These days, there’s hardly a company or an individual for whom textual information isn’t playing a key role in what they do and how they do it. We are all surrounded by gigabytes of textual data to put to use: from the mundane archive folder of our emails, we often lose ourselves in, to the large bodies of legal documents a lawyer needs to go through to ensure a company’s actions comply with the relevant laws, policies and regulations.

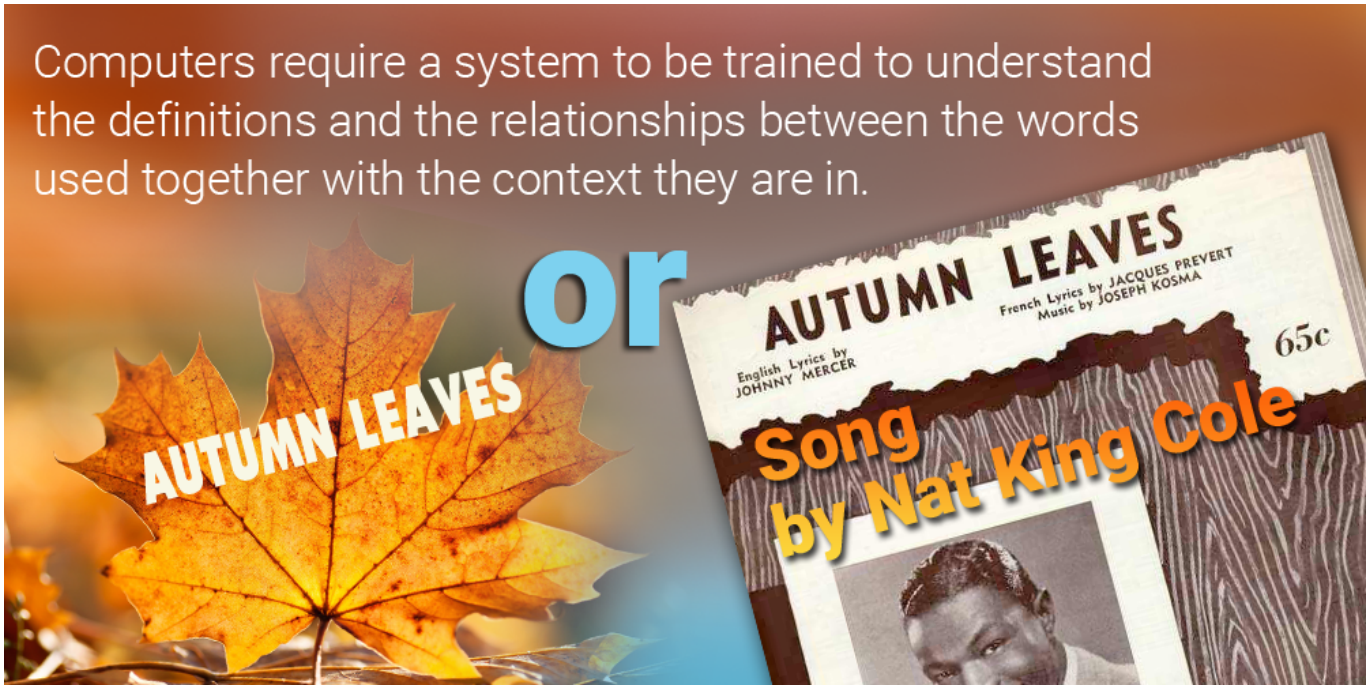
The truth is, we can’t possibly manipulate these gigantic quantities of text streams in a lifetime. The best we could try is to teach our machines to do that for us, efficiently and accurately.

Autumn Leaves and Text Analytics

Like autumn leaves carried away, digital words and numbers dash through our private and public lives. They swirl across so many different channels in so many different contexts that our traditional methods and tools for coping with the written word are becoming increasingly inadequate. In a digital swirl of texts, new methods find their places in our approaches towards taming the written word. One such method is Text Analytics.

Text analytics processes, or simply put, the techniques used to make the meaning of textual resources machine-readable, processable and further ready for analysis and action, are different and vary according to the needs and the scales of the solutions required. Their challenges, however, are similar: work out a means for our language – an elaborate construct, ambiguous and context-dependent in its nature – to be formalized and presented in an unambiguous way to a computer system.

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



Let's get back to "autumn leaves" and see what all of the above means in practical terms. For those of us who love jazz, "autumn leaves" can directly bring to mind a song by Nat King Cole. But what about a computer? How is a system supposed to figure out whether this is the name of a song, or a sentence about the end of the autumn season, or a tag about a bunch of autumn leaves?

Computationally, the tasks of disambiguating words, detecting a reference to a song and understanding the sentence's syntax are extremely complex. They require a system to be trained to understand the definitions and the relationships between the words used together with the context they are in. This is where Text Analytics comes into play.

The main goal of Text Analytics is to bring textual content to a point where it is represented as data and thus accessible for machines to process and make sense of. Share on X

This might involve syntax, token and sentence parsing, keywords and named entities detection and classification, as well as text summarization and trustworthiness and veracity analytics. It is through this "weaving" of data into text, which is the result of Text Analytics, that:

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

- texts become readable for machines;
- information retrieval is improved;
- connectivity on multiple levels is achieved;
- knowledge is turned into highly-manageable chunks.

Coupling Text Analytics with a Knowledge Graph

Understanding a text, or let’s be more accurate and say processing a text as to extract certain meaning out of it, presents many challenges to a machine: from identifying where the words start and end, through detecting phrases and sentences, all the way to determining what the entire text is about, based on the people, things, events and places, mentioned in it.

The first two are what Text Analytics deals with. The latter is the hardest for it often requires referring to knowledge that is outside the given text. Such knowledge, available for machines to access, is what a Knowledge Graph is all about.

Here’s a Graph, Go Figure!

If Text Analytics allows for a machine to have basic reading skills, Text Analytics coupled with a Knowledge Graph makes for a machine capable of recognizing people, organizations, locations and general encyclopaedic concepts in a text.

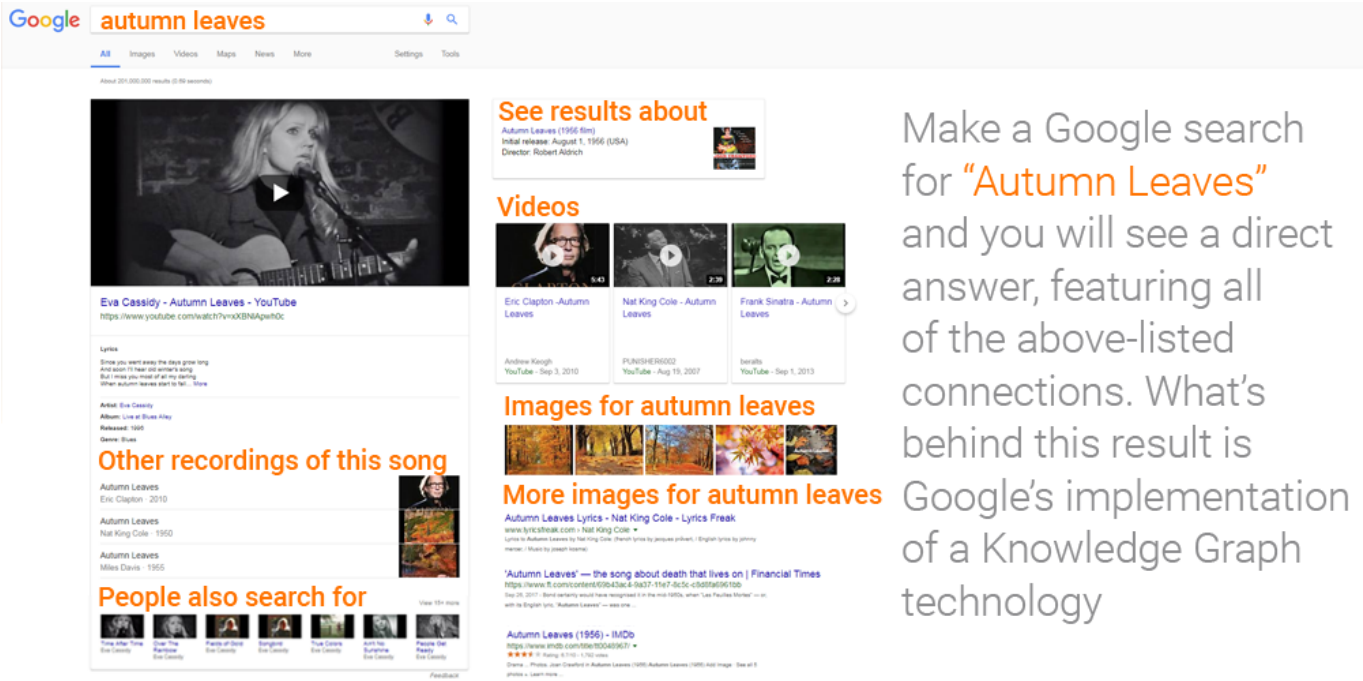
The Knowledge Graph represents a collection of interlinked descriptions of entities – real world objects, events, situations or abstract concepts. When a critical mass of concept descriptions are linked together in a big Knowledge Graph, they allow computers to interpret them and derive context and awareness, similar to the one that people develop in specific domains by means of education – formal or informal.

Knowledge Graphs serve machines to connect the dots. For example, a Knowledge Graph could contain the explicit statement that “Autumn Leaves” is a song by the artist Nat King Cole, recorded by other artists such as Eva Cassidy, Edith Piaf, Erik Clapton, etc.

Case in point, make a Google search for “Autumn Leaves” and you will see a direct answer,

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

featuring all of the above-listed connections. What's behind this result is Google's implementation of a Knowledge Graph technology – a graph database that provides structured and detailed information about a given entity and the connections it enters with other entities.



As Atanas Kiryakov explained in a webinar touching on the workings behind a Knowledge Graph (see Graph Analytics on Company Data and News), a big Knowledge Graph can provide all flavors of context, which computers need to be able to recognize an entity in a text. Such richness is the result of a complex dynamics of relationships between machine-readable context and Text Analytics and involves technologies related to:

- **Classification** – grouping entities into classes, categories, etc. they belong to. Knowing that Paris Hilton is a Person helps us disambiguate a reference to her in a context like "Paris Hilton was arrested" so that it is clear that this is not about the Hilton hotel in Paris.
- **Related entities** – if an article mentions multiple cities in Texas, this is evidence that "Paris" refers to Paris, TX and not the capital of France.
- **Differentiating features and similar nodes** – where similarity acts as an implicit

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

relationship.
- **Importance and popularity** – if Einstein is mentioned in a text, without much context suggesting alternatives, the machine should assume it is a reference to the most popular entity with this name – Albert Einstein, the genius physicist.
- **Co-occurrence** – when entities are typically mentioned together.

The Knowledge Graph and the Enterprise

Within an enterprise context, a Knowledge Graph can be used to enhance Text Analytics in several ways. For example, a machine’s reading skills can be significantly improved when the system is fine-tuned to read Wikipedia, the news, various databases and sets of documents.

This can happen with the creation and, further, with the extension of an underlying Knowledge Graph with specialized datasets that will fine-tune the system’s reading skills with respect to specific types of text and the sort of facts that should be extracted from them. For instance, once built, a Knowledge Graph can further be augmented with proprietary databases, documents and other data about products, employees, clients and suppliers over time.

At Ontotext, the Text Analytics processes coupled with the creation and maintenance of a Knowledge Graph are proven to be of immense help in the fields of:

- Publishing and Media – a system can process terminologies in various domains such as engineering, technology, etc.
- Financial services – a system can process macro- and microeconomic concepts, financial instruments, money amounts, etc..
- Life science and Healthcare – a system can be fed with knowledge about genomics (genes, mutations, assay), drug products (vendors, brand names, ingredients), etc.

The Meaningful Way Out of the Textual Heaps

Despite being overwhelmed by texts flying around us at electric speed, we can still get better results faster and search our content more easily with minimum effort and maximum clarity and meaning. We only need to conceive of Text Analytics in the broader context of teaching

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

machines not only to read for us but to read for us better, that is, with more awareness of the context around the text itself.

Knowledge Graphs put information in context and allow its better interlinking, interpretation, analytics and reuse. Share on X

Tons of diverse documents online can be easily automatically retrieved, with the needed information extracted from them, combined and made sense of. And this is how, in concert with having information collected from different sources about a place or a person, Text Analytics makes the power of the billions of words and numbers swirling across our digital spaces work for us in a meaningful way.

For wouldn’t it be wonderful to have a machine read all the latest laws or sift through a thousand of documents to find an answer to a question of ours?

It would be! And it is.

Want to learn more about knowledge graphs and their application in the enterprise?