Author: Jarred McGinnis

The history of computing has largely been a history of humans forced to follow the whims of our machines. No humans used binary until Leibnitz. Even Leibnitz didn't find much use for it. There were attempts at decimal computers in order to be closer to the numeral system used by humans but they didn't last. We are forced to use binary computers, because computers use electricity and voltages can vary. To make more reliable computers, they were designed with two states, with (aka 1) or without (aka 0) voltage. We've been bound to communicate with computers in the rather ugly and verbose manner of zero and ones ever since.

I don't think anyone on the Apollo missions would have chosen punch cards, a technology originally used for programming Victorian industrial looms, to take man to the moon fifty years ago, but that's what they had to do. The progress of human-computer interactions is the slow and steady movement toward higher levels of abstraction and less bound by the underlying implementation. The keyboard enabled users to type in letters of a word in a human language rather than raw binary. The mouse turned our screens into a metaphor for a two dimensional plane. Humans were required to make the conceptual shift to understand that a move on the x-axis of our desks meant the cursor would move on the z-axis on our screens. Humans are clever like that. We are adaptable where machines are rigid.

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

Bits and Bytes into Knowledge and the Virtuous Cycle

As the physical interactions between humans and computers move toward the more human-centric and intuitive models such as gesture computing, augmented reality and embedded computers, an important gap remains between man and machine, understanding. Knowledge-driven human-computer interaction is closing that gap with ontologies, formal definitions of knowledge and inference. Knowledge graphs have become essential to getting machines to understand the needs of humans.

Text Analysis for Content Management solutions are at the forefront of knowledge-driven computing. By using a knowledge graph database like GraphDB and natural language processing (NLP), content becomes connected, dynamic, meaningful and contextual. This enables the automation of knowledge tasks as well as empowering analytical tools for human experts to discover insights and make decisions.

There are a number of approaches Ontotext uses to transform documents and data of all flavours (e.g. structured and unstructured) into a format that is accessible as interconnected knowledge. The solution used will always depend on the use case and the data of the client.

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

Ontotext has years of experience working with clients to tailor a solution that best addresses the organization's needs. This has resulted in a proven methodology and tools to deliver these solutions.

Text Analysis

Our content is understandable to us humans because we understand concepts like words, parts of speech and even the meaning of the blank space between a set of letters. It is still all zeros and ones to the machine. Text analysis is a generic term for various processes that order and structure the vagaries of human language into a format interpretable to computers. Ontotext's unique offering is to integrate a pragmatic approach to text-analysis with GraphDB's inference to close the gap between our language and the computers.

Often the volume of content and specific data mining challenges requires Ontotext to employ machine learning techniques. This process often, but not always, involves human subject matter experts going through samples of documents and performing the same task that the machine would be expected to perform. This acts as the gold standard by which the machine learning algorithm is trained. It's a highly intensive manual process but can result in high levels of accuracy and precision.

The distinguishing feature of Ontotext is that these NLP processes are integrated with a knowledge graph. The extracted entities and their relationships are put into the database providing a formal definition of those entities and their relationships, which can be reasoned about. The computer not only knows 'Joe Biden' is a thing called a 'Person' but that he has a relationship to another entity called 'The United States'. By virtue of being defined as a person, the database reasons additional information about 'Joe Biden' not explicitly stated in the source content. The beauty of this approach is that the newly extracted entities and their relationships are added to the knowledge graph, which in turn increases the performance of NLP on subsequent documents.

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

Content Management

To a machine our most valuable assets and most important documents are a gray goo of bytes. And yet, to companies that grey goo is black crude in the ground. Companies can and do see the worth in their content and data, but it requires the collection, refinement, management and delivery to extract that value. Depending on the requirements of the business case, the content management solutions that Ontotext delivers can be categorized into a number of tasks:

- Document Classification – categorizing your documents against a standard or your own taxonomy;
- Named Entity Recognition – extracting different classes of entities such as people, places, and companies;
- Relationship Extraction – discovering relations between entities such as ‘works for’, ‘located in’, ‘employees of’.
- Recommendation Services – providing contextual or behavioural recommendations;
- Semantic Search – providing faceted, free-text and knowledge-based search capabilities.

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

All these solutions are tackling the semantic gap between the human and computer views of content. To stay with the crude oil metaphor, if text analysis is the extraction then content management is the refinement and the pipelines delivering the product. Solutions like *document classification* ensure that the right content gets to the right audiences. Academic publishers often work across a broad range of disciplines and the ability for the computer to automatically understand the differences between papers on 'Yukawa Coupling' and 'Coupled Map Lattices' is valuable for the user experience and satisfaction.

Name entity recognition and *relationship extraction* have been shown to be vital for intelligence tools for publishers in the finance industry due to the entity-richness of their content and the monetary value to their subscribers to identify relationships before others. *Recommendation services* and *semantic search* are inarguably increasingly valuable as the exponential growth of content continues but at any given moment, we are searching for 'that one thing' we need.

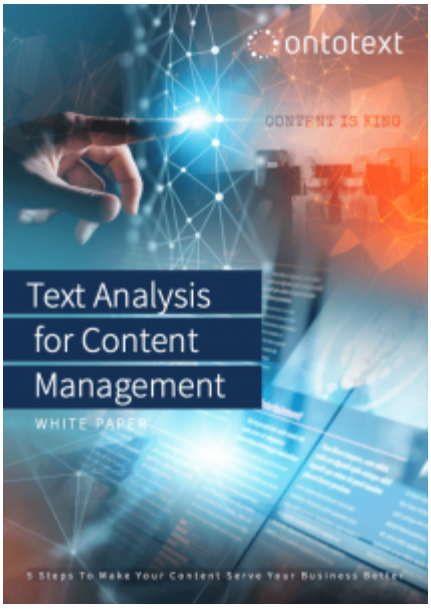
Human-Computer Collaboration

As the means of human-computer Interaction becomes closer and closer to the way humans actually interact with each other, this will necessitate the machines understanding the world that concerns them. The physical interactions are becoming closer and maybe even too close.

Knowledge-driven computing such as Ontotext's content management solutions will be essential to closing the semantic gap between us. It will no longer be us teaching ourselves the idiosyncrasies of computers or interpreting outputs with respect to the real world. With semantic technology we will move toward human-computer collaboration rather than just interaction.

Do you want to learn more about Ontotext's content management solutions?

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



White Paper: Text Analysis for Content Management
Learn how we can make your content serve you better!

Download Now

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