BMC Bioinformatics

ORAL PRESENTATION   **Open Access**

Pitfalls in applying text mining to scientific literature

Jean-Marc Neefs

Numbers and data mining are easy. Our numerical system counts 10 digits, any combination is possible, and every measured value can be captured in a number. Large quantities of measures can be analysed efficiently using incredibly powerful calculators, and resulting information can be shown is simple clear graphs.

Text is hard. Hundreds of letters and millions of different combinations can be used in the personal interpretation of information, in words and phrases that reflect one's personality rather than objective measurements. Depending on context and language, the same expression carries totally different information, or no meaning at all.

Text Mining requires 'education' at different levels: for providing information, to capture, to store and to retrieve that information, and to interpret results of the mining process.

I will provide a few examples of a few text mining tools in daily practice.

Submit your next manuscript to BioMed Central and take full advantage of:

• **Convenient online submission**
• **Thorough peer review**
• **No space constraints or color figure charges**
• **Immediate publication on acceptance**
• **Inclusion in PubMed, CAS, Scopus and Google Scholar**
• **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

Correspondence: JNEEF1@its.jnj.com
Janssen Pharmaceutica, 2340 Beerse, Belgium

BioMed Central