Text Analytics for Context Intelligence: How to get more relevant insights from your unstructured data.
Following last month’s article ‘Unstructured Data – Analytics: next Frontier’ a lot of questions came on how textual analysis helps making sense of unstructured data.

The question has been around for some years.

Unstructured data represents roughly 70% to 80% of all data available to enterprises. Text analytics and context intelligence technologies such as Squirro allow users to extract meaning from unstructured data.

With today’s abundant computing power and emphasis on algorithms ever more precise statistical approximations are calculated. The resulting patterns are easily worked with in order to discover relationships and analyse unstructured content.

The ability to identify the most relevant information in unstructured data produces tremendous benefits: Cutting down on the amount of time knowledge workers dedicate to finding the intelligence that matters, enabling entirely new levels of decision making.

What is text analytics?

Text analytics is the process of analysing unstructured text, extracting relevant information, and transforming it into structured information that can then be leveraged in various ways. ¹

Because of the explosion of electronic data available the capacity to extract relevant information from large unstructured data sets becomes increasingly crucial.

Text analytics is based on the extraction of information implicitly contained in collections of documents or similarity-based structuring and visualisation of large sets of texts. ²

There are many principles and techniques used in text analytics and here we will focus on two main ones, which are at the core of the Squirro technology.

- Feature selection:

According to Wikipedia feature selection is: “the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. Redundant features are those, which provide no more information than the currently selected features, and irrelevant features provide no useful information in any context. Feature selection techniques are a subset of the more general field of feature extraction.” ³

- Entity extraction:

It is a subtask of information extraction that seeks to locate atomic elements in text. It is often associated with entity recognition in which those atomic elements are then classified into predetermined categories. ⁴

- Phrase detection:

The aim of phrase detection is to extract from texts sequences of words, which occur together more times than we would expect co-occurrence due to chance. ⁵

Let’s use the following text to illustrate the three principles and see how they are integrated in Squirro.

“Women’s entrepreneurship has hit a media tipping point. The question is: Is it just a passing media fad that will soon be a blip on the radar screen, or is it actually a real, fundamental economic force that’s reshaping the world? I think it’s safe to say that it’s the latter. Women-owned entities in the formal sector represent approximately 37% of enterprises globally — a market worthy of attention by businesses and policy makers alike.

(...) Entrepreneurial activity creates growth and prosperity — and solutions for social problems. And today’s trends show that women will be a driving force of entrepreneurial growth in the future.” ⁶

By using feature selection and phrase detection Squirro is able to create a “Smart Filter” for the above text in which it identifies the main features of the text and gives them specific weight.

This is graphical representation of the selected features as part of a “Smart Filter”, where the most representative key elements form the basis of the filter and their relationship to each other and to the entire text give a specific weight to each (Figure 2).

Entity extraction is exemplified by the phrase Global Entrepreneurship Monitor (GEM). Where GEM is associated with Global Entrepreneurship Monitor, the atomic element.

These principles contribute to the quality of the “Smart Filter” Squirro develops for any piece of unstructured data.
What are the benefits of text analytics and “Smart Filters”

- Reduction of research time:

Using simple key word based search to look for information about a company, a brand or a market requires a lot of time. Searches need to be made in different languages and results need to be parsed for homonyms and duplicates in order to achieve some level of precision.

Instead text analytics provide the toolset to implement a more powerful enterprise search, where you do not have to manually phrase search queries (e.g. using Boolean operators), therefore the time needed to search through unstructured documents is drastically reduced.

These principles permit the creation of “Smart Filters” in Squirro that can be applied to any document to see which ones are the best matches.

By incorporating this technology in Business Intelligence dashboards or in Customer Relationship Management systems (CRM) for example users do not need to make any searches anymore, instead Squirro reads the elements selected in those systems to create “Smart Filters” to deliver within a dashboard or a CRM instance the most relevant information.

Information updates are made in real-time and users get all they need within one workspace, thus saving them up to 90% of time in the search for important insights.

Technologies like Squirro provide an easy to use environment to extract and curate knowledge from unstructured textual sources and deliver relevant insights for your business.

- Simplification of processes:

The corollary of research time reduction is the simplification of processes whereby information is gathered and shared within a company and between co-workers.

Text analytics and context intelligence in particular improve the way work is done by providing better information.

Not only the information is delivered in real-time, but by using “Smart Filters” the information is more precise and more relevant.

Conclusions:

Text analytics is at the core of Context Intelligence technologies like Squirro. It enables the access to the most relevant elements of unstructured textual data.

It is those elements that permit the understanding of that data and therefore empower knowledge workers to work with it to get better insights.

The consequence for companies is that they can now utilise all the unstructured data they have at hand. Use the insights from that data to have better understanding of the information they generate or consume, therefore having better and more effective decision-making processes.

---

2 Text Mining - Knowledge extraction from unstructured textual data, Martin Rajman and Romaric Besançon, http://iawww.epfl.ch/Publications/Archive/RajmanBesancon98a.pdf
About Squirro

Squirro is the leader in Context Intelligence, combining structured and unstructured data to provide the ‘Why’ behind the data. Squirro brings the relevant context from the sea of information directly to your regular workplace.

What Business Intelligence systems did for numbers, Squirro does for content: make unstructured data usable. ‘So What?’ - because achieving this reduces searching time by 90% and allows for better, more effective decision-making.

The highly skilled Swiss team of search experts has been working together for over 10 years to create a precise software engineering solution that delivers a real-time, self-learning embedded 360° context radar.

Squirro: Your data in Context.

Contact

Squirro by Nektoon AG
Badenerstrasse 120
8004 Zurich
Switzerland

Tel CH: +41 44 586 98 98
Tel US: +1 650 353 93 68
Email: info@squirro.com
Web: www.squirro.com