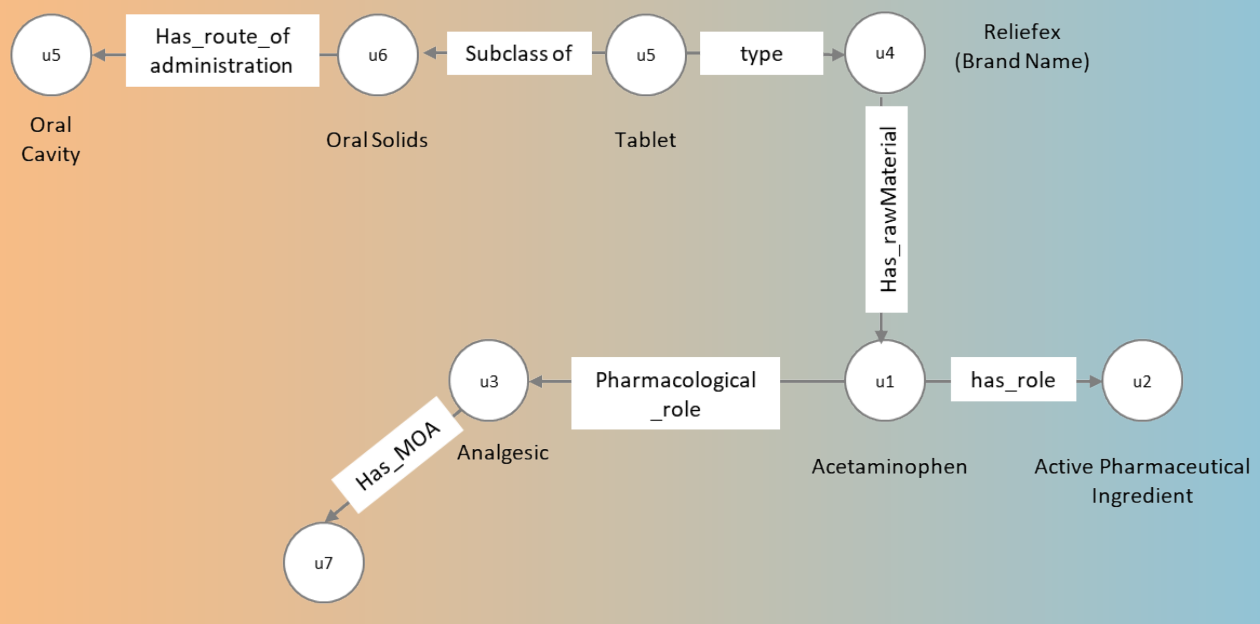

Fig. 1 – A Knowledge Graph

Ontologies are the backbone of KGs, providing a shared understanding of concepts and relationships. Ontologies enable advanced reasoning by allowing systems to understand class hierarchies and relationships, making it possible to infer associations between two data points. Picture this: If Reliefex is an instance of a tablet and tablets fall under oral solids, it is only natural that Reliefex is an oral solid as well (though there is no relationship between Reliefex and oral solids). Another key attribute of an ontology is its ability to address ambiguity and misunderstanding by aligning context and concepts with the precise meaning of words. By parsing relational information and concepts, it not only enhances query understanding but also significantly improves the quality of information retrieval.

Phases of knowledge graphs

The development of KGs unfolds into four critical phases: Data Acquisition, Knowledge Processing, Knowledge Graph Construction, and Smart Applications.

Data Acquisition:

Data acquisition involves extracting metadata from key enterprise and manufacturing applications like Enterprise Resource Planning (ERP), Manufacturing Execution System (MES), Supervisory Control and Data Acquisition (SCADA), Laboratory Information Management System (LIMS), etc. Additionally, entities and relationships are drawn from unstructured data like SOPs, training manuals, and operator notes with advanced AI/ML and NLP techniques. For structured sources, entities are identified and linked through multiple data integration platforms and APIs. This multi-faceted approach ensures data is gathered from multiple sources, providing a comprehensive view of the landscape.

Knowledge Processing:

Knowledge engineers and experts meticulously model different class hierarchies and define relationships and constraints to create domain ontology. This process involves rigorous validation to ensure conceptual clarity and consistency at each stage.

Depending on the objective, the ontology can be customized by adjusting the level of granularity. For example, simple tasks can benefit from a coarse-grained ontology, which provides adequate performance and speed due to its light size. On the other hand, tasks involving in-depth domain understanding like decision-support systems benefit from a fine-grained ontology that provides a high degree of information. It is important to note that as the size increases, the search space of the knowledge graph also grows, making it difficult to get accurate results. The ontology's correctness and quality are evaluated against five key criteria i.e. accuracy, clarity, completeness, computational efficiency, and consistency.

Knowledge Graph Construction:

High-quality KGs are constructed by mapping extracted entities and relations with respective ontologies. KGs are tested by evaluating their performance against a host of intricate graph queries such as factoid queries (fact-based questions), aggregate queries (summary values), and verification queries (verify specific information) by combining advanced AI/ML algorithms and Gen AI capabilities, the model learns about entities in natural language constructs and executes reasoning over knowledge graphs to yield explainable and precise responses. Using queries and algorithms such as graph-based pattern matching, instance matching, and graph walk, the model facilitates knowledge discovery, logical deduction, and advanced reasoning capabilities.

Smart Applications:

In manufacturing, knowledge graphs facilitate data integration by breaking down silos and creating a cohesive and interoperable model. This ensures that information from various sources, such as ERP, MES, and LIMS, is seamlessly combined, enhancing data visibility and analytics. Querying these graphs further allows fine-grained access to data. In information retrieval systems, ontologies are pivotal in enhancing user queries by semantically enriching text with domain-specific entities and attributes. This strategic augmentation significantly boosts the recall and precision of retrieval engines. Additionally, it mitigates the risk of hallucinations by grounding Large Language Model (LLM) responses in established structured knowledge.

Furthermore, knowledge graphs elevate AI by providing conversational reasoning capabilities, allowing production supervisors and operators to interact with the system more intuitively and enhancing the overall user experience.

KGs in pharma manufacturing

KGs excel at performing root cause analyses by traversing through interconnected nodes to uncover intricate dependencies within the manufacturing process. Strict quality control measures are implemented at each stage to ensure the safety, efficacy, and consistency of production.

Consider manufacturing process of tablets, that involves blending active ingredients with various excipients such as binders, disintegrants, pigments, etc., followed by milling, granulation, and compression into tablets. In this context, a quality control associate finds out that batch XX0012 shows an unexpected dissolution trend. Traditionally, the operator must check multiple systems to identify the root cause, which could be because of excess compression force, high binder volume, etc. This process takes time and can impact productivity and batch release.

On the other hand, using knowledge graphs, AI agents can traverse along the edges to visit related nodes and ascertain the root cause. The agent can extract the hardness and dissolution rate observations from the LIMS system and verify the patterns associating these entities with the root cause. By linking these findings with the compression force of the rotary tablet press, the AI agent can infer that the elevated compression force is the root cause of the slower dissolution rate.

Figure 2 showcases the path taken by the AI agent to establish the relationship between the compression force and the dissolution rate. The process reveals how intricate dependencies within the manufacturing process can be uncovered and analysed with KGs.

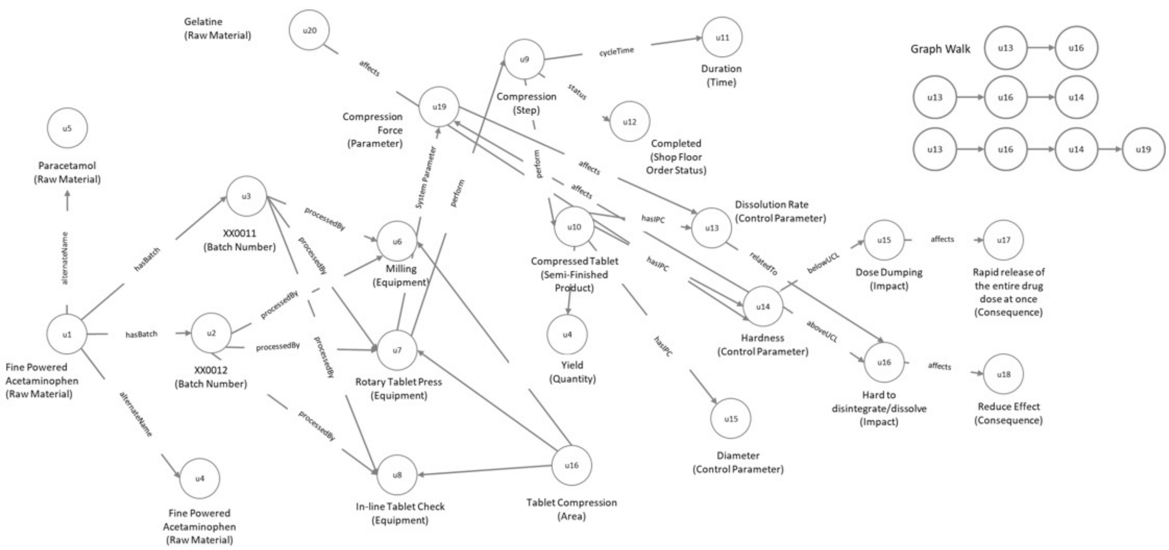

Fig.2 Root cause analysis for batch deviation

At LTIMindtree, we are transcending the conventional approach of merely creating wrappers around available LLMs or directly integrating raw data into RAG systems. Instead, we embed reasoning and domain-specific knowledge within these models by complimenting Gen AI with rich biomedical and industry semantics. By leveraging knowledge graphs, we are augmenting the capabilities of AI agents to align with the complexities of the life sciences industry.

Conclusion

Semantically integrating structured and unstructured information can significantly transform how we interact with data. By amalgamating data from the life sciences domain, processes, and technology, KGs provide a robust data foundation for Gen AI models, ensuring transparency and evidence-backed responses. Ontologies improve the logical thinking and reasoning abilities of Gen AI models by facilitating deductive reasoning and inferencing. Furthermore, KGs can be continuously enriched with new data, ensuring that AI systems remain current with the latest scientific discoveries and technological advancements.

Author profiles

Siddharth Joshi
Principal Director—Life Sciences

Siddharth is a seasoned professional with over 25 years of experience in IT, supply chain, and manufacturing. His successful stewardship of major accounts, coupled with collaborative engagements with CIOs and VPs of Fortune 100 companies, has resulted in the delivery of numerous transformative programs. With a strong background in both domain and IT, Siddharth has excelled in digital consulting, solution architecture, program management, and application development and maintenance.

Freny Rambhia
Senior Business Analyst—Life Sciences

Freny collaborates closely with clients, offering valuable insights, guidance, and support to deliver innovative solutions. With a proven track record of addressing diverse customer needs, she actively contributes to projects that drive business outcomes, consistently exceeding expectations. Her passion lies in exploring and studying new technologies.