

Design and Implementation of a University-wide Expert Search Engine

Ruud Liebrechts and Toine Bogers

ILK / Tilburg centre for Creative Computing, Tilburg University, Tilburg, The Netherlands



TILBURG UNIVERSITY

1 Introduction

Problem

The current Tilburg University information sources are inefficient and ineffective for locating experts and/or thesis supervisors.

Current information sources

- *Webwijs* – database of university researchers (part of the UvT Expert Collection (Balog et al., 2007))
- UvT Repository – publications and theses archive
- UvT website search engine

Objective

Design, development, and evaluation of a university-wide expert search engine

- that supports both expert and supervisor finding
- better than the current systems
- in a bilingual setting (Dutch and English)
- with high satisfaction of both users and university researchers

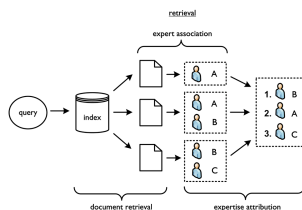
2 Data & Methodology

Data

- 28,641 publications (UvT Repository)
 - 19.3% with full text
 - Approx. 50% Dutch, 50% English
- 6,656 Bachelor and Master's theses (UvT Repository)
 - 43.7% with full text
 - Approx. 75% Dutch, 25% English
- 1,944 experts (UvT address book)

Algorithm

- Document-centric expert finding (Balog et al. (2006))



- Document search side done using Indri 4.4

3 System-based evaluation

Query set

Created 120 Dutch and 120 English queries with relevance judgments based on known associations between documents, topics, and experts.

Optimized parameter settings based on these query sets. Best-performing settings gave MAP of 0.6757 and NDCG of 0.7755.

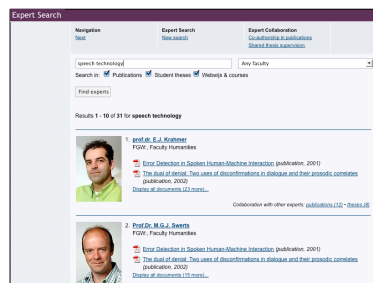
Some findings

- Data sources
 - Publications contribute most towards expert finding
 - All sources combined lead to best performance
- Field weighting (e.g. title or full text) did not help
- Citation counts did not improve performance

4 Expert-based evaluation

Prototype evaluation with real UvT researchers

- 30 researchers from all faculties
- To obtain real relevance judgments
- To measure satisfaction and gather feedback



Survey

- List one of their expertise areas
- Rate their own expertise, and list and rate colleagues
- Evaluate the search engine's performance

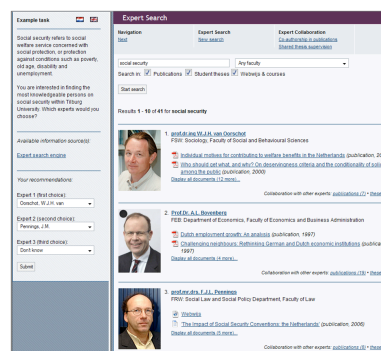
Results

- High satisfaction: 3.8 avg. on 5-point Likert scale
- Updated MAP of 0.80 and NDCG of 0.81 with new relevance judgments

5 User-based evaluation

Research questions

- How do users interact with the search engine?
- Does the search engine outperform the current sources?



Experimental Setup

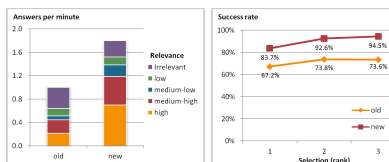
Participants were asked to perform 3 expert finding and 3 supervisor finding tasks using either the expert search engine (new system) or any combination of the current information sources (old system). Systems to use were randomly assigned.

Participants (101 in total)

- Internal group: 57 university students
- External group: 44 high school students

Results

- On average 2 queries per task
- All data sources were used in 80% of the queries
- Few clicks on evidence links to documents
- Most clicks on top 3 candidates
- Users tend to stay on the first results page



6 Conclusions

Performance

- Efficiency and effectiveness of new system higher than old system
- Very high user satisfaction (4.1 / 5)

Participant groups

- External group worse using old system than internal group
- Both groups performed better with new system
- No more learning curve for external users with new system!

Try it out yourself!

http://ls0135.uvt.nl/~ruud/search_html.php

Future Work

Possible avenues for future work

- Incorporating additional data
 - Project proposals
 - Press releases
- Exploring user interface issues
- Incorporating contextual data (Hofmann et al. (2008))
 - Media experience
 - Organizational hierarchy

References

References

- K. Balog, T. Bogers, L. Azzopardi, M. de Rijke, and A. van den Bosch (2007). *Broad Expertise Retrieval in Sparse Data Environments*. In Proc. of SIGIR '07, pp. 551-558.
- K. Hofmann, K. Balog, T. Bogers, and M. de Rijke (2008). *Integrating Contextual Factors into Topic-Centric Retrieval Methods for Finding Similar Experts*. In Proc. of the SIGIR 2008 ICHER workshop, pp. 29-36

Further information

Please contact A.M.Bogers@uvt.nl. More information on this and related projects can be obtained at <http://ilk.uvt.nl/>. The UvT Expert Collection can be found at <http://ilk.uvt.nl/uvt-expert-collection/>. An online PDF version of this poster can be found at <http://ilk.uvt.nl/~toine/publications/ecir2009-poster.pdf>.