

## 12/20/09 - Semantic players

I am not here to talk about the little semantic browser thing, which I know all of you are expecting (and I promise a first test version of the app in the next few days, Christmas holidays should help in giving me some time to prepare it right); I'm here to discuss some impressions on the Microsoft Semantic Engine presentation.

Yes, I said "Microsoft", and I'd like to survive to that phrase, and no, I did not change to the dark side of computer science; it happened that, a couple of weeks ago, I was browsing the sessions of this year's Professional Developers Conference, when I found [this one](#), about a so-called "Microsoft Semantic Engine": as a semantic maniac, I had to download the video and see what the "enemy" is preparing in this field, especially because I think that KDE is the first desktop environment to introduce semantics for its apps, and if a new player is coming into the field, then it's good to see what he is doing. So, finally today I found some time to take a look at it, and here are some impressions.

### **In general**

It looks like a technology introduced especially for business users, which is a good place to start: if the semantic desktop works (and here I'm talking about Nepomuk itself: Microsoft never used the term "semantic desktop"), it could really help Linux itself in becoming a business player heavier than how it is now.

For evidencing even more the business aspect, they built their engine on a relational database, which is pretty shocking IMHO: yes, in this way you can integrate with data warehouses, but the underlying code is a hell if you do not use any of the semantic languages/engines/technologies and their facilities (RDF(S), OWL, Jena, Virtuoso, Pellet). Triples, anyone?

### **Clustering**

This engine also uses clustering and data mining techniques, together with (un)supervised learning algorithms, to deal with all the metadata that various crawlers pick up from documents (more on this later), and this reminded me of a discussion we had in Freiburg, about clustering for obtaining meaningful facets and terms for the semantic browser: for now, that prototype application will work with just three facets, but when that number will grow, things will surely become interesting...

### **File crawler**

The crawling part is quite straightforward: each document is analyzed many times, at different levels, for getting keywords or other meaningful informations; there is an OCR part for images, as Scribo does, and a really interesting audio analyzer, which tries to extract also the tempo and the key from the music (if it is a music audio file, of course), and that's really interesting from a technical point of view, at least for a musician like me.

## **Conclusions**

In the end it has been an interesting video, with some really interesting informations on how other software vendors are dealing with the semantic technologies; I think we (as KDE) have a great advantage on these "competitors" in this field, and we need to keep up the good work and integrate Nepomuk more and more into all the applications: the future is coming, and we are right on the bleeding edge.