

A PROJECT MANAGER TELLS HIS STORY

SCIENTISTS AS TDM-CUSTOMERS

Stephane Schneider, IT project manager at INIST-CNRS



WHEN DID YOU HEAR ABOUT TEXT AND DATA MINING FOR THE FIRST TIME?

This was at school, where I got a graduate degree in computational linguistics. However, the tools and methods have greatly evolved since then. In particular with the advent of complex numeric methods for big data manipulation and the possibility to mix them with traditional NLP technics.

WHAT DO YOU CURRENTLY DO WITH TDM?

I am an IT project manager, specialized in document application design that uses text mining for the analysis and search of scientific documents. At the moment, we are about to develop a new application that will explore a corpus on the ISTE platform. This platform (www.istex.fr) is hosted by INIST-CNRS and has 21 million digital documents (journal articles, book chapters, text corpora etc.) from all disciplines. We focus on geoscience and want to test the feasibility of using an OpenMinTeD workflow for this kind of scientific corpora.

WHY DO YOU THINK TDM IS IMPORTANT?

Because of the technical advances in recent years, our modern world is flooded with information. TDM can help to sort and analyse information and in that way keep us from drowning.

I think researchers could be the first 'TDM customers', as they are by nature big consumers of information. Worldwide, more than 75 million articles are published (Univ Ottawa). Rapidly growing volumes of scientific information make it increasingly complex to find relevant knowledge. TDM can help scientists with this.



I like the idea that the computer contributes to scientific progress, by making sense of and better understanding language.

HOW DO YOU SEE THE FUTURE OF TDM?

The context is favourable, even though there are still barriers. Look, for example, at the legal context in France: the conditions for using TDM make things very complex. Disseminating TDM tools is also a problem. We have to introduce a software management plan, in order to improve the access to and reuse of TDM software.

TDM is a highly technical domain, but most good results come from work of a scientist who is also a TDM specialist. This cross-pollination has to be promoted more. INIST-CNRS would be a good meeting place for this.

WHAT DO YOU THINK IS THE BIGGEST CHALLENGE?

On one hand, text and data miners have to help researchers who do not know what to do with TDM. They can show what's possible, and how TDM can provide solutions. On the other hand, text and data miners need to pay more attention to users, and offer services based on real user experiences.



This resource is published under a Creative Commons Attribution Licence