Detecting an 'Anomalous State of Knowledge' for Proactive Information Filtering

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

Information filters for the world wide web (WWW) require that users formulate a more or less stable information need. Often, however, an 'anomalous state of knowledge' (Belkin & Kwasnik, 1986) occurs: the user needs information but is not aware of this. With the rapid growth of information available on line, it is likely that this need could be filled, if only there were a mechanism to detect this need on the user's behalf. We implemented such a mechanism in an agent architecture, as part of the proactive information filter PROFILE (Hoenkamp, 1997). The mechanism subsumes three modules:

Module 1: a microtheory of the user. Most filters and search engines express the user's information need at the linguistic level (as keywords). PROFILE, however, uses a deeper semantic representation. We use the *Stanford Ontology Server* to encode the user's knowledge in a taxonomy and axioms. Only the knowledge that pertains to a recurring information need is encoded, hence we refer to it as a 'microtheory' (cf. CYC).

Module 2: WordNet. WORDNET is an on-line lexical reference system to use under program control (Miller, 1995). English words are organized into synonym sets, each representing one underlying lexical concept, or 'word sense'. Concepts in the microtheory of the user may point to word senses, providing words (nouns in PROFILE's case) to express the concept.

Module 3: Spectrum analysis. Traditionally, information retrieval relates documents via the words they contain. However, several researchers have proposed techniques to capture the content documents have in common. Examples include factor analysis, clustering, and the recent 'latent semantic indexing' or LSI (Landauer & Dumais, 1997). As they all rely on the computation of the set of eigenvalues or *spectrum* of a matrix, I use the term *spectrum analysis* as a generic for such techniques. Thus, the common ground of a collection of documents is characterized by its spectrum.

Figures 1 and 2 explain how the modules cooperate.

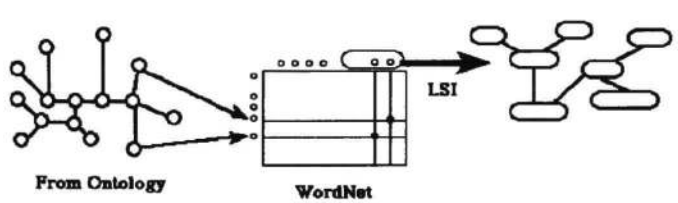

Figure 1: Detecting an anomalous knowledge state step one. (1) express concepts in the microtheory as words via WORDNET, (2) query WWW, and (3) apply spectrum analysis (e.g. LSI, clustering).

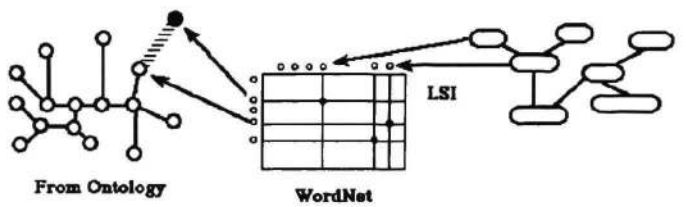

Figure 2: Detecting an anomalous knowledge state step two: (1) feed back representative words from the spectrum, (2) look up the corresponding word senses, (3) word senses originally absent, suggest lacunae in the microtheory.

References

Belkin, N. J., & Kwasnik, B. H. (1986). Using structural representations of anomalous states of knowledge for choosing document retrieval strategies. In *Proc. of the 9th International Conference on Research and Development in Information Retrieval* (pp. 11–22). ACM.

Hoenkamp, E. (1997). Semantic expansion for proactive information filtering. In *Proc. of the 19th Annual Conference of the Cognitive Science Society* (p. 951). Ablex.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*(2), 211–240.

Miller, G. (1995). Wordnet: A lexical database for english. *Communications of the ACM, 38*(11), 39–41.