

ADVANCED SEARCH TECHNIQUES BASED ON AN EUROWORDNET APPLICATION, THROUGH DEMONSTRATIONS OF THE OMNIPAPER PROJECT RESULTS, AND TECHNOLOGICAL APPROACHES AS IMPLEMENTED IN THE OMNIPAPER PROTOTYPE, INCLUDING SOAP AND WSDL

JAN ENGELEN¹; MARKUS SCHRANZ²

¹Katholieke Universiteit Leuven, Belgium

²Presstext, Austria

This workshop will be devoted mainly to demonstrations and the presentation of results obtained through a novel search technique developed in the European project Omnipaper. This smart way to search has been applied for news sources and the project OmniPaper has tested a prototype that offers 'semantic search' capabilities for Europe's newspapers on the Net. Furthermore, this prototype permits multilingual access to the reams of online newspapers available through News Brokers.

The IST project's first prototype was made available to the public in November 2003, with users able to use the semantic, or 'smart search', for news articles from a set of almost 2000 English-language articles from UK-based The Daily Telegraph. The present day prototype handles multilingual queries, has a novel visualisation and manipulation unit for refining the search queries and now provides the articles in real time (i.e. also today's articles). Smart search means that the search engine first tries to understand your query on a semantic level. It will look for keywords in your query and ask you to specify certain words. If for example your query contains the keyword 'bush' the system asks you what you mean by it: 'a large wilderness area', 'President Bush' or something else. So the idea is that the user gives semantic feedback on his/her query, yielding more accurate search results.

This search mechanism is based on an English-language ontology that operates as a separate knowledge layer on top of distributed news archives. The search engine uses this knowledge layer as an access gate to the underlying information. The use of ontology for semantically searching is quite new but even more the technique of combining searching and navigating in a graphical way is a promising technique that might become standard in future search engines.

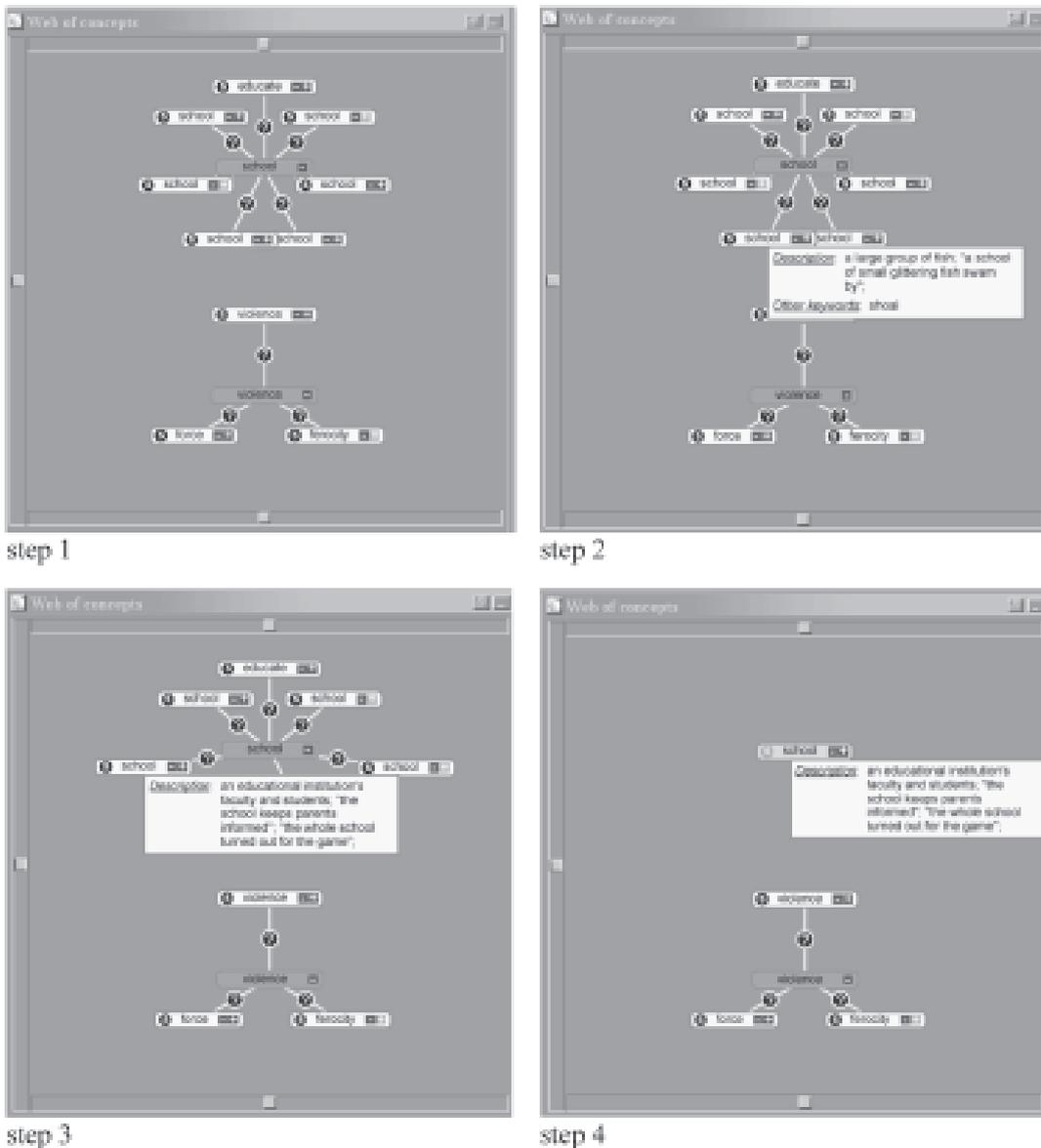
The OmniPaper consortium plans to commercialise these new technologies as an extra functionality offered to news clipping services. In a first phase, these services will offer the smart search function to their customers, while in a second phase, news publishers will be able to resell news from other sources with the OmniPaper technologies for distributed news publishing. A fundamental aspect is that all news brokers, partners in the Omnipaper initiative, remain fully responsible for their content which will be accessed using SOAP technology and therefore no central article database is needed.

IN THIS WORKSHOP

- a) a short overview of the Omnipaper project will be given;
- b) a demonstration of the Omnipaper visual approach to semantic search will be given;

c) several technologies (SOAP, WSDL, SVG), will be explained without becoming too technical;

d) a typical system architecture deployed for distributed information retrieval will be explained and discussed, as applied to the Omnipaper project.



EXAMPLES

1. An example of a **semantic search** is given below.

Suppose we want to look for info on “Violence in schools”. Therefore we can enter the query “violence AND school” in the search box.

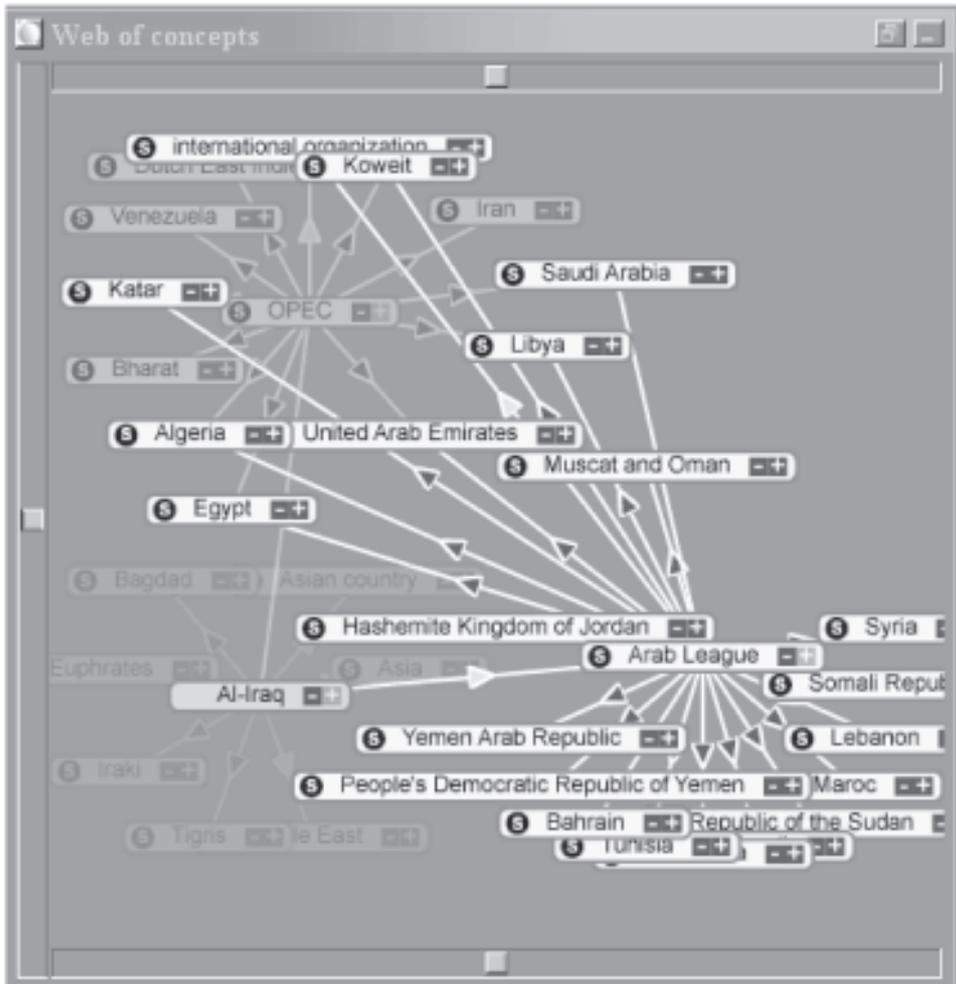
- step 1: The graphical result of the query. The system found 7 different meanings (=concepts) for the query keyword “school”.

- step 2: The user can get more information on each meaning by *mouseover*. A *tool tip* is shown with a more detailed description of the meaning and other keywords associated with this meaning. This meaning does not match the meaning of the user so he can remove this meaning by clicking the red “minus” button. The meaning is removed from the view.

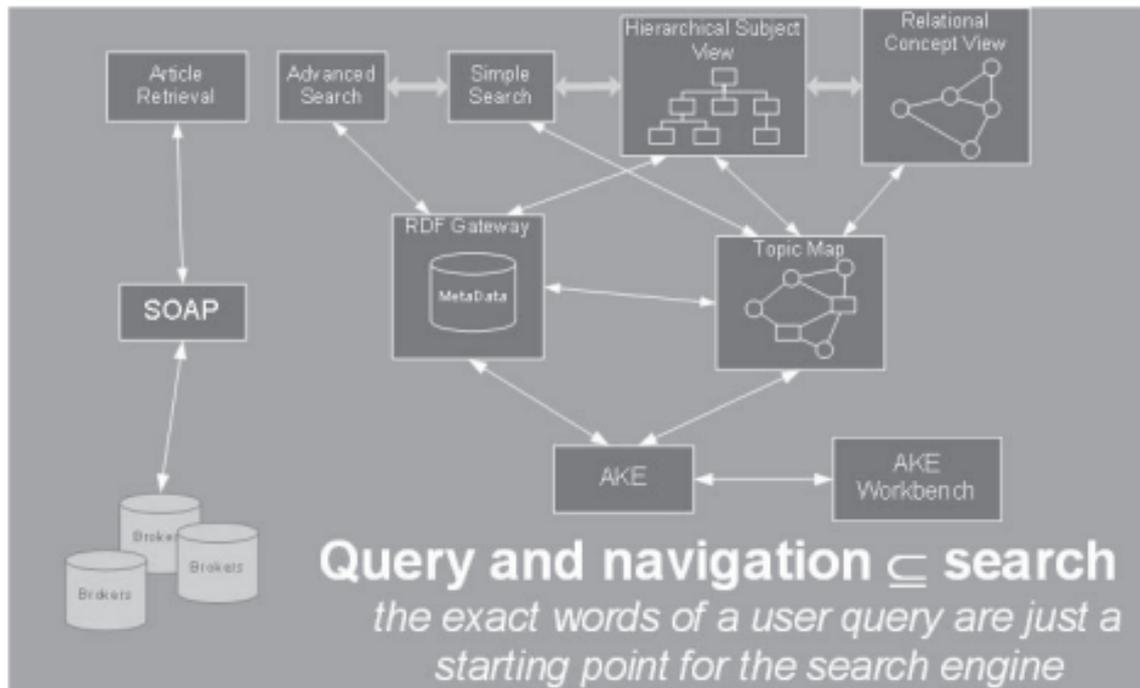
• step 3: The user found the meaning he was looking for. He will now select this meaning by clicking the blue “select” button.

• step 4: As a result of this selection, all other meanings have disappeared and the green query keyword has been replaced with the matching meaning.

For each meaning (represented by yellow rounded rectangles) the user is able to navigate to related meanings by clicking the green “plus” button. This is called an *expansion*. The following figure illustrates the result of a user looking for concepts related to the query keyword “Iraq” after some expansions.



2. Overview of the Omnipaper distributed architecture



ACKNOWLEDGEMENT

The developments within the Omnipaper project are funded by the consortium participants and co-funded by the European Commission, DG Information Society, as part of the 5th Framework programme of Research and Development.