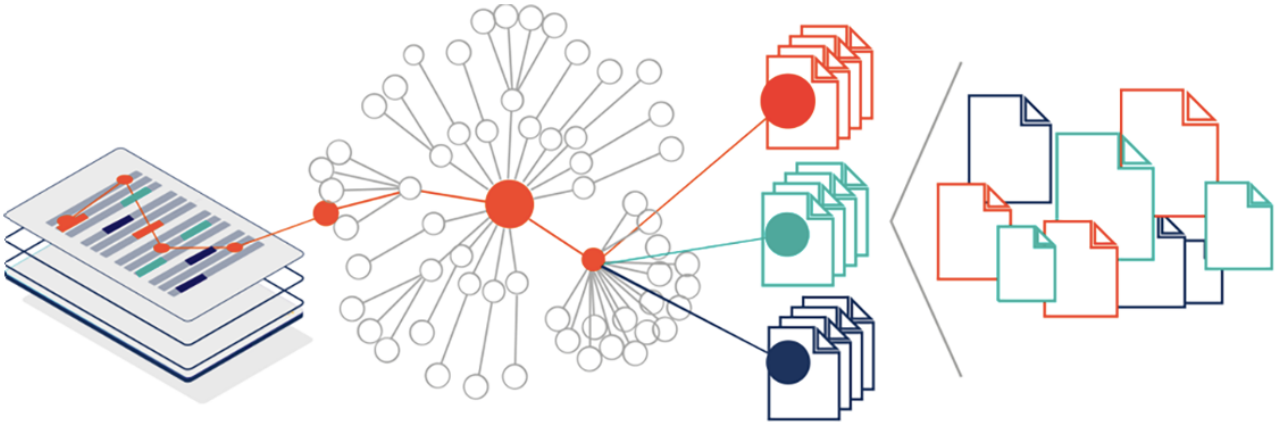
Author: Ivaylo Kabakov

Although text analysis can sound too complex at times, it's easy to see that understanding unstructured content brings a lot of business benefits. At Ontotext, we have over 20 years of experience in natural language processing and we have proven that knowledge graphs are very beneficial for solving text analysis challenges, thus helping with content management in general.

On one hand, the data stored in a knowledge graph can be used as input for text analysis and can improve this task significantly. On the other hand, the interconnectedness of the concepts in a graph can serve as additional context to help infer new knowledge and make it easier for machines to interpret natural language.

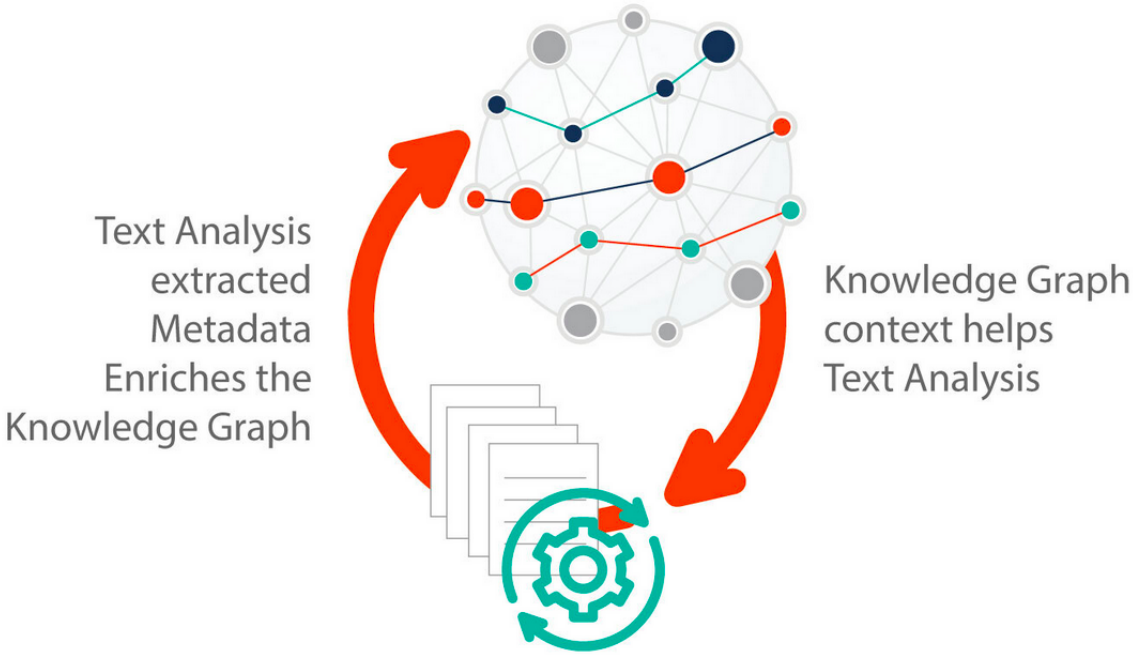
Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com



On top of that, text analysis can use all this information to extract new concepts and relationships from unstructured content and feed that information back to the knowledge graph, turning this natural symbiosis into a self-enforcing loop. This can be utilized in a broad range of tasks we often need to solve when dealing with content management.

Read on to learn more about what they are and how we solve them!

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



The Business Value: The Top Benefits Of Our Approach

So, what's the business value we claim to deliver by offering text analysis with knowledge graphs?

Over the years, we've built a plethora of content management solutions, which have proven to bring a number of benefits:

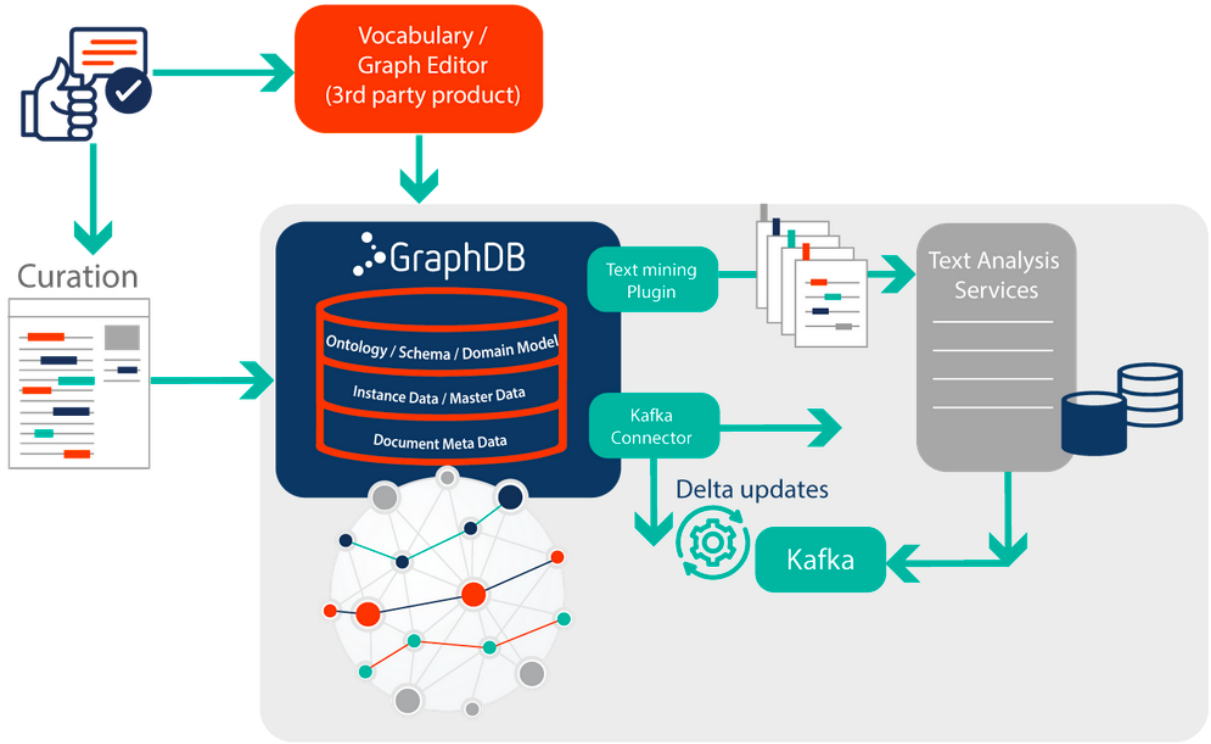
- By linking your knowledge to concepts from a knowledge graph, we enable you to **discover knowledge locked away** in heterogeneous content (which may be stored in multiple document or content stores across your organization, creating silos and making it practically impossible to get the full picture).
- By repackaging your content for different distribution channels (to fit different personas, business objectives, etc.), we allow you to **reuse and repurpose this content** instead of creating more and more new pieces.
- By interlinking and organizing your content through classification, we make it **easy to consume**.
- By using recommendations, we make your content explorable in a more linear way and

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

increase your user engagement. (With many of our customers, we have proven that this is quite an observed result!)
- Last but not least, by exploiting the interconnectedness between concepts in a big knowledge graph, we empower you to **get insights** that help you gain much more from your data.

The Secret: A Knowledge Graph and Text Analysis That Complement Each Other

All of our text analysis solutions stand on the shoulders of other Ontotext products.



First of all is our flagship product GraphDB – a highly scalable and robust RDF database for knowledge graphs. You probably know that it now has a text mining plugin, which enables the integration of third-party text analysis services. It also features a Kafka connector that

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

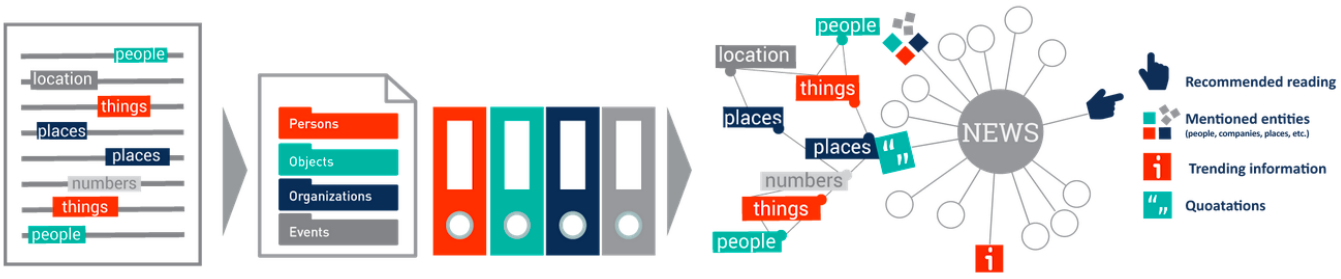
allows easy processing of RDF updates coming from any external systems.

Another (new) member of our product portfolio that contributes to our text analysis offerings is the Ontotext Metadata Studio. It enables all the human-in-the loop-activities you would need when working with text analysis.

We also often complement our products with some of our partners' offerings to provide an end-to-end text analysis solution. This includes Semantic Web Company's PoolParty and metaphacts' metaphactory, to mention a few.

Text Analysis Basics: Tasks and Approaches

When building a text analysis solution, there are various content-centered tasks we usually have to tackle. Some of the ones we encounter most often include document classification, named entity recognition, relationships extraction, recommendation services and semantic search.



Although there are generally two approaches to solving these tasks – rule-based or machine learning based, real-life use cases frequently require a combination of both. So, throughout the years, we have built a technological arsenal that enables us to integrate both rule-based expression logic and machine learning components. This enables us to deliver the best possible result for the task at hand.

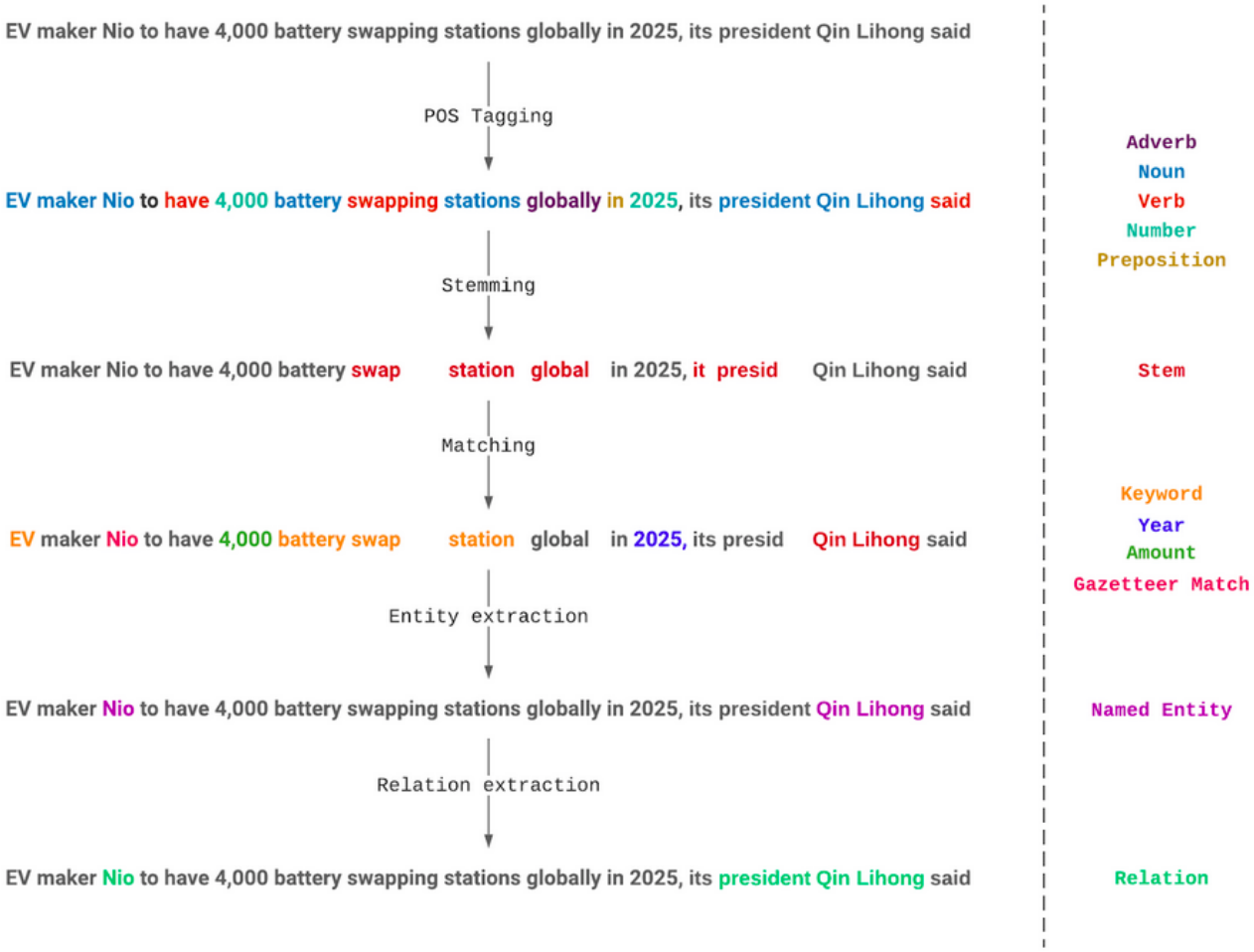
The Nuts and Bolts: Building A Text Analysis Pipeline

Usually, what we have to do when solving a text analysis task is to build a pipeline – a set of successive steps, where each subsequent step depends on the outcome of the previous one.

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

First, we may need to do some pre-processing such as sentence splitting, part-of-speech tagging, morphological analysis, etc. Then, we may need to match keywords or named entities against dedicated gazetteers already ingested in the knowledge graph. Or we may need to do named entity linking to find out, for example, who exactly a person is from a certain knowledge base. Finally, we may need to do relation extraction to determine the relations between a person and an organization or between organizations like in cases of C-level role changes, merger and acquisition events, asset deals, etc.

For example, the following is a sentence from a news article about one of Tesla's competitors, the Chinese electric vehicle (EV) maker NIO Inc. and this is what happens when we parse it through our pipeline.



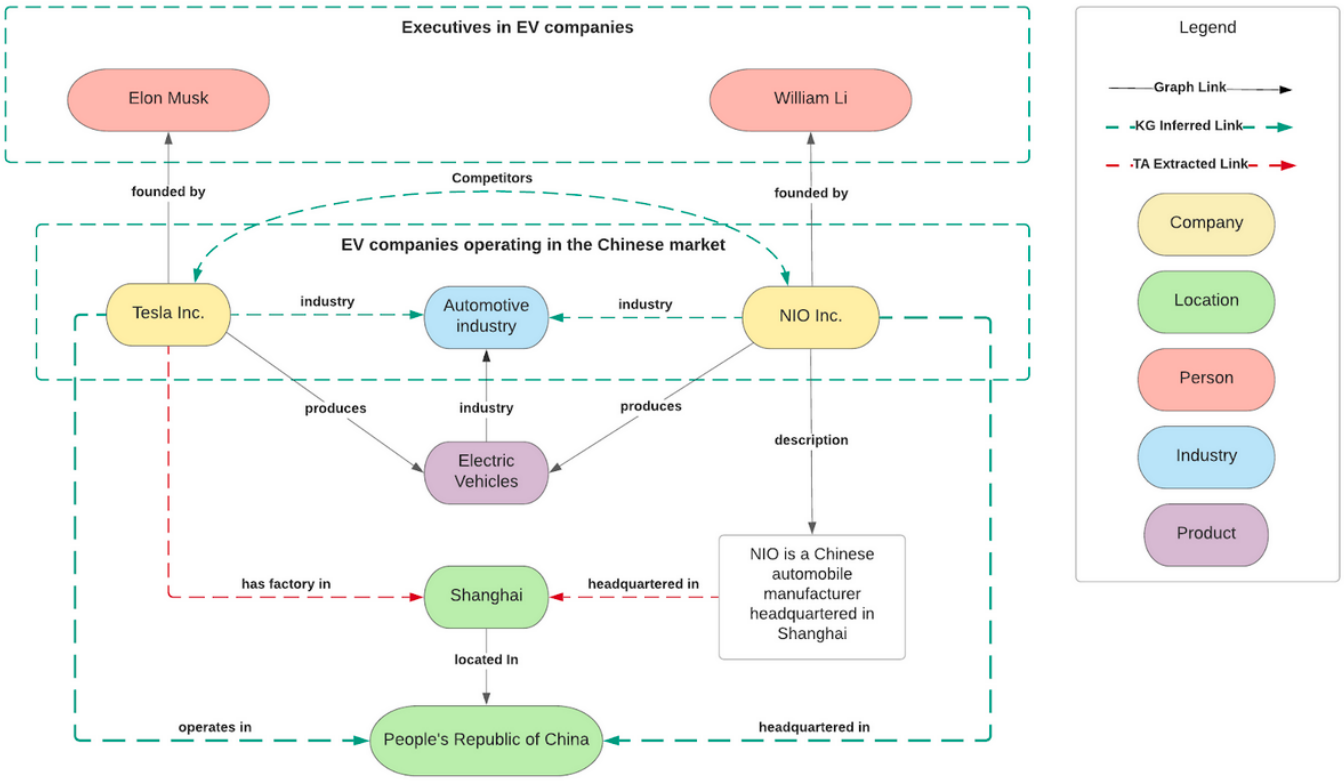
Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

As a result of the parsing, you can see which parts of this sentence are verbs, nouns, etc. Or how the different expressions of some words are reduced to their root form for easier processing. You can also discover certain keywords, years, amounts, etc. or see the outcome of name entity and relation extraction.

Behind the Scenes: Simple Graph Inference

Now, let's have a look at what happens at the level of the knowledge graph.

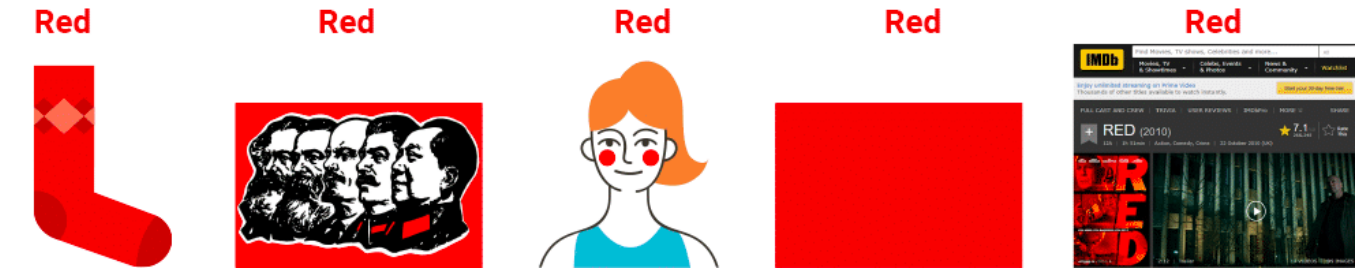


As you can see from this diagram, if a dataset contains both Tesla and NIO, we can process the descriptions of the companies through a text analysis pipeline and obtain additional facts to enrich the knowledge graph. Based on the explicit facts in the graph, we can also infer that NIO is located in Shanghai. We know what Shanghai is because it links to the GeoNames ID of that city and we can also infer that it's located in the People's Republic of China.

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

All this allows us to answer questions like: “Who are all the companies from the knowledge graph working on EVs and operating in the Chinese market?”. We would be able to answer such a question (and “free of charge”!) only because we have extracted from the articles that Tesla is building a factory in Shanghai and that NIO is headquartered there. Or we can answer questions like “Who are the executives of all EV companies?” or “Who are the executives of all EV companies operating in Asia?”. All of this might not be possible if that same data was modeled in a relational database.

Finally, there’s also the challenge of disambiguating general types of entities (such as people, organizations and locations), which often trip machines over. For example, most people interested in baseball will easily understand that the news title “Red Sox Tame Bulls” refers to a baseball match. However, lacking their background knowledge, machines will generate several linguistically valid interpretations, which are very far from its intended meaning. And, by the way, people not interested in baseball will not fare much better.



To Sum It Up

Text analysis is a big topic and to have useful results, you need to have the know-how, the technology, the processes, the ability to operationalize it and maintain it, etc.

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

ontotext

Over the years, we have built custom solutions for many of our clients but we have pre-packaged offerings as well. The latter can greatly accelerate the delivery of such solutions as well as lower the level of their complexity. We believe that this can be achieved best by providing horizontal technological offerings in three separate tiers:

- **Essentials** – a more naïve but very explainable text analysis offering, providing full control over the outcome;
- **Expert** – an offering building on Essentials and extending it with metadata preservation and search capabilities;
- **Enterprise** – an offering building on Expert with added machine learning, re-training capabilities, content-centric services like enterprise search recommendations, etc. and more.

Stay tuned for further details!

Book a Discovery Call

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