ONTOFORCE

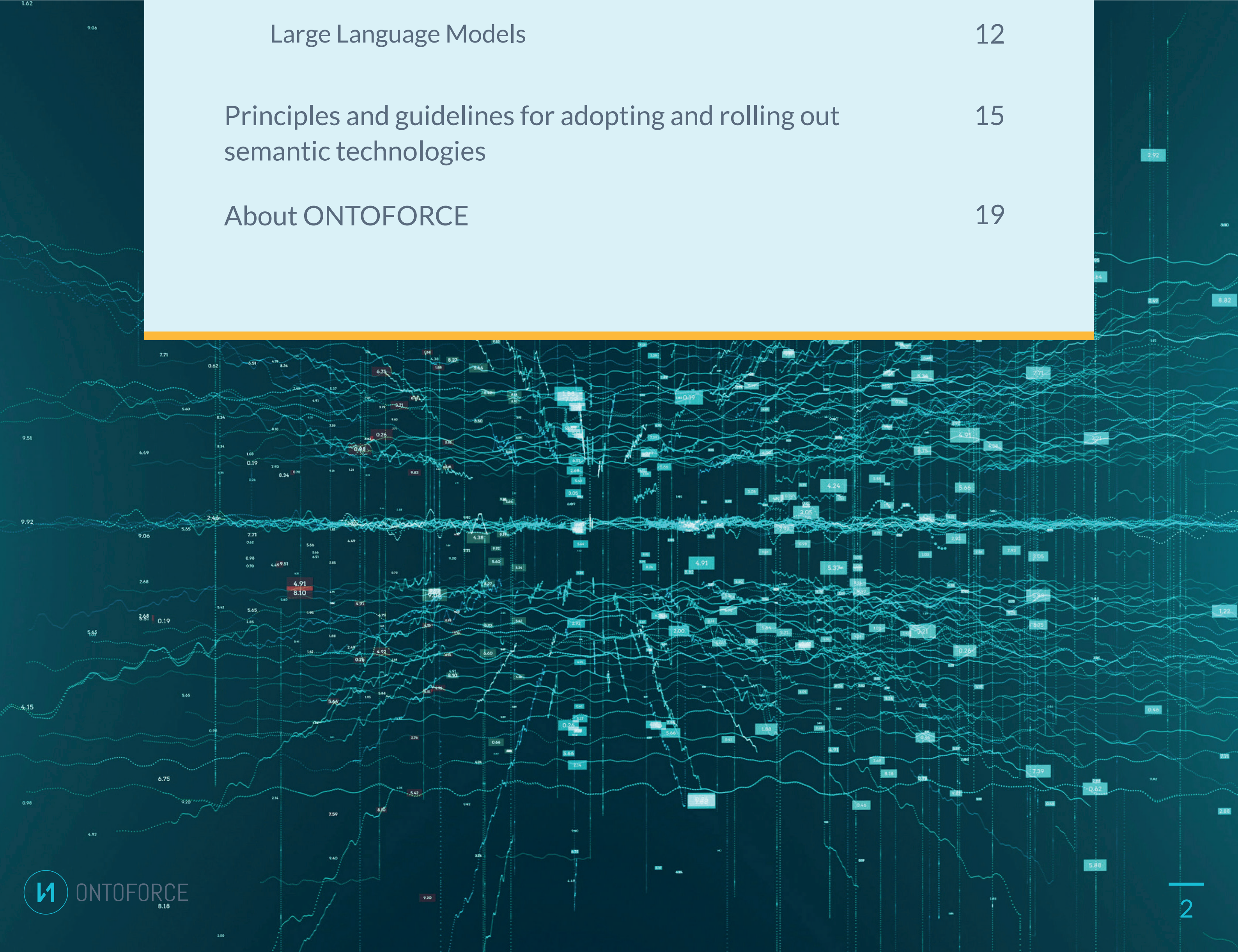
Linking data to unlock insights: the building blocks of semantic technology

AN INTRODUCTION TO SEMANTIC TECHNOLOGY FOR THE LIFE SCIENCES INDUSTRY

ebook

ontoforce.com
hello@ontoforce.com

Table of contents

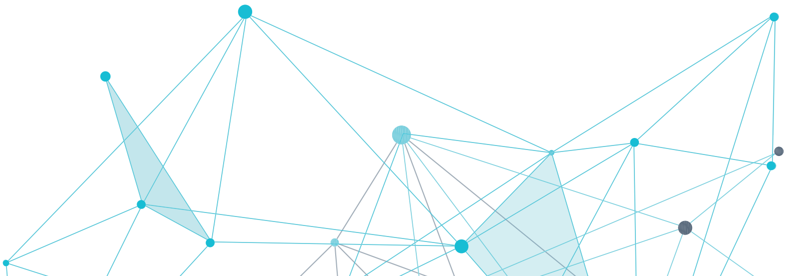


Introduction

Semantics refers to the branch of linguistics and logic concerned with meaning. It focuses on the interpretation of words, phrases, signs, and symbols within language, and how these elements convey messages, ideas, and concepts in different contexts. Semantics is crucial for understanding human language. Semantic technology is grounded within the field of semantics and refers to a range of tools, standards, and methodologies designed to enable machines to understand and interpret the meaning or context of information on the web or within data sets, much like humans do.

In the life sciences industry, semantic technology plays a pivotal role in managing and interpreting the vast amounts of complex and heterogeneous data generated by drug development activities. Semantic technology provides a framework for integrating, sharing, and analyzing data across various biological, chemical, and medical domains. This enables researchers and healthcare professionals to uncover hidden relationships and patterns within data, facilitating advancements in drug discovery, personalized medicine, and patient care.

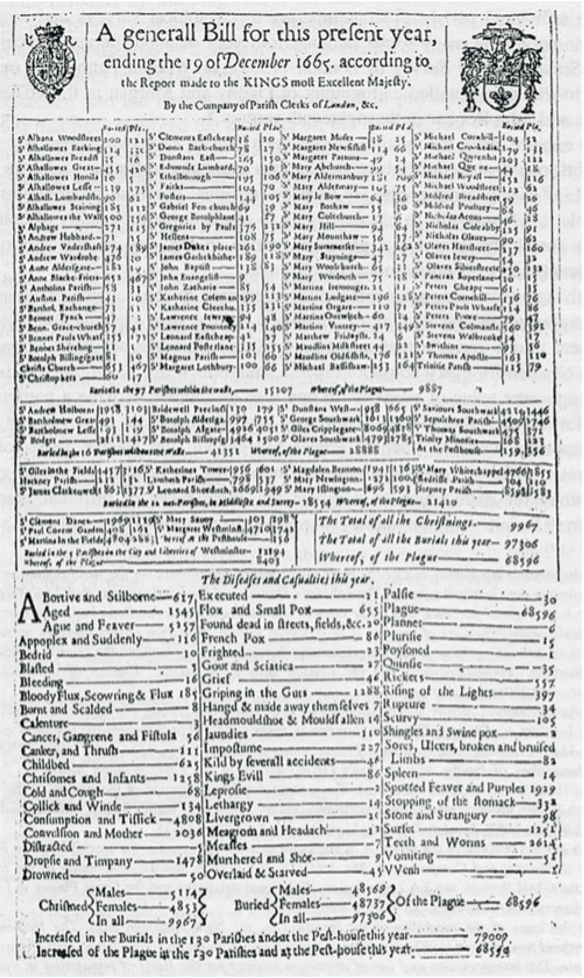
In this eBook, we'll have a look back at the origins of semantic technology, explore the relevant modern applications of semantic technology for the life sciences industry, and dive into essential insights for organizations looking to embark on their journey with adopting and rolling out the technology.



The origins of semantic technology

The history of semantic technology starts a few hundred years ago, although it wasn't really technology at that point. Around the mid 1500s, cities in Europe were becoming larger and larger. London provides an insightful example. The city grew from approximately 50,000 to 500,000 people in the span of just a few decades. This jump in population caused considerable problems for government officials as they tried to figure out how to appropriately accommodate this growth. They were concerned with housing people, treating sick people, burying people, and so on. These concerns and planning challenges brought an increased interest in medical statistics, particularly, an interest in what people were dying of.

To investigate this further, the government in London employed lay people to attend the funerals that were taking place around the city and the city limits to ask questions to the relatives of the deceased about the cause of death. The descriptions of the various causes of death captured were vivid but weren't necessarily useful from a scientific perspective, such as "found dead in fields" and "teeth in worms." This recording and publishing of the numbers of deaths broken down by cause and age in London and throughout other major cities support early public health and city planning efforts, and represent an early foundation to medical statistics, later paving the way for semantic technology.



A generall Bill for this preſent year,
ending the 19 of *December* 1665. according to
the Report made to the KINGS moſt Excellent Majeſty.

By the Company of Pariſh Clerks of *London*, &c.

	Buried	Pla.		Buried	Pla.		Buried	Pla.		Buried	Pla.
St Albans Woodſtreet	200	121	St Clements Eaſtcheap	38	20	St Margaret Moſes	38	25	St Michael Cornhill	104	52
St Alhallowes Barking	514	330	St Dionis Back-church	78	27	St Margaret Newfiſhſt	114	66	St Michael Crookedla.	179	133
St Alhallowes Breadſt	35	16	St Dunſtans Eaſt	265	150	St Margaret Pattons	49	24	St Michael Queenhit	203	122
St Alhallowes Great	455	426	St Edmonds Lumbard.	70	36	St Mary Abchurch	99	54	St Michael Quee ne	44	18
St Alhallowes Honila	10	5	St Ethelborough	195	106	St Mary Aldermanbury	181	109	St Michael Royall	152	116
St Alhallowes Leſſe	239	175	St Faiths	104	70	St Mary Aldermary	105	75	St Michael Woodſtreet	122	62
St Alhall. Lumbardſtr.	90	62	St Foſters	144	105	St Mary le Bow	64	36	St Mildred Breadſtreet	59	26
St Alhallowes Staining	185	112	St Gabriel Fen-church	69	39	St Mary Bothaw	55	30	St Mildred Poultrey	68	46
St Alhallowes the Wall	500	356	St George Botolphlane	41	27	St Mary Colechurch	17	6	St Nicholas Acons	46	28
St Alphage	271	115	St Gregories by Pauls	376	232	St Mary Hill	94	64	St Nicholas Coleabby	125	91
St Andrew Hubbard	71	25	St Hellens	108	75	St Mary Mounthaw	56	37	St Nicholas Olaues	90	62
St Andrew Vndershaft	274	189	St James Dukes place	262	190	St Mary Summerſet	342	262	St Olaues Hartſtreet	237	160
St Andrew Wardrobe	476	308	St James Garlickhithe	189	118	St Mary Stayning	47	27	St Olaues Iewry	54	32
St Anne Alderſgate	282	197	St John Baptiſt	138	83	St Mary Woolchurch	65	33	St Olaues Silverſtreete	250	132
St Anne Blacke-Friers	652	467	St John Euangeliſt	9		St Mary Woolnoth	75	38	St Pancras Soperlane	30	15
St Antholins Pariſh	58	33	St John Zacharie	85	54	St Martins Iremonger.	21	11	St Peters Cheape	61	35
St Auſtins Pariſh	43	20	St Katherine Coleman	299	213	St Martins Ludgate	196	128	St Peters Cornehill	136	76
St Barthol. Exchange	73	51	St Katherine Creechu.	335	231	St Martins Orgars	110	71	St Peters Pauls Wharfe	114	86
St Bennet Fynch	47	22	St Lawrence Iewry	94	48	St Martins Outwich	60	34	St Peters Poore	79	47
St Benn. Grace-church	57	41	St Lawrence Pountney	214	140	St Martins Vintrey	417	349	St Stevens Colmanſtr.	560	391
St Bennet Pauls Wharf	355	172	St Leonard Eaſtcheap	42	27	St Matthew Fridayſt.	24	6	St Stevens Walbrooke	34	17
St Bennet Sherehog	11	1	St Leonard Foſterlane	335	255	St Maudlins Milkſtreet	44	22	St Swithins	93	56
St Botolph Billingſgate	83	50	St Magnus Pariſh	103	60	St Maudlins Oldfiſhſtr.	176	121	St Thomas Apoſtle	163	110
Chriſts Church	653	467	St Margaret Lothbury	100	66	St Michael Baſſiſhaw	253	164	Trinitie Pariſh	115	79
St Chriſtophers	60	47									

Buried in the 97 Pariſhes within the walls, — 15207 *Whereof, of the Plague* — 9887

St Andrew Holborne	3958	3103	Bridewell Precinct	230	179	St Dunſtans Weſt	958	665	St Saviours Southwark	4235	3446
St Bartholmew Great	493	344	St Botolph Alderſga.	997	755	St George Southwark	1613	1260	St Sepulchres Pariſh	4509	3746
St Bartholmew Leſſe	193	139	St Botolph Algate	4926	4051	St Giles Cripplegate	8069	4838	St Thomas Southwark	475	371
St Bridget	2111	1427	St Botolph Biſhopſg.	3464	2500	St Olaves Southwark	4793	2785	Trinity Minories	168	123
									At the Peſthouſe	159	156

Buried in the 16 Pariſhes without the Walls — 41351 *Whereof, of the Plague* — 28888

St Giles in the Fields	4457	3216	St Katherines Tower	956	601	St Magdalen Bermon.	1943	1363	St Mary Whitechappel	4766	3855
Hackney Pariſh	232	132	Lambeth Pariſh	798	537	St Mary Newington	1272	1004	Redriffe Pariſh	304	210
St James Clarkenwel	1863	1377	St Leonard Shoreditch	2669	1949	St Mary Iſlington	696	593	Stepney Pariſh	8598	6583

Buried in the 12 out-Pariſhes, in Middleſex and Surrey — 28554 *Whereof, of the Plague* — 21420

St Clement Danes	1969	1319	St Mary Savoy	303	198
St Paul Covent Garden	408	261	St Margaret Weſtminſt.	4710	3742
St Martins in the Fields	4804	2883	Whereof at the Peſthouſe		156

Buried in the 5 Pariſhes in the City and Liberties of Weſtminſter — 12194
Whereof, of the Plague — 8403

The Total of all the Chriſtnings — 9967
The Total of all the Burials this year — 97306
Whereof, of the Plague — 68596

The Diſeaſes and Caſualties this year.

Abortive and Stilborne	617	Executed	21	Palſie	30
Aged	1545	Flox and Small Pox	655	Plague	68596
Ague and Feaver	5257	Found dead in ſtreets, fields, &c.	20	Plannet	6
Appoplex and Suddenly	116	French Pox	86	Pluriſie	15
Bedrid	10	Frighted	23	Poyſoned	1
Blaſted	5	Gout and Sciatica	27	Quinſie	35
Bleeding	16	Grief	46	Rickets	557
Bloody Flux, Scowring & Flux	185	Griping in the Guts	1288	Riſing of the Lights	397
Burnt and Scalded	8	Hangd & made away themſelves	7	Rupture	34
Calenture	3	Headmouldſhot & Mouldfallen	14	Scurvy	105
Cancer, Gangrene and Fiſtula	56	Jaundies	110	Shingles and Swine pox	2
Canker, and Thruſh	111	Impoſtume	227	Sores, Ulcers, broken and bruiſed Limbs	82
Childbed	625	Kild by ſeverall accidents	46	Spleen	14
Chriſomes and Infants	1258	Kings Evill	86	Spotted Feaver and Purples	1929
Cold and Cough	68	Leproſie	2	Stopping of the ſtomack	332
Collick and Winde	134	Lethargy	14	Stone and Strangury	98
Conſumption and Tiſſick	4808	Livergrown	20	Surfet	1251
Convulſion and Mother	2036	Meagrom and Headach	12	Teeth and Worms	2614
Diſtracted	5	Meaſles	7	Vomiting	51
Dropſie and Timpany	1478	Murthered and Shot	9	Wenn	1
Drowned	50	Overlaid & Starved	45		

Chriſtned	Males	5114	Buried	Males	48569	Of the Plague — 68596
	Females	4853		Females	48737	
	In all	9967		In all	97306	

Increaſed in the Burials in the 130 Pariſhes and at the Peſt-houſe this year — 79009
Increaſed of the Plague in the 130 Pariſhes and at the Peſt-houſe this year — 68590

William Farr – the birth of medical statistics

In the UK, this type of documentation work continued for several hundred years until the middle of the Victorian era, when medical statistics first took off in earnest. William Farr, who is considered as one of the founders of medical statistics, pioneered the approach to how deaths are formally recorded in the UK, along with other medical statistics as well. Farr developed the first proper ontology for standardizing causes of death and organizing diseases in a way that was vaguely scientific and actually categorized rather than just listing in alphabetical order.

The origins of ICD

Throughout the 19th century, various efforts were made to start up organizations to make an international list of diseases that could be applied across several different countries. However, challenges arose due to war and strife in this period throughout Europe. Thus, it took around 40 years for Farr's list to finally become ratified by the International Statistics Institute thanks to the efforts of the chief statistician, Jacques Bertillon. Farr's list became the basis of the International Classification of Diseases (ICD), which is still around today and is in its 10th revision.

The rise of computing and data

The first half of the 20th century saw an unprecedented growth in technology, industry, and science. New international institutions arose, including the World Health Organization, which took over management of the ICD, and the United Nations. These institutions collected disease records on a much greater scale than ever seen before, creating an unprecedented amount of data. Alongside this, the dawn of computing arrived, together with early efforts in artificial intelligence, causing a resurgence of interest into how these disease records could be implemented using structured language that would allow computers and algorithms to perform basic reasoning on them.

Dendral – the first application of a knowledge base in biology

Dendral was a groundbreaking project in the field of artificial intelligence (AI) and is considered one of the earliest expert systems. Developed in the 1960s at Stanford University, Dendral was designed to analyze chemical mass spectrometry data to deduce the possible molecular structures of organic compounds. This pioneering system combined rules-based reasoning with a database of known chemical structures to make predictions, and effectively laid the groundwork for computational biology and the application of AI in scientific research. Moreover, Dendral is the first instance of using a knowledge-based system in biology and is perhaps the first instance of any expert system of any kind.

The rise of bioinformatics – data begins to explode

Following Dendral's demonstration of how computational methods can solve complex biological problems, researchers and scientists were inspired to further explore the potential of applying computer technology to biological data across the 70s and 80s. The advent of high-throughput sequencing technologies and the exponential growth of biological data necessitated advanced computational tools for data management, analysis, and interpretation, further propelling the growth of bioinformatics.

Since early biological data sources were generally developed independently and used bespoke, tailored formats and terminologies, it was very difficult to integrate and share data between them in a way that allowed that data to be manipulated algorithmically. On top of these technical considerations, understanding complex biological processes requires more than simple keyword-based queries. These factors together drove demand for the development of sophisticated algorithms and software to facilitate analyzing genetic sequences, predicting protein structures, understanding complex biological networks, and more.

1971
Protein Data Bank founded

1971
'Online' version of MEDLINE launched

1972
First sequencing of a complete gene

1974
SQL Developed

1976
First Relational Database

1976
First sequencing of a complete genome

1978
Intel 8086 Processor

1983
PCR invented

1987
First automated sequencing machine

1992
Nucleic Acid Database

1996
PubMed launced

The fundamentals of semantic technology

In today's world, semantic technology encompasses a suite of tools and methodologies designed to enhance the way computers understand and process the meaning of data, text, and web content - akin to human comprehension. At its core, semantic technology leverages data, ontologies, knowledge graphs, and natural language processing (NLP) to create a rich, interconnected framework that allows for more sophisticated data interpretation, retrieval, and analysis.

Semantic technology enables machines to understand the context and relationships within data, facilitating more accurate search results, data integration, and the automation of reasoning tasks. By imbuing data with meaning and making it machine-readable, semantic technology paves the way for advanced applications in artificial intelligence, information management, and beyond, transforming vast amounts of raw data into actionable knowledge.

Let's dive further into some of the more relevant semantic technologies and concepts for life sciences companies.

Ontologies

An ontology, within the context of semantic technology, is a formal representation of knowledge as a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain and to enable knowledge sharing and reuse among computers and humans. Ontologies are a crucial component of semantic technology because they provide a structured framework that allows explicit specification of the meaning of terms and the relationships between them.

There are various ontologies in the life sciences realm, such as Gene Ontology, Sequence Ontology, and Medical Subject Headings (MeSH). These ontologies promote consistency in the preferred terms within the field. In doing so, they enable indexing of content and content retrieval through browsing or searching.

An ontology is a formal representation of knowledge within a particular domain.
It generally consists of three things:

1. **Concepts** are the entities or "things" relevant to the domain.

Examples: a disease, a gene, an active substance, a person, an institution.

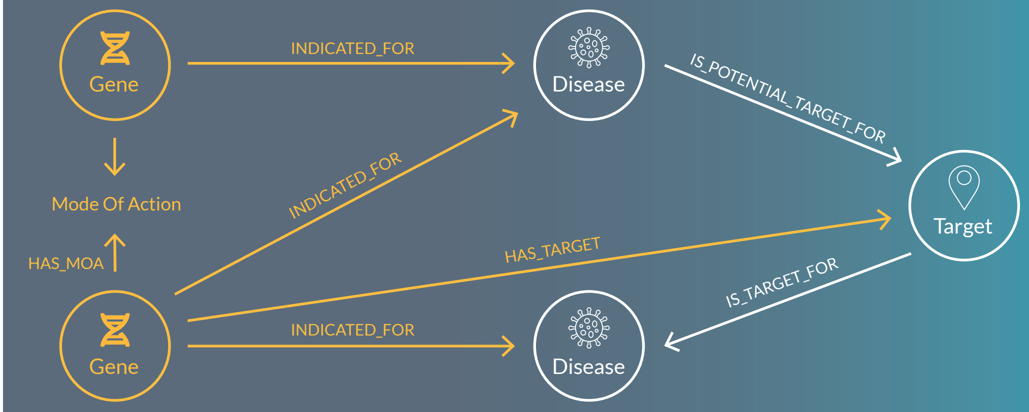
2. **Attributes** are characteristics or properties that these entities possess.

Example: for the concept of a person, an attribute might be a name.

3. **Relations** are the ways in which the entities interact or relate to each other.

Example: the concept of a drug and the concept of a disease may be related to each other as the drug is a treatment for the disease.

Drug Repurposing Example



Knowledge graphs

In a knowledge graph, data is consolidated from various databases and sources into a unified system where this data is primarily represented through nodes (a point in a network system at which pathways intersect or branch).

For life sciences companies, these nodes could be the key entities within their data, encompassing a wide range of elements like genes, people, organizations, and diseases. Nodes are interconnected via edges allowing for relationships to be established between them, such as linking a clinical study to its sponsoring organization or linking drug-target interactions.

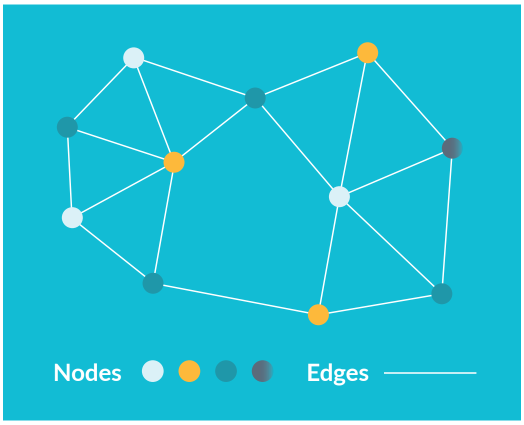
To provide further depth and context to a knowledge graph, nodes and edges can be enriched. Nodes can be given attributes, such as descriptive details or quantitative properties. For instance, a node representing a person might have attributes like name, age, and occupation. Edges can be annotated or labeled within the graph to represent types of relations. For example, in a graph where an edge connects two nodes representing people and organizations, the edge annotation could specify the nature of this connection, such as “employee of” or “founder.”

WHEN TO USE KNOWLEDGE GRAPHS

Know your use case
Knowledge graphs work well for many use cases, but often so does a structured query language (SQL) database, which is sometimes a simpler solution for the task at hand.

Don’t add everything
When a knowledge graph is the right solution, not all instance data needs to be integrated into it. It often makes more sense to store this data elsewhere and reference it.

There can be multiple knowledge graphs
It’s often better for performance, value, and usability to have multiple knowledge graphs, each serving a distinct purpose. As long as they share a common language they can be integrated as necessary in the future.



What is the difference between an ontology and a knowledge graph?

The key difference between a knowledge graph and an ontology is the use or inclusion of different types of data within the structure. An ontology represents knowledge across domains at a high level, essentially providing the key concepts of a domain and how they relate to each other. In the health domain, for example, data encoded in an ontology like ICD or SnoMed, consists of diseases and conditions, and the relationships between these entities. In this way, an ontology is a rather rigid structure.

On the other hand, knowledge graphs tend to incorporate more instance data. For example, within the healthcare domain this is data like patient records, l lab results, or data from specific clinical trials. In addition to this, a knowledge graph can also integrate full text and other types of data. A knowledge graph, therefore, is utilized for more complex querying and analytical type tasks, while an ontology is more useful in looking up straightforward information.

While both ontologies and knowledge graphs aim to represent domain knowledge, there are key differences:

Scope: a knowledge graph includes instance data (an ontology can be included as nodes or in attributes of a knowledge graph).

Purpose: knowledge graphs are geared more towards querying and analysis.

Flexibility: knowledge graphs can integrate diverse kinds of data.

How a knowledge graph uses an ontology

A knowledge graph utilizes an ontology as its foundational framework to define the types, properties, and interrelationships of the data elements within its domain. The structure provided by ontologies enables knowledge graphs to effectively organize, integrate, and interpret complex datasets. By leveraging ontologies, knowledge graphs can infer new knowledge through reasoning, enhance data interoperability, and ensure consistency across diverse data sources. For example, in a life sciences knowledge graph, an ontology might define the relationships between diseases, symptoms, treatments, and patient demographics, allowing for sophisticated queries that can support research or clinical development. The use of ontologies thus empowers knowledge graphs to provide a rich, semantic understanding of the data, facilitating advanced analytics and insights.

Natural language processing

Natural language processing (NLP) is at the intersection of computer science, artificial intelligence, and linguistics. NLP is focused on enabling computers to understand, interpret, and generate human language in a way that is both meaningful and useful. NLP involves a range of techniques and technologies designed to bridge the gap between human communication and computer understanding, facilitating the processing of large volumes of natural language data. This includes tasks such as speech recognition, language translation, sentiment analysis, and chatbot development. By analyzing the structure and meaning of words and sentences, NLP systems can extract insights, identify patterns, and respond to queries in natural language.

When employed on unstructured data, NLP can pull out meaningful information and insights from text that lack a predefined format. By applying techniques like sentiment analysis, entity recognition, and topic modeling, NLP transforms unstructured data into structured, actionable knowledge, enabling automated processing and analysis for various applications.

Alongside the growth of ontologies and knowledge graphs, there's been a corresponding growth in NLP. Typically, a knowledge graph creates demand for NLP services as there are several instances in the lifecycle of a knowledge graph where NLP's capabilities provide major value for a knowledge graph and its users.

NLP techniques enable the extraction of structured information from unstructured text, which can then be integrated into a knowledge graph. This synergy allows for the dynamic expansion of the knowledge graph with information extracted from a wide range of text sources, including scientific literature, web content, and internal documents.

NLP applications

- Text mining and information retrieval
- Categorization and annotation
- Deep learning and semantic role labelling
- Ontology building and annotation

Additionally, the combination of NLP with knowledge graphs facilitates more intuitive and natural language-based interactions with the data, enabling users to query and analyze the graph using conversational language. This integration not only significantly improves the accessibility and usability of the information stored in knowledge graphs but also enhances the capability for semantic search, information retrieval, and automated reasoning, providing deeper insights and supporting more informed decision-making processes.

Large language models

Large language models (LLMs) have become a major topic throughout recent years. LLMs are not strictly semantic technologies in the traditional sense but rather advanced applications of machine learning and NLP that have significant implications for semantic analysis and understanding. While semantic technologies typically involve the use of ontologies, knowledge graphs, and reasoning to understand and infer the meaning of text based on structured relationships, LLMs approach semantics through the statistical patterns and relationships learned from vast amounts of text data. LLMs can contribute to semantic technologies by enhancing their capabilities, especially knowledge graphs.

What is the difference between an LLM and a knowledge graph?

The primary difference between an LLM and a knowledge graph lies in their approach to handling and understanding information. An LLM is a type of artificial intelligence model that learns from vast amounts of text data to understand and generate human-like text. It operates based on statistical patterns and relationships within the data it has been trained on, enabling it to perform a wide range of language-related tasks without explicit programming for each task.

On the other hand, a knowledge graph organizes data into entities and their interrelations, based on a structured framework or ontology. This structure allows for the explicit representation of knowledge in a way that is understandable both by machines and humans.

While LLMs excel in generating and understanding language through learned patterns, knowledge graphs provide a clear and interconnected map of facts and relationships, offering precise and explainable insights into the data they represent.

90%
50%
Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam

What can be accomplished when leveraging an LLM with knowledge graph within a pharmaceutical organization?

Some application areas include:

1. Question answering

LLMs can be trained to understand natural language queries and retrieve relevant information from knowledge graphs.

By combining the capabilities of LLMs with the rich relationships and semantic connections captured in knowledge graphs, pharmaceutical researchers can obtain precise answers to specific questions, enabling faster information retrieval and decision-making.

2. Data exploration

Knowledge graphs enable exploratory data analysis by capturing relationships and connections between different data elements.

LLMs can leverage these connections to provide insights and recommendations based on patterns and correlations found within the graph. This can help identify potential target-drug associations, drug-drug interactions, or identify new research areas based on existing knowledge.

3. Knowledge curation and expansion

LLMs can assist in the curation and expansion of knowledge graphs. By analyzing large amounts of textual data, LLMs can extract relevant information and populate or update the knowledge graph with new entities, relationships, and attributes.

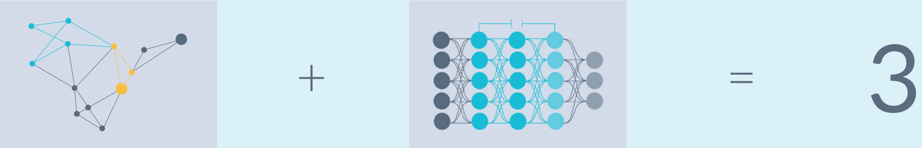
This enables the continuous growth and enrichment of the knowledge graph, enhancing its utility in drug discovery and development efforts.

One could think that the rise of LLMs indicates that knowledge graphs will sooner or later be made obsolete and be, in a sense, replaced by LLMs. When taking a full inventory of what's possible with each technology, it's clear that this is unlikely to be the case, as both systems have advantages and disadvantages that make them better suited for certain applications.

For example, while LLMs are powerful at understanding and prediction, they are not designed to store knowledge or to allow this knowledge to be corrected or governed, making them rather slow and inaccurate when it comes to retrieving specific knowledge.

Knowledge graphs on the other hand, are reliable for producing (and reproducing) specific, factual results while providing full transparency on how a query returned that result. While neither could replace the other, combining the two technologies opens up the possibility of getting the best of both worlds.

Large language models complement knowledge graphs



Knowledge graphs

- Predefined terminologies
- Strong for long tail
- Reliablility: no bias of hallucinations
- Security
- Reproducibility and transsparency
- Low run-time cost
- Easy to fix mistakes
- Easy data visualization

Large language models

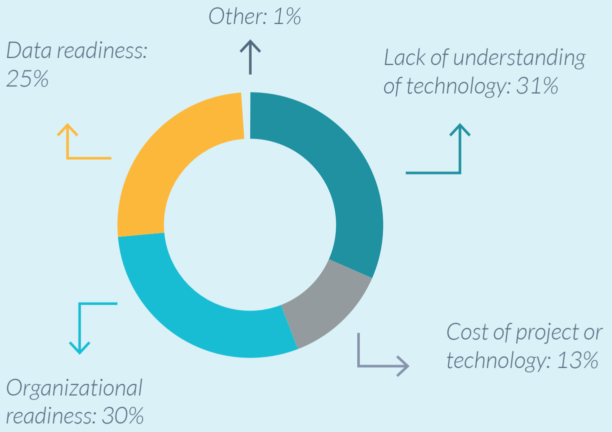
- Unmanaged terminologies
- Strong for common things (e.g. diseases)
- Contextual understanding
- Conversational interface
- Human readable answers
- Strong with ambiguity

Combine both to strengthen semantic understanding and accurancy

Principles and guidelines for adopting and rolling out semantic technologies

Leveraging semantic technology enables an organization to harness the full potential of its vast and diverse data sets, from genomic sequences to clinical trial data, and everything in between. However, not all organizations may be ready to roll out the use of semantic technology. Below are several general guidelines to follow to help make the roll-out and adoption of semantic technology as seamless as possible.

ONTOFORCE webinar audiences noted that a lack of understanding of the technology and organizational readiness are some of the main obstacles for getting started with semantic technologies.



Results of our poll *"What are the main obstacles for you getting started with semantic technologies?"*

September 2023

General principles for getting started with semantic technology

Define objectives: understand what you want to achieve with semantic technology

Defining clear objectives before implementing semantic technology is paramount for any organization aiming to harness its full potential. This preparatory step ensures that the deployment of semantic technologies is aligned with the organization's strategic goals, whether they involve enhancing data interoperability, improving information retrieval, or facilitating advanced analytics. Clear objectives guide the selection of appropriate semantic tools and methodologies, which is essential as there are many exciting possible applications to pursue. If an organization is too eager, without proper planning, it might end up biting off more that it can chew – jeopardizing a successful roll-out and adoption plan.

Clear objectives also enable an organization to tailor the development of the semantic layer to meet specific needs and provide a benchmark against which the success of the implementation can be measured. Moreover, having well-defined goals helps in effectively communicating the purpose and expected benefits of the semantic technology initiative to stakeholders, securing their support and fostering a collaborative environment.

Start small and focused: identify a smaller project or dataset where semantic technology can provide immediate value

Starting small may not seem exiting but it can be key for organizations who are just beginning their journey. Consider a particular research project, a subset of data, or a specific business process that could benefit from the implementation of semantic technology. Starting off with this smaller area and showcasing the value that semantic technology brings serves as a proof of concept that can guide future endeavors.

Perform a gap analysis: determine exisiting tools and skills and what's needed

Performing a gap analysis helps identify the disparity between the current state and the desired future state of technological capabilities. This process involves a thorough assessment of the existing tools, technologies, and skills within the organization, as well as those required to successfully deploy and utilize semantic technology.

By understanding what resources are already available and what additional investments are needed, organizations can make informed decisions about training, hiring, and technology acquisition. This strategic approach ensures that the implementation of semantic technology is both efficient and effective, minimizing redundancies and focusing resources on areas that offer the greatest potential for enhancing data interoperability, analysis, and insight generation.

Involve domain experts: identify in-house experts or find contracted ones that can ideally embedded in teams

An often-over-looked step is involving domain experts. Ideally, these experts can be tightly embedded in an organization's teams. For example, a health data company may embed clinicians in product teams on a day-to-day basis who can then work with and guide the development of semantic technology. In this way, the clinicians can provide helpful and specific domain knowledge to drive development. This level of insight is very hard to replicate without a close level of integration between the domain and technology experts.

Plan for scaling: be wary of prematurely optimizing your solution at the expense of getting things done

It is important to think about scaling your solution from the start. However, it's important to not dive into scaling too fast. If the history of semantic technology teaches us anything, it's that there will always be need for scaling. Thus, thinking and planning about scaling shouldn't come at the expense of actually getting things done with a solution. It is still essential to keep an eye towards the future and ensuring that a solution is interoperable is one way to do that. With an interoperable solution, data can easily be extracted or accessed by different systems in the future.

Monitor and update: semantic technologies require ongoing attention

Monitoring and maintaining semantic technology is crucial for ensuring its continued effectiveness and relevance in handling an organization's evolving data needs. Regular updates and optimizations help in adapting to new data sources, changing standards, and emerging technologies, thereby maintaining the integrity, accuracy, and usability of the semantic framework for advanced data analysis and decision-making processes. However, the monitoring and maintenance phases are not as exciting as the building phases. While it might be relatively easy to get buy-in for building a shiny new project, it can be harder to get others to think about monitoring and maintaining a solution. Consider how to engage the right individuals and teams for these necessary and ongoing processes.

Data requirements for getting started with semantic technology

Structured and consistent

Ensure there is a well-defined model and structure. The more that data is structured, the easier it is to use and extract value from. safety profiles.

Identified

Each instance should have a unique identifier following a sensible schema. Enforcing data standards pays enormous dividends down the line.

Annotated

Ideally data should be annotated with semantic metadata or linked to ontologies.

QA and validation

Have a plan for cleaning, enriching or normalizing data over time.

Governance/security

Consider access control, data governance, compliance, GxP, etc.

Provenance

Ensure provenance is captured from day one. Data loses its value without proper provenance tracking.

Advice for smaller companies:

Be quick, agile, and leverage what's existing

- **Leverage open source and community tools:** there are several open-source tools and platforms available for creating and managing ontologies and knowledge graphs. For a smaller company, there are great advantages in being able to quickly adopt something that's existing in the larger ecosystem. Take advantage of your agility.

- **Use existing ontologies and standards:** consider adapting or extending an existing ontology rather than starting from scratch.

- **Prototype rapidly:** create a minimum viable product (MVP) or prototype to validate the utility of the semantic technology for your specific needs to ensure you're on the right track.

- **Collect feedback and iterate:** gather user feedback to understand if the system is meeting its objectives and what could be improved.

- **Make decisions quickly:** work directly with someone who can make key decisions without delay to help the project stay within the allotted timeframe and budget.

- **Look for external funding:** many opportunities exist, from grants to customer pilots to partnerships. Leave no stone unturned when searching for funding.

Advice for larger companies:

Treat it as a change management process, but slice the problem

Larger companies might find success in approaching their project from a change management perspective from the start. . This is best done by slicing the project into chunks. Ensure that the first part of the project can not only be tackled in a six-to-nine-month timeframe, but that it will also provide decent, tangible results. These results can then be leveraged for additional funding for the next parts of the project. Other aspects to keep in mind:

- **Build a cross-functional team:** ensure the project team has a combination of technical acumen and domain expertise, and don't forget the business side.

- **Find champions:** executive buy-in will be critical for bringing teams on board, enabling budgets, implementing standardization, and supporting with stakeholder management.

- **Needs assessment:** having a well-defined problem statement that fits with the strategic objectives of the organization and is ideally linked to tentpole strategic goals in the organization.

- **Pilot programs:** find smaller-scale projects linked to strategic goals that can deliver value in a 6-9 month timeframe.

- **Avoid decision paralysis:** implementing a regular steerco meeting to identify issues and blockers can help the decision-making process.

About

ONTOFORCE

For more than a decade, ONTOFORCE has addressed the problem many life sciences companies struggle with when bringing together structured and unstructured data to create new insights. These insights lead to accelerated drug discovery, more in-depth insights into research, real-world evidence, optimized clinical trial research, and a faster go-to-market.

Our vision

In a future where the rate of breakthroughs will be determined by the accessibility of knowledge, ONTOFORCE is leading the charge towards a life sciences industry without barriers to data, where empowered stakeholders across the business can collaborate to bring drugs to market faster.

Our mission

Through our intuitive data and technology solutions, we transform complex data into actionable insights, streamline drug development, and accelerate new treatments to patients.

Transform data into knowledge

www.ontoforce.com

hello@ontoforce.com