

Expertise in Workgroups

(English abstract of 'Expertise in Werkgroepen')

Willem Thoonen

August 22nd, 2006

Summary

Expertise and experts are concepts that are tricky to define. Many definitions have been proposed in the literature and they each highlight different aspects. Sternberg (1994) [4] defines expertise as superior analytical and creative abilities, and the ability to process and apply new information faster than non-experts. A century of psychological research into the field has focused mainly on skill at chess, which can be readily measured and subjected to laboratory experiments [3]. Expertise in softer, non-competitive areas such as knowledge organization and dissemination in universities and commercial organizations has been researched using the basic toolkit consisting of interviews, questionnaires and social network analyses.

In more recent years, automatic methods for expertise extraction en expert identification have been proposed such as the link-based analysis of e-mail messages by Campbell et al. (2003) [2] and the Enterprise Search track of TREC 2005 [5]. Another automatic method for extracting expertise within workgroups called authoritative re-ranking by Bogers en Van den Bosch (2006) [1] uses term co-occurrence metrics and TF-IDF to assign term weights that represent an author's expertise and use these weights to calculate query-dependent expert scores for each author. In addition, they use these expert scores to successfully re-rank search results within the workgroup.

However, their entire approach is optimized on the combination of both tasks: the quality of the expertise extraction is implicitly evaluated by looking at the quality of the re-ranking. In this thesis, the expertise extraction component of their authoritative re-ranking approach is evaluated separately. An online expertise questionnaire was constructed and presented to the current members of the ILK workgroup (one of the focus workgroups in the original authoritative re-ranking research). This questionnaire consisted of four different parts, each with different goals.

In the first part, participants were asked to provide at least 5 terms that they thought represented their expertise. On average, respondents provided 7 terms which were

often bigrams and trigrams that were good descriptors of research areas but did not occur often in the automatically extracted lists of expertise terms.

The second part of the questionnaire had the participants evaluate automatically extracted terms that were supposed to represent their authority. Two different combinations of extraction settings were used: the optimal settings from [1] and stricter settings that produced more bigrams and trigrams and less noise. Participants favored these terms more than the terms that were optimal for re-ranking.

In the third part, workgroup members were asked to rate each of their publications with regard to how well they represent their expertise. There appears to be a negative correlation between these ratings and author position: first authors of a paper rated that paper as more representative than authors further down in the list. Participants also indicated that they believe that publications offer a good indication of expertise in general.

In the final part of the questionnaire users were presented with ten queries and asked to identify the current ILK experts on the query topic and rank them according to the perceived expertise on that topic. Compared to the automatically extracted expert rankings, ILK members more clearly identified a small number (around 3) of core experts whereas the computer identified 5 experts on average with more variation. However, the top expert was identified in the top 3 of experts by the computer for all ten queries.

This result and the results of the comparison in the other three questionnaire parts seem to suggest that automatic expertise extraction and expert identification is possible. However, humans and computers use different kinds of terms (general bigram and trigram descriptors versus large collections of specific single word terms).

References

- [1] T. Bogers and A. van den Bosch. Authoritative Re-ranking of Search Results. In *Proceedings of the 28th European Conference on Information Retrieval (ECIR 2006)*, vol. 3936 of *Lecture Notes on Computer Science*, pages 519–522, Berlin, April 2006. Springer Verlag.
- [2] C. Campbell, P. Maglio, A. Cozzi, and B. Dom. Expertise Identification using Email Communications. In *Proceedings of CIKM2003*, pages 528–531, New Orleans, LA, 2003.
- [3] P. E. Ross. The expert mind. *Scientific American*, July 2006.
- [4] R. Sternberg. *Expertise in Context*, chapter Cognitive Conceptions of Expertise, pages 149–162. MIT Press, Cambridge, MA, 1994.
- [5] TREC. TREC Enterprise Track, 2005. Visited: October 2005, http://www.ins.cwi.nl/projects/trec-ent/wiki/index.php/Main_Page.