Semantic Expansion for Proactive Information Filtering

Eduard Hoenkamp (HOENKAMP@ACM.ORG)
Nijmegen Institute for Cognition and Information (NICI)
Nijmegen, the Netherlands

Publishing on the internet is virtually for free. This has led to a potentially unlimited flow of information of varying quality, the so-called *information overload*. For electronic mail, newsgroups, and web pages, *information filters* have been designed in hopes to let only the relevant part be delivered to the user. Here traditional information retrieval falls short: it assumes that users know what information they need, and how to formulate this need. But with the steep increase in production and distribution, the information may be accessible in principle, but users may not know of its existence. And even if they knew, they might not be able to represent their needs in a machine accessible form. To approach this problem we started the PROFILE project, which investigates: • How to get the information to the users who need it • At the time they need it • In a form they can use. As in all filter projects, we assume that the user has a fairly stable information need. However, the filter is not reactive, filtering out irrelevant information from the incoming stream. Instead it is *proactive*, using the information need to actively search and fill the user in on missing information. A major departure from current trends in information filters is to leave the linguistic level, and to rely on a deep conceptual model of users and their information needs. This abstract only deals with one part of PROFILE (Hoenkamp, Schomaker, van Bommel, Koster, & van der Weide, 1996): the part that pertains to the representation of the information need and how to satisfy it.

Filters employ a *user profile*, a representation of the user's more or less stable information need. The representation is typically in the form of (weighted) keywords, necessary for the indexes of search engines on the web. Similarly, users typically have to translate their information need into keywords, iterating until a reasonable amount of documents is presented. (An excellent example of users adapting to the technology instead of the other way around.) We decided that the information need, and hence the user profile, should be represented on a deeper semantic level. Just as some keyword based systems use *linguistic expansion* on phrases using thesauri (e.g. WORDNET), we defined *semantic expansion* on the semantic representation. This offers the following advantages:

- a semantic representation can be expressed in different languages, thus not limiting retrieval to one particular language,
- the information need may sometimes be far more effectively rendered by other media such as sound or graphics,
- retrieval may be based on a semantic expansion of the current semantic representation.

At the semantic level, it is also easier to add context, to introduce the user's goals, and to combine several conceptual structures into one.

In the actual implementation of our system, we had to choose an ontology in which to express the user model and the user's information needs. We decided to use the *Stanford Ontology Server*, which provides a flexible representation language to define the user's information needs. The system provides three desirable services that we did not readily find in CYC: 1. distributed groups can build and share ontologies, 2. ontologies can be converted to representations that one typically finds in other filter projects, and 3. it offers a protocol that lets our program for semantic expansion directly interact with the ontologies at the server. For example, when the program needs to expand the particular semantic structure representsing a user's current information need, it can first ask the ontology server for related structures, and then translate these into queries to the net.

One important expansion mechanism we currently experiment with, uses of the concept of *basic level objects*. Somewhat simplified: in a hierarchy of objects, an object at the basic level adds many features to the level above (e.g. from 'furniture' to 'chair'), but very few are added to the level below (e.g. from 'chair' to 'kitchen chair'). Whether a basic level is constant over people or situations is heavily debated. What is important, however, is that a given basic level is not restricted to words describing a common category; it also extends into other domains such as movement, shape, or usage. For example if we needed to know about kitchen chairs, we might extend to the basic level, to find documents about chairs, but also to find pictures of things to sit on, or that look like a chair etc. In the actual implementation we investigate similar mechanism for compound concepts. This way we could already retrieve unexpected but relevant information.

References

Hoenkamp, E., Schomaker, L., van Bommel, P., Koster, C., & van der Weide, T. (1996). *Profile - A Proactive Information Filter* (Tech. Rep. No. CSI-N9602). Computing Science Institute, University of Nijmegen.