
Petri Vasara
Jaakko Pöyry Consulting (Finland)
April 21, 2004

Who:

An executive at the Finnish global management consultancy Jaakko Pöyry Consulting, and also an active researcher, managing and creating projects e.g. around data mining. Special areas include weak signals, technology surveillance and strategies, environmental trends and industrial policy and politics. Has an M.Sc in Technical Physics, a Lic.Tech. in Computer Science and a Dr.Tech in neural networks from Helsinki University of Technology. Private e-mail address: Petri.vasara@pp.inet.fi

Why text analysis:

In an earlier research project, we used neural networks to process pictures, data and text simultaneously. It was during the .com era, and the example application was clothes sold over the Internet. Each object, i.e. piece of clothing, had a pictorial representation, a textual representation and a numerical evaluation. Here, we did our first textual mapping. Before that, I had done some semantic analysis on environmental discussion over the Internet – and found my own name linked by the computer to surprising people and issues. This was the start of the decision to compile a toolkit for textual analysis – some purchased, some custom-made. Currently, our datamining of text using neural networks is done using proprietary algorithms, and T-LAB is the main tool for semantic analysis. Most of our work deals with material spidered from the Internet or generated, in electronic format, by different types of organizations. Thus, news items, journal articles, newsletters, opinion pieces, patents are all among the variform data analysed. To a lesser degree, we have analysed people's replies to interviews. There is a simple reason: the interview material is often so limited, that using mathematical tools is overkill.

Why T-LAB:

In collecting the toolkit, I went through and tested a large set of tools. T-LAB seemed to me to have e.g. the following advantages: the largest amount of analyses, continuous improvement of the software, suitable graphs and tables for further processing, stability - and personal contact at the vendor site.

Research done:

I'm writing this, coming from a seminar on hybrid media (i.e. the area of combinations and convergence of different digital and paper-based media), where I presented the conclusions of a study. As the definition of hybrid media is far from a consensus, a textual analysis of texts by researchers in the area was undertaken with T-LAB to find the common elements and verify the different schools of thought in the area as part of the

study. Many of the studies and analyses done have been for corporate clients and cannot be detailed here. One which was made public deals with structuring the discussion (media, experts, newspapers, NGOs and other activist groups) around the RFID (Radio Frequency ID) technology, its trends and perception. This “smart tagging” technology has many privacy advocates in uproar, while it promises to make possible great advances in information flow in the supply web.

Overall, T-LAB is mostly used by me for structuring highly unstructured discussion; for trying to separate the important from noise; for detecting patterns where none seem apparent.