ontotext

Semantic information extraction is a boring name for a fascinating task: pulling out meaningful data from textual sources.

Semantic Information Extraction and the New Digital Disorder

In his book “Everything is Miscellaneous – the power of the new social order”, David Weinberger writes:

> When you have ten, twenty, or thirty thousand photos on your computer, storing a photo of Aunt Sally labeled “DSC00165.jpg” is functionally the same as throwing it out, because you’ll never find it again.
>
> Cit. Everything is Miscellaneous – the power of the new social order, p. 13

Now, take this understanding and extrapolate it on a corporate plane, where every day thousands of emails, customer service records, presentations, logs of calls, supplier lists, employee records and a myriad of other texts and text chunks flow around the business unutilized.

Functionally, they all end up where Aunt Sally’s photo does: in the trash. Or at least filling someone’s computer, deemed to oblivion.

The way out of this oblivion is semantic information extraction – a brave new trail to blaze, where textual sources are utilized properly and they fuel data-driven visions, conclusions and discoveries.

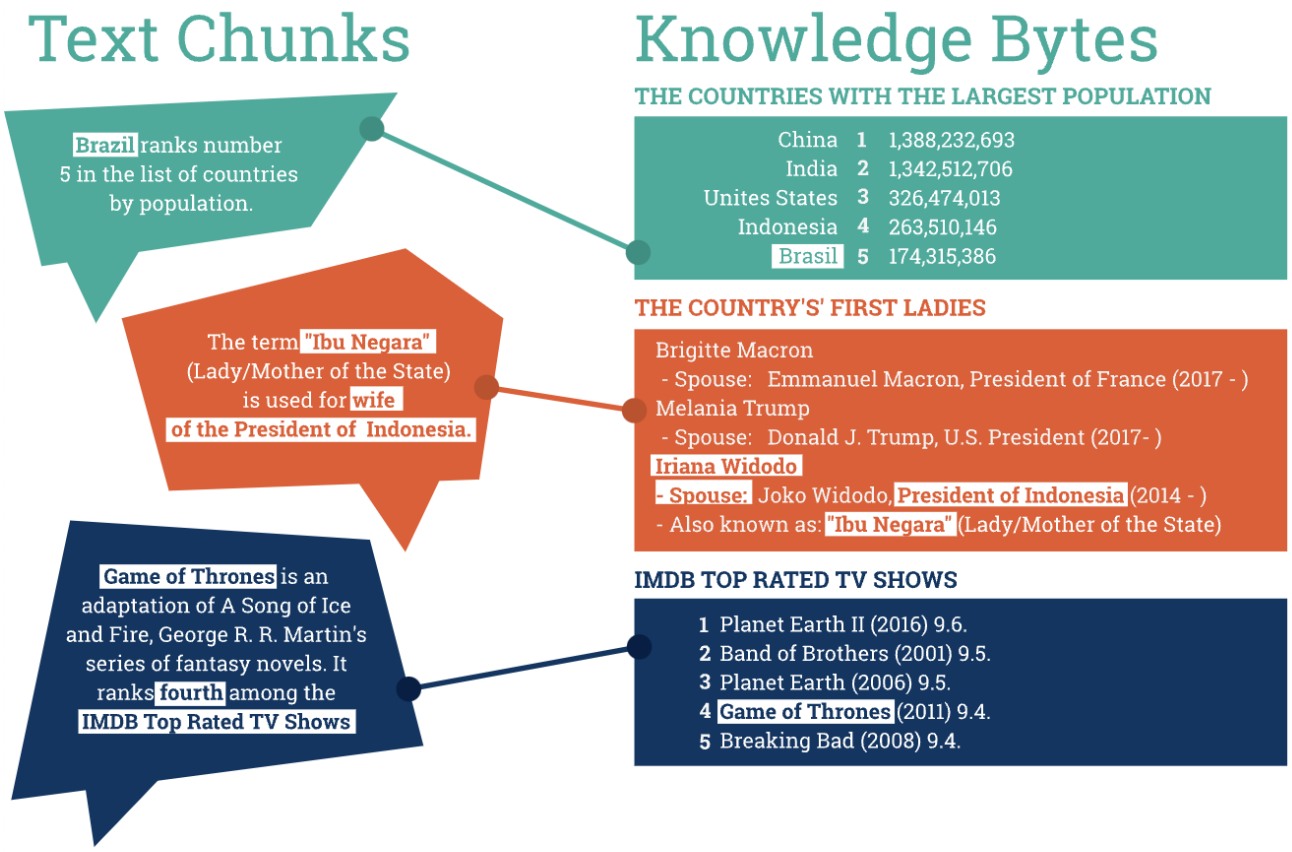
With semantic information extraction, text chunks become data bits, data bits become semantic metadata and semantic metadata become knowledge bytes – data pieces, ready to be leveraged for insights, decisions and actions.

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

ontotext

From Text Chunks to Data Bits: Traditional Information Extraction

Traditional information extraction turns text chunks into data bits, which involves finding and classifying pre-specified names in texts in order to extract and gather clear, factual information.



Typically, information extraction is applied to free-flowing textual sources, such as legal acts, medical records, social media interactions and streams, online news, government documents, corporate reports. By translating these into structured, machine-readable data, information extraction enables content classification, integrated search, content management and delivery.

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

ontotext

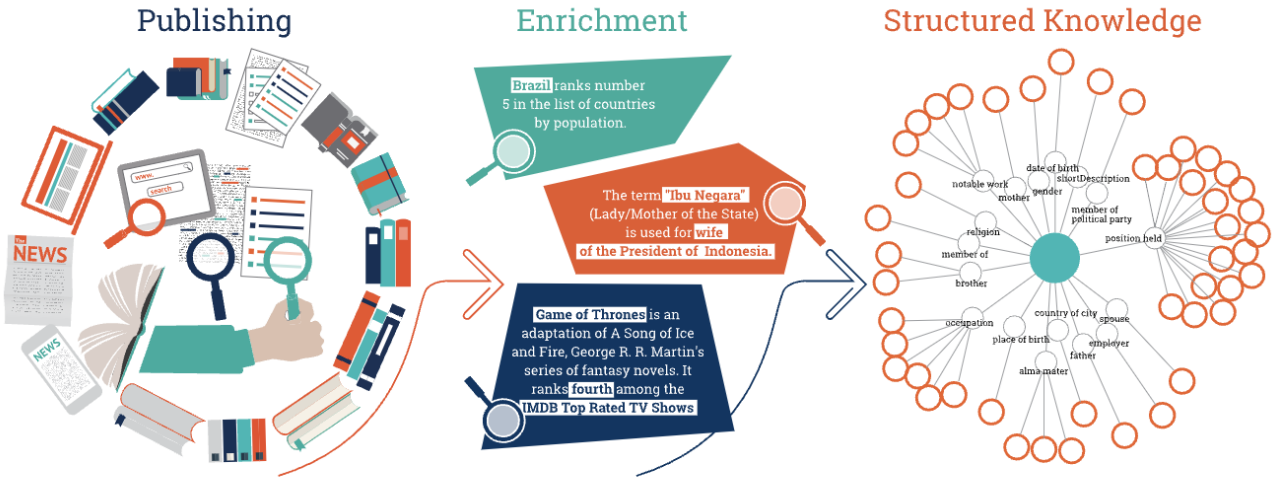
As useful and valuable as this process is for many tasks that benefit from automation such as gathering structured information from multiple sources, media monitoring, drug discovery, scientific research and more, information extraction can be even more powerful.

Integrating semantic technologies in the traditional information extraction process tames the powers of the content hurricanes our digital world exposes us to and uses their force to create knowledge.

The Road Less Travelled: From Data Bits to Knowledge Bytes

For data bits to become knowledge bytes, semantic information extraction comes into play. Share on X

To the traditional information extraction, where texts are transformed into data pieces, it adds another layer of richness in the representation of texts as data, turning them into semantic metadata, that is, into knowledge bytes.



Semantic information extraction, also referred to as semantic annotation or semantic enrichment, makes the shift to the next level by including semantics to the information

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

ontotext

extraction process. Thus textual sources are not only converted into machine-processable facts, but further enriched with machine-readable links, references and relationships.

With semantic information extraction, capturing and making sense of all sorts of data is much more effective than other alternative approaches. It is not a silver bullet for enterprise knowledge management, but is still a powerful tool for connecting, integrating and analyzing – where data bits become knowledge bytes.

Linked Data For Lean Enterprise Data Management from **Ontotext**

From Knowledge Bytes to Everywhere

Semantic Information Extraction revolutionizes the way we think about textual sources. It helps seeing texts, scattered across the web and corporate intranets – every document, every business record, every email – as an asset.

Put together, these small pieces of information from a disparate range of textual sources add up to a 360-degree view of an organization, of its content and its context. This opens up many opportunities for interactive representation and use of content, as well as for a super efficient search that enables accomplishing certain tasks in minutes.

To mention just a few applications for which semantic information extraction lays the foundation:

- integrated search across all sorts of textual data;
- automatic relationship discovery;
- content recommendation;
- discovery of references to concepts and entities;
- integration of disparate and seemingly unrelated sources.

Turning texts into data bits allows algorithms to enter the processes of risk management, fraud detection, retrieving of facts and statistics, investigating connections, keeping up with compliance standards, tracking consumer behavior and much more.

The moment textual sources are translated into the language of semantic metadata and

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

ontotext

further structured into a Knowledge Graph, the overwhelming digital mazes of content suddenly transforms into a well-structured organized space with integrated data pieces, ready to become understanding, actionable information and ultimately knowledge.

Want to learn transform text-associated hurdles into content-related opportunities?

White Paper: Text Analytics for Enterprise Use

Download Now

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