ontotext

Knowledge graphs, while not as well-known as other data management offerings, are a proven dynamic and scalable solution for addressing enterprise data management requirements across several verticals. As a hub for data, metadata, and content, they provide a unified, consistent, and unambiguous view of data scattered across different systems. Using global knowledge as context for interpretation and a source for enrichment, they also optimize proprietary information so organizations can enhance decision-making and realize previously unavailable correlations between their data assets.

Organizations already know the data they need to manage is too diverse, dispersed, and at volumes unfathomable only a decade ago. This often leaves business insights and opportunities lost among a tangled complexity of meaningless, siloed data and content. Knowledge graphs help overcome these challenges by unifying data access, providing flexible data integration, and automating data management. The use of knowledge graphs

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

has an enormous effect on various systems and processes which is why Garner predicts that by 2025, graph technologies will be used in 80% of data and analytics innovations, up from 10% in 2021, facilitating rapid decision-making across the enterprise.

Knowledge Graphs Defined and Why Semantics (and Ontologies) Matter

According to Wikipedia, a knowledge graph is a knowledge base that uses a graph-structured data model or topology to represent and operate on data. The heart of the knowledge graph is a knowledge model – a collection of interlinked descriptions of concepts, entities, relationships, and events where:

- Descriptions have formal semantics that allow both people and computers to process them efficiently and unambiguously
- Descriptions contribute to one another, forming a network, where each entity represents part of the description of the entities related to it
- Diverse data is connected and described by semantic metadata according to the knowledge model

By creating a common semantic description, a knowledge graph enables a higher level of abstraction that does not rely on the physical infrastructure or format of the data. Sometimes referred to as a data fabric, it delivers a unified, human-friendly, and meaningful way of accessing and integrating internal and external data. Using semantic metadata, knowledge graphs provide a consistent view of diverse enterprise data, interlinking knowledge that has been scattered across different systems and stakeholders.

With the help of natural language processing (NLP), text documents can also be integrated with knowledge graphs. Given that many researchers say that between 75-85% of an organization's knowledge is locked in static documents, tremendous value and wisdom are being missed. NLP pipelines benefit enormously, as sophisticated text analysis methods can be used when combining machine learning with knowledge graphs. Knowledge graphs are also essential for any semantic AI and explainable AI strategy.

Ontologies are equally important, as they represent the backbone of the formal semantics of

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

a knowledge graph. As the data schema of the graph, they serve as a contract between the developers of the knowledge graph and its users regarding the meaning of the data. A user could be another human being or a software application needing to interpret the data in a reliable and precise way. Ontologies ensure a shared understanding of the data and its meanings. When formal semantics are used to express and interpret the data of a knowledge graph, there are several representation and modeling instruments:

- **Classes:** Most often, an entity description contains a classification of the entity concerning a class hierarchy. For instance, when dealing with general news or business information, there could be classes that include Person, Organization, and Location. Persons and organizations can have a common super-class agent. Location usually has numerous sub-classes, e.g., Country, Populated place, City, etc.
- **Relationships:** The relationships between entities are usually tagged with types, which provide information about the nature of the relationship, e.g., friend, relative, competitor, etc.
- **Categories:** An entity can be associated with categories that describe some aspect of its semantics, e.g., “Big Four consultants” or “XIX century composers.” A book can belong simultaneously to all these categories: “Books about Africa,” “Bestseller,” “Books by Italian authors,” “Books for kids,” etc. Often categories are described and ordered into a taxonomy.
- **Free Text:** It is possible to add “human-friendly text” to further clarify design intentions for the entity and improve search.

Knowledge Graphs in Resource Description Framework (RDF)

Resource Description Framework is a standard for describing web resources and data interchange, developed and standardized with the World Wide Web Consortium (W3C). Aside from RDF, the labeled property graph (LPG) model provides a lightweight introduction to the management of graph data. LPGs often win the hearts of developers if data needs to be collected ad hoc and graph analytics are performed in the course of a single project with the graph being discarded afterward. Unfortunately, the technology stack around LPGs lacks standardized schema or modeling languages and query languages, and there are no provisions for formal semantics and interoperability specifications (e.g., no serialization

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

formats, federation protocols, etc.).

While RDF allows statements to be made only about nodes, RDF-Star allows one to make statements about other statements and in this way attach metadata to describe an edge in a graph such as scores, weights, temporal aspects, and provenance. All in all, knowledge graphs, represented in RDF, provide the best framework for data integration, unification, linking, and reuse, because they combine the following:

1. **Expressivity:** The standards in the Semantic Web stack – RDF(s) and OWL – allow for a fluent representation of various types of data and content: data schema, taxonomies, vocabularies, all sorts of metadata, reference, and master data. The RDF-star extension makes it easy to model provenance and other structured metadata.
2. **Formal semantics:** All standards in the Semantic Web stack come with well-specified semantics, which allow humans and computers to interpret schema, ontologies, and data unambiguously.
3. **Performance:** All specifications have been thought out and proven to allow for efficient management of graphs of billions of facts and properties.
4. **Interoperability:** There is a range of specifications for data serialization, access (SPARQL Protocol for end-points), management (SPARQL Graph Store), and federation. The use of globally unique identifiers facilitates data integration and publishing.
5. **Standardization:** All the above is standardized through the W3C community process, to make sure that the requirements of different actors are satisfied –from logicians to enterprise data management professionals and system operations teams.

However, it’s important to note that not every RDF graph is a knowledge graph. For instance, a set of statistical data, e.g. the GDP data for countries, represented in RDF is not a knowledge graph. A graph representation of data is often useful, but it might be unnecessary to capture the semantic knowledge of the data. It might also be sufficient for an application to just have the string “Italy” associated with the string “GDP” and the number “$1.95 trillion” without needing to define what countries are or what the Gross Domestic Product of a country is.

It’s the connections and the graph that make the knowledge graph, not the language used to represent the data. A key feature of a knowledge graph is that entity descriptions should be interlinked to one another. The definition of one entity includes another entity. This linking is

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

how the graph forms (e.g., A is B; B is C; C has D; A has D). Knowledge bases without formal structure and semantics, e.g., Q&A “knowledge base” about a software product, also do not represent a knowledge graph. It is possible to have an expert system that has a collection of data organized in a format that is not a graph but uses automated deductive processes such as a set of “if-then” rules to facilitate analysis.

Knowledge graphs are not software either. Rather a knowledge graph is a way to organize and collect the data and metadata to meet criteria and serve specific purposes which, in turn, is used by different software. The data of one knowledge graph can be used in multiple independent systems for different purposes.

Knowledge Graphs and Real-Time Data Management

The demands on our data have pushed traditional approaches to data management past their limits. There are copious amounts of data, more every day, and it all needs to be processed, understood, and made useful. It needs to be reliable and done in real-time regardless if it is coming from internal or external sources. After all, the value of data depends wholly on the ability to leverage its use. This is a lesson that organizations are learning quickly as they seek to reduce development and maintenance costs and come to appreciate the advantages and revenue to be gained by intelligently managing organizational data. Today’s data ecosystems are also global.

Knowledge graphs can deal with their diversity and the lack of centralized control because it is a paradigm suited to the global data ecosystem that includes every organization. Better yet, as the information and an organization’s understanding and needs from that information change, so does the knowledge graph. The data represented by a knowledge graph has a strict formal meaning that both humans and machines can interpret. That meaning makes it usable to a human but also allows automated reasoning to enable computers to ease some of the burden. With knowledge graphs, organizations can change, prune, and adapt the schema while keeping the data the same and reusing it to drive even more insights.

Years ago, we moved away from the buzzword of Big Data to Smart Data. Having unprecedented amounts of data pushed the need to have a data model that mirrored our complex understanding of information. To make data smart, machines could no longer be bound by inflexible and brittle data schemas. They needed data repositories that could

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

represent the real world and the tangled relationships that it entails. All this needed to be done in a machine-readable way with formal semantics to enable automated reasoning that complemented and facilitated human expertise and decision-making.

Knowledge graphs expressed in RDF provide this as well as numerous applications in data and information-heavy services. Examples include intelligent content, packaging, and reuse; responsive and contextually aware content recommendation; automated

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com