There's been a lot of criticism that knowledge graphs are too complex. Of course, they are complex. But so are rockets. Yet, you don't expect to be able to get to the moon on a bike, do you? Unless you already have ET riding with you.

In Computer Science, we are trained to use the Occam's razor – the simplest model of reality that can get the job done is the best one. So, why do we recommend knowledge graphs, which are perceived to be complex, to our customers? The reason is that the inherent complexity of big enterprises is such that this is the simplest model that enables them to "connect the dots" across the different operational IT systems and turn the diversity of their business into a competitive advantage.

In this post, which is a matured version of my opening keynote at Ontotext's Knowledge Graph Forum 2023, I will start with evidence about the impact of complexity on the growth and efficiency of big enterprises. Next, I will explain how knowledge graphs help them to get a unified view to data derived from multiple sources and get richer insights in less time. Finally, I will elaborate on the major design patterns and the requirements for an enterprise-wide knowledge graph management platform.

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

Recognizing the inherent complexity of global businesses

> Complexity isn't bad in and of itself. Often, it's a by-product of business growth. The larger an organisation, the more complex it will naturally be, needing more people and technologies to serve a growing customer base – PwC, Is your organisation too complex to secure?

Often, an enterprise starts with one thing it does well and then adds more business lines to expand the market. This requires new tools and new systems, which results in diverse and siloed data. A customer research from Okta, found that since 2016, the average number of applications deployed in organizations has increased by 24%. It also established that organizations of over 2,000 employees use close to 200 internal applications. This scales to thousands of different IT systems and tools for some of the big enterprises we work with. In such an environment, it is a massive challenge to integrate information across various

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

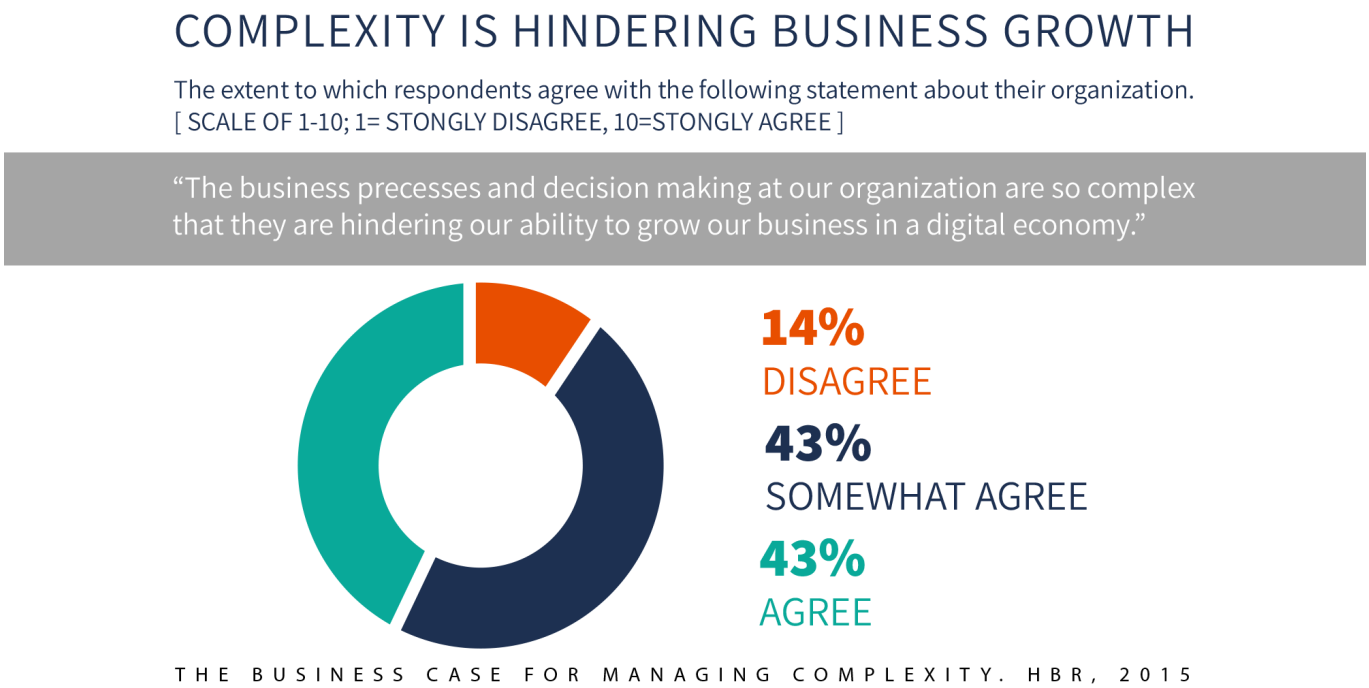
functions and units and their specific operational IT systems. Still, enterprises must connect the dots across their business, if they want to be competitive. As organizations grow, they face a range of inefficiencies – for example, management costs go up and decision-making slows down due to deeper hierarchies. Large enterprises must compensate for this by leveraging synergies, economies of scale, cross-selling, etc. And each of these gains requires data integration across business lines and divisions.

Limiting growth by (data integration) complexity

Most operational IT systems in an enterprise have been developed to serve a single business function and they use the simplest possible model for this. Any type of metadata or universal data model is likely to slow down development and increase costs, which will affect the time to market and profit. As a result, most of the data integration work usually takes place after the initial system design and implementation and it's considered to be… somebody else's problem.
As diversity grows, integration gets harder and the web of interconnected data, technology, and processes becomes very convoluted and difficult to manage. It's a challenge to determine even simple things like how elements impact each other or which ones are vital to the business and which aren't. This is confirmed by a Harvard Business Review analysis showing that complexity hinders business growth. It covers close to a thousand organizations and over 80% of the people agree that complexity is an issue.

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

ontotext



A slightly different angle to the same problem is discussed in a recent Ontotext blog post. For most organizations being competitive means they need to adapt and be able to source all relevant information. So, they become very data-driven. But instead of getting ahead, they are still missing out on key insights for better business decisions or more efficient processes.

> ...in the race to become data-driven, most efforts have resulted in a tangled web of data integrations and reconciliations across a sea of data silos that add up to between 40% – 60% of an enterprise's annual technology spend. We call this the Bad Data Tax.

So, how to manage this complexity better?

Maintaining integrity and leveraging complexity

As global businesses have to analyze data across different systems, they need a holistic view that allows them to perform deeper analysis and get richer insights. No matter what machine learning or graph algorithms are used, they cannot uncover dependencies if the

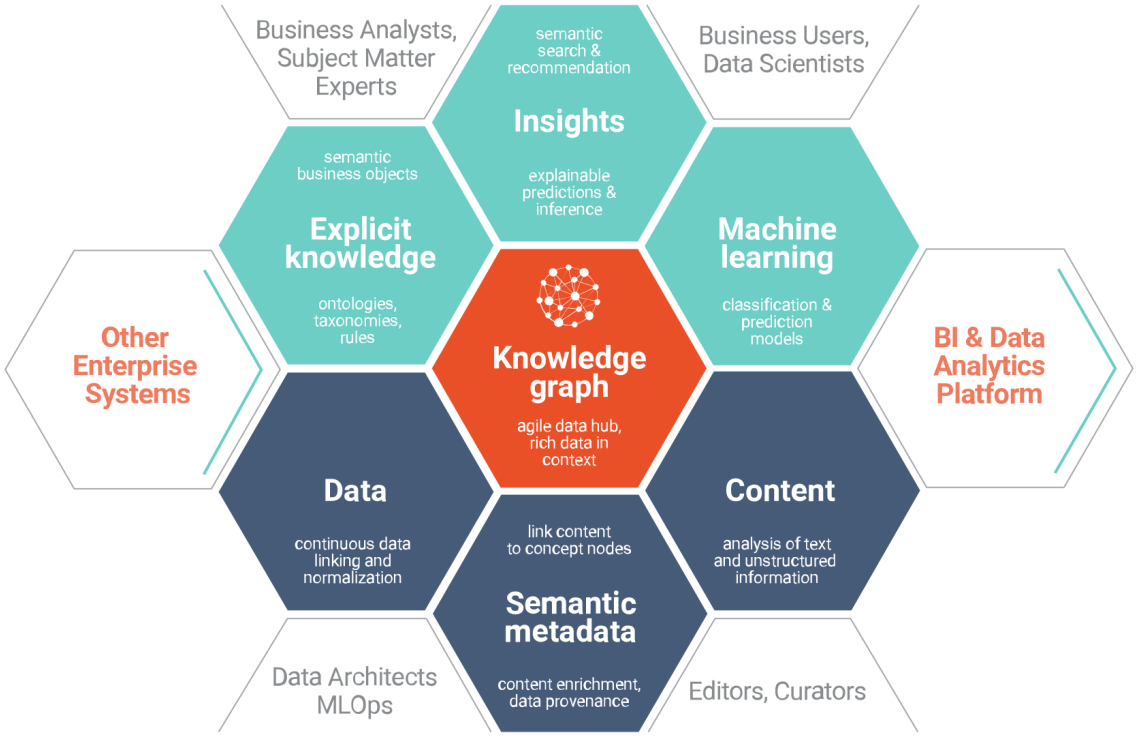
Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

corresponding “signals” are missing. And, quite often, these signals are spread across multiple systems.
Knowledge graphs are designed to provide a unified view to diverse data. They can lower the entropy in the convoluted IT ecosystems of big enterprises, as Tony Seale pointed out in a recent post. So, to maintain integrity and leverage complexity, big organizations need an enterprise-wide knowledge graph platform.

As global businesses have to analyze data across different systems, they need a holistic view that allows them to perform deeper analysis and get richer insights. Share on X

The need for a knowledge graph platform

At a very high level, a knowledge graph platform has to connect the data into a reusable graph to power deep analytics. This will allow enterprises to derive insights from their disparate databases, integrate external knowledge for better interpretation of their data, and incorporate information extracted from unstructured content.

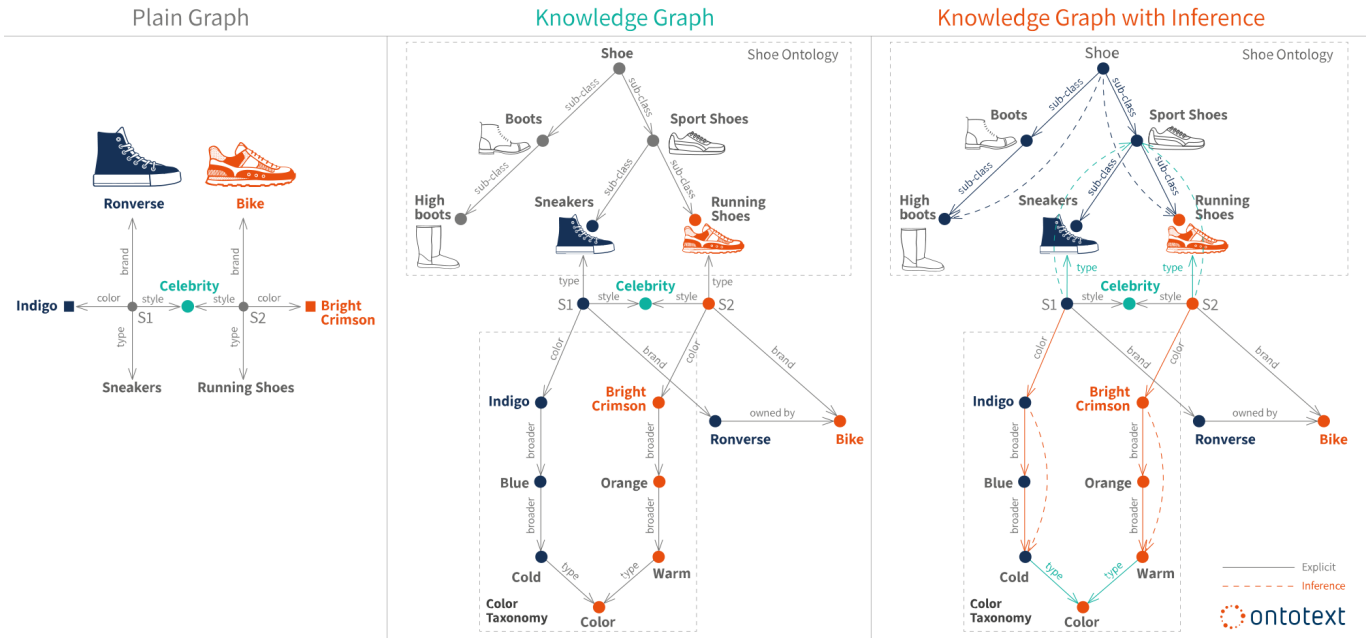


In order to integrate structured data, enterprises need to implement the data fabric pattern.

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

It will enable them to get data shared and consumed via marketplaces, while data ownership is distributed. This is the data mesh paradigm that helps them improve both the usability and the quality of their key data assets. When unstructured content is considered, knowledge graphs and text analysis have been "dancing together" for quite some time. We can use text analysis to populate and enrich knowledge graphs but also use knowledge graphs as a source of context and domain knowledge to improve the performance of text analysis pipelines. In both cases, semantic metadata is the glue that turns knowledge graphs into hubs of data, metadata, and content.

When we say semantic metadata, what we mean is having rich machine reasonable descriptions of information. The diagram below illustrates this in a simplified form. It shows two pairs of sports shoes and how their description can vary in richness. When there's just data without context, we can't get many insights out of it. For example, we wouldn't be able to derive insights about the kind of people who prefer shoes in cool colors because in a plain graph color attributes like "indigo" and "bright crimson" are only strings. To be able to derive such insights from data, they need to be put in the context of a color taxonomy. Then it will show that indigo is a type of blue, which is a cool color, while bright crimson is a type of orange, which is a warm color.



At the end of the day, we'll be looking at the same sports shoe catalog, but in a knowledge graph, it will be enriched with ontologies and taxonomies. The context and the extra

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

knowledge that comes with it enable deeper analytics. And, if we have a semantic graph database, we can also use inference, to interpret the meaning of the different relationships. We can also materialize "shortcuts" that can dramatically improve both the analytical power of the queries and speed up their evaluation. This will allow us to perform quicker slicing and dicing and to get richer results in less time.

The context and the extra knowledge that comes with ontologies and taxonomies enriching a knowledge graph enable deeper analytics. Share on X

RDF or LPG or … both?

When using graph technology, architects need to decide whether to use the RDF technology stack or labeled property graphs (LPGs). Each of these models has its advantages and disadvantages depending on the task at hand. Overall, RDF graphs are much finer-grained and enable better data governance and flexibility, while LPGs have proven to be more efficient in some graph analytics tasks.

Use Case	RDF	LPG
Data scientists working on concrete data analytics project in an enterprise *"At present LPG solutions seem to offer more options, and perhaps better performance - although properly evaluating performance is a very complicated issue."*	+	++
Architects designing enterprise data management systems *"Schema support and metadata management are crucial aspects of enterprise data management systems. RDF has advantages in both of these areas"*	++	-
Architects designing content management systems *"The need to manage URIs and metadata, as well as perform SEO functions, makes this use case an ideal match for RDF"*	++	-
Data governance and system operations teams *"Dependency tracking can be dealt with equally well by LPG and RDF solutions. When it comes to metadata management, however, we believe this is an area in which RDF solutions have traditionally emphasized, and excelled in"*	++	-
Developers building applications using such database engines *"The RDF data model can be used ... However, LPG has a significant head start there."*	+	++

Source: Graph Database Market & Technical Analysis 2021. Linked Data Orchestration, George Anadiotis

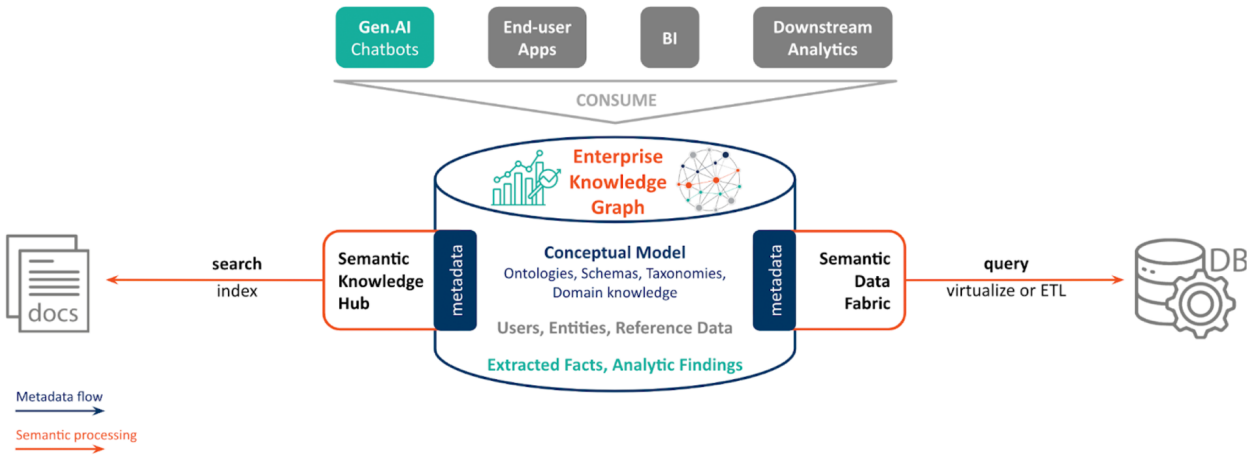
The distribution of use cases best served by RDF graphs and LPGs has been well-known for years. What's new is that people have just started to realize that if they want to have an enterprise-wide knowledge graph platform that serves all of these use cases, they need both

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

RDF and LPG support. So, it's not either/or, it's both/and.

Looking at the big picture

In the following progression of diagrams, I will present an outline of an enterprise-wide knowledge graph platform and the interplay between the different tools, engines, and legacy systems.



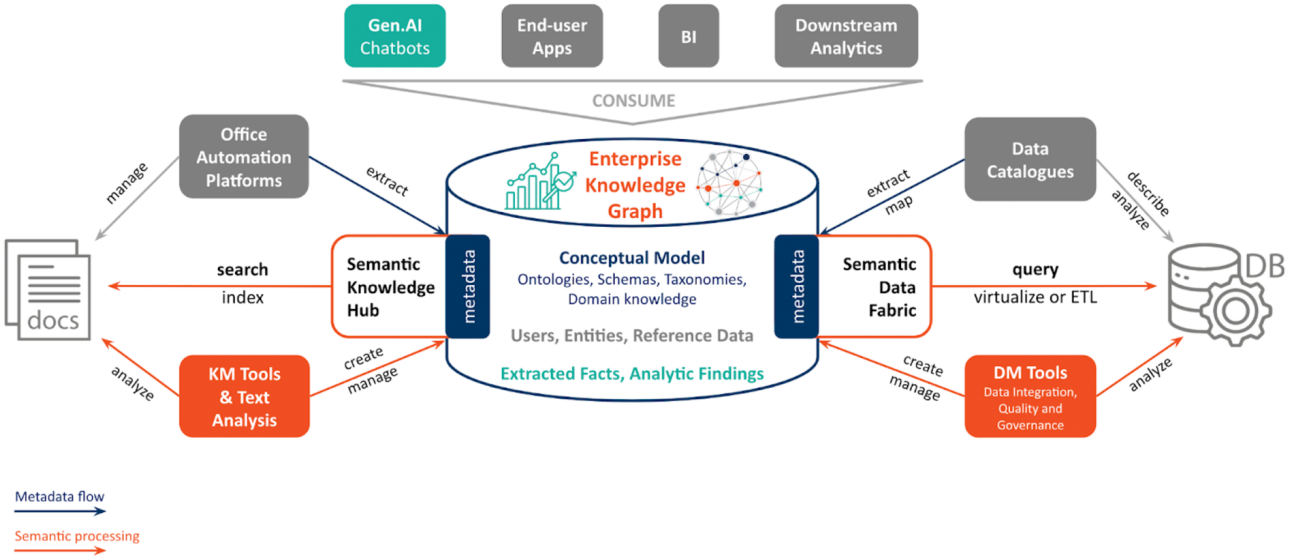
As a start, such a platform would need to support two major design patterns: **semantic knowledge hub** and **semantic data fabric**. We call knowledge hub the pattern of using knowledge graphs to better manage documents and unstructured content. This helps improve the way relevant documents are found, recommended, and so on. Data fabric, on the other hand, is the pattern that provides better unified access across multiple databases and its ultimate objective is to enable querying them all as if they were one database.
Both use cases use semantic metadata, which describes information sources with respect to a unified **conceptual model**, that includes ontologies, data schema, taxonomies, reference data, or other domain knowledge. In this way, knowledge graphs make it easy to discover, analyze, and interpret information based on its meaning, even when it's sourced from hundreds of IT systems, as Gregor Wobbe, Head of Data Architecture of UBS, presented at KGC 2023.

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

ontotext

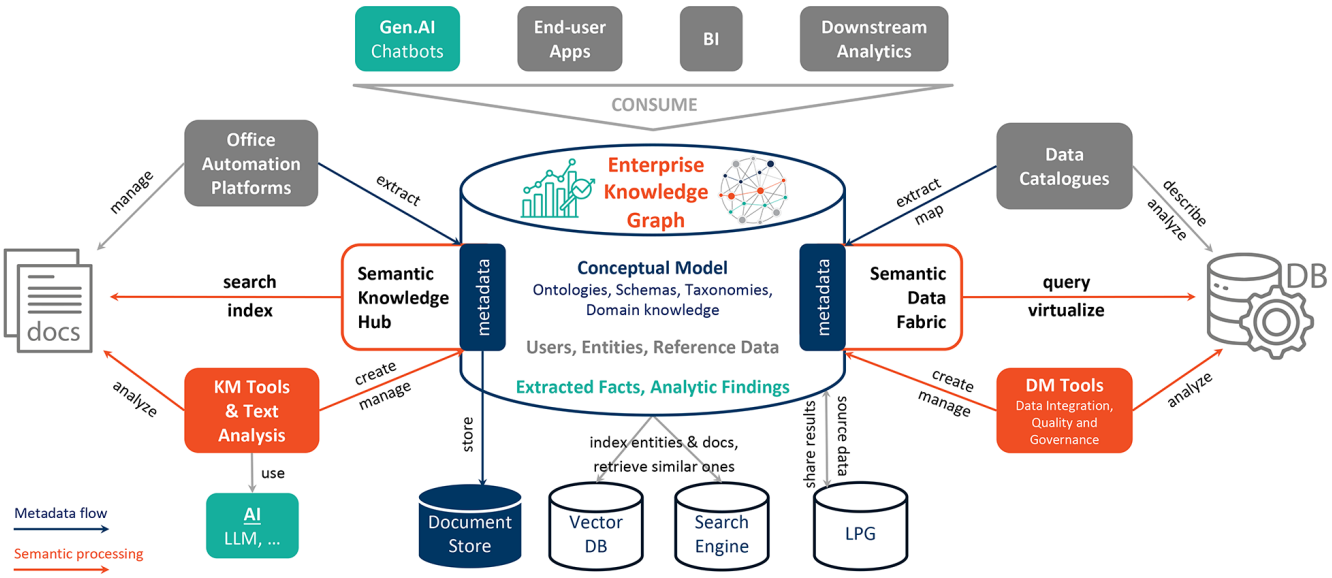
The required ingredients

If we want to implement a semantic knowledge graph hub, we need **knowledge management** and **text analysis tools**. They sift through documents, generate metadata, and store it in the knowledge graph. To implement a data fabric pattern, on the other hand, we need **data management tools** for data integration, data quality, and data governance. They analyze the individual databases and create metadata describing the different data sources so that we can query them as if they were one database.



Of course, a lot of metadata is already available and we must learn to use it. This includes **office automation platforms** like Microsoft SharePoint or Google Suite as well as **data catalogs** that already describe the different data assets. The only thing we need to do is to "semantify" this metadata by mapping it to our conceptual model.
Finally, to be able to manage and exploit this knowledge graph efficiently, we need a range of different engines and AI models. These include **large language models and other machine learning tools** for text analysis, **document stores** for storing big volumes of document metadata, **full-text search engines** and **vector databases** for retrieving similar entities and similar documents, and **LPG engines** for complex graph analytics.

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



Choke points and how we address them

So, what are the difficulties in making such a knowledge graph platform work?
To start with, we need a database engine with predictable performance across multiple workloads. It has to have support for semantics at scale to perform inference and data validation. It has to be able to manage and share standards-based metadata and offer a range of graph analytics capabilities. It also needs LLM and vector database integration to properly interact with various AI tools and models.
We at Ontotext have dedicated 20 years to developing GraphDB so that it meets all these requirements. I am proud to share that GraphDB is the first database engine that has passed both the Social Network Benchmark and the Semantic Publishing Benchmark of the Linked Data Benchmarking Council (LDBC). The former is about graph analytics and the latter is about semantic metadata management. GraphDB can handle both with excellence, which positions it as the most versatile graph database engine.
In terms of text analysis, we need an ecosystem that allows us to experiment with multiple text analysis models and services, including LLMs. We need to be able to measure the accuracy of these services and, once in production, manage the quality of the results. To deliver on the promises for meaning-based content management, we also need to be able to link the entities and the concepts mentioned in the documents to the right ones in our

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

knowledge graph, which is called entity linking. This task is something that LLMs don't do well.
To address these challenges, we offer Ontotext Metadata Studio. It integrates with GraphDB and other Ontotext products as well as the knowledge management tools of our partners. It provides out-of-the-box ready integration with many annotation services and text analysis services. We can also offer four new separate entity-linking models, so we can choose the most efficient one for our customers' specific use cases.

To wrap it up

We started this post with the inherent complexity of knowledge graphs and the need for a knowledge graph platform in big enterprises. We know that it can be overwhelming to try and do everything all at once. So here's our advice. If you are a big enterprise, start with implementing a data fabric and choose your first use cases carefully. If you are a small to medium enterprise, it might be more practical to get access to your unstructured documents with the knowledge hub pattern.

Still not sure how to start?

Book a Discovery Call

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