

OmniPaper Smart Information Retrieval Prototype

Bert Paepen

Katholieke Universiteit Leuven, Department of Electrical Engineering,
Research Group on Document Architectures,
Kasteelpark Arenberg 10, 3001 Heverlee, Belgium.
bert.paepen@esat.kuleuven.ac.be

DEMONSTRATION

The OmniPaper project has implemented three information retrieval prototypes in the area of electronic news publishing. One prototype uses SOAP as communication protocol between the central system and a number of distributed news archives. The second prototype uses an RDF metadata database, enabling direct metadata queries to the central system. Finally the Topic Map prototype uses query expansion and semantic linking for smart metadata search. The Topic Map prototype enhances the search experience by implementing a knowledge layer that combines the semantic content of a lexical database, consisting of concepts and keywords, with a metadata-set of newspaper articles.

After developing and testing three smaller prototypes, the OmniPaper consortium has combined these prototypes in one. In this final prototype a kind of “enhanced full-text search” engine is implemented. This means that the prototype is an interface on top of existing search engines. It does three main things: 1) add semantic query support for enhanced searching, 2) enable multilingual queries and 3) provide a graphical “web of concepts” as a semantic aid to the search operation.

In this prototype, querying and navigation are considered as alternative methods to find relevant information. Both interact with each other and together they produce a combined user experience that can be expressed as *find what you were looking for and then browse away from it*. In fact, the prototype considers both querying and navigation as a kind of search action and tries to integrate both. Topic maps have traditionally been used for navigation purposes, but with the emergence of dedicated query languages, they have also become a useful tool for querying information.

In concrete, keywords in a query are looked up in a dictionary and shown to the user in a graphical way. In the background, the keywords are translated and expanded to related terms. These expanded queries are sent to the underlying full-text search engine(s) in all requested languages. In the graphical view (“web of concepts”) users can redefine their query words, resulting in an updated query and result set. Figure 1 shows the initial web of concepts for the query “poll Indonesia”. The word “Indonesia” is recognized in only one concept, “Dutch East Indies”, whereas the word “poll” has many different meanings. If you select the concept “canvass”, this word is replacing the word “poll” in the original query. After selection the concept “canvass” can be expanded to its related concepts: “enquiry” is more general while “straw vote” is more specific.

Figure 1: initial web of concepts

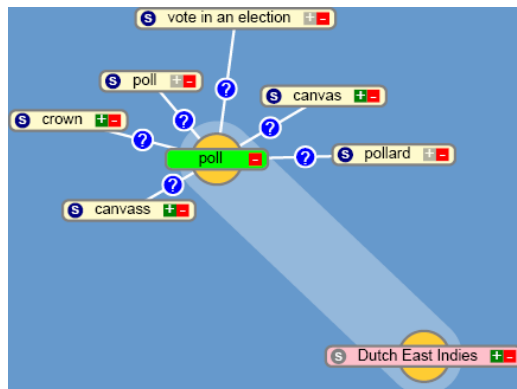


Figure 2: web of concepts after selection and expansion

