

Relevance and Authority: Development of Information Retrieval and Ranking Algorithms for Social Search Tools

Keywords: Information Retrieval, Human-Computer Interaction, Social Search, Collaborative Computing, User Modeling, Natural Language, Content Analysis and Indexing, Data Mining

Introduction: The proportion of professionals tasked primarily with collecting, handling, and distributing information is growing at an unprecedented pace. The proliferation of information sources on the web poses two problems for knowledge workers and casual searchers alike: Among all possible search results and information artifacts, which are the most relevant and authoritative? Domain experts make skillful judgments of relevance and authority, but those of us without this expertise are left overwhelmed and overburdened by even simple search tasks. In the current state of the art, information seeking is approached as a solitary affair, yet we know that domain experts exist in individual's peer and extended social networks.

Proposed Study and Hypothesis: The present work intends to study domain experts, both with respect to (1) qualitative study of the evaluation criteria used by these experts to identify high-quality, relevant sources and (2) data mining and modeling to identify domain experts within a given social network. These findings will then be used to prototype social search tools which use relevance and credibility rankings to order search results and also highlight connections in the searcher's network who may provide additional help.

Background: Research has already demonstrated that an over-abundance of information sources can overload workers' cognition and threaten their productivity [1]. Major corporations, such as Microsoft, Google, IBM, and Intel have begun to address this problem by forming the Information Overload Research Group (see iorgforum.org). Certain users who are experts in the domain over which they are searching develop techniques for culling relevant information sources from the rest, and a variety of research studies, ranging from focused search tasks with individual users [2] to large-scale, log-based studies [3] have demonstrated that these techniques translate into better performance when searching for information on the web. An example of one of these strategies is the use of more domain-specific technical words in queries [3].

Social technologies represent a way for domain novices to overcome this gap by giving them access to the cognitive and informational benefits provided by their social connections. Research literature has shown that social interactions during search activities can play an important role in search performance [4,5]. For instance, social interaction can address the 'vocabulary problem' [6] mentioned regarding knowledge of domain-specific jargon. In my own research, I have observed the informational and cognitive benefits of social interaction during online information seeking [7]. We found that subjects who had difficulty navigating in a novel domain gained significant benefits from finding and reaching out to experts in their social networks; the ability to locate experts, however, varied significantly, and with better social tools, it is possible that all participants could have learned and performed better.

Objectives and Methodology: This research will be composed of three parts: a qualitative study, data mining and modeling, and construction and evaluation of a prototype search system. The first (qualitative) study will examine the evaluation strategies employed by experts to identify relevant, credible sources of information. This will involve direct observation of search

behaviors in a laboratory setting using think-aloud/teach-aloud protocols and semi-structured interviews. In the second component, I will use data mined from public social tools such as blogs, micro-blogs, and social networks and employ language modeling techniques such as LDA to characterize topical interests in order to computationally establish areas of domain expertise. Finally, I will use results from these studies to direct the development of a search system that (1) re-ranks results using algorithms based on relevance and authority metrics and (2) uses relevance and authority measures to suggest individuals within one's social networks who may be domain experts. These prototypes will be tested with users in order to gain insight into its strengths and weaknesses and then developed into a robust, scalable system for wide deployment, following an established pattern of human-computer interaction research.

Anticipated Results: This is a key time for research into algorithms for identifying relevant, credible sources of information in an individual's social network. Widespread adoption of social tools has led to an explosion of information being produced and transmitted via networks, and effective technologies will be required to help people filter signal from noise. This same rapid proliferation of public data also makes the proposed data mining extremely feasible. In addition, the benefits provided specifically by social interactions with domain experts when searching for information has already been demonstrated, and tools facilitating this interaction would greatly support learning and searching tasks on the internet. Current tools for social search rely primarily on keyword matching and chronological ranking, making them unsuitable for most social search tasks; the research proposed will provide a novel framework for finding and presenting results.

As mentioned in my personal statement, access to high-quality information has the potential to transform lives; as I saw in the pediatric clinic, providing new mothers with information about potential neurological and developmental disorders allows them to observe their child's behavior in order to determine whether further intervention is needed. By pursuing the goal of eventually building a publicly accessible search tool, I believe that this research has the potential to benefit society as a whole, from new mothers to students to knowledge workers.

References:

1. Whittaker, S. and Sidner, C. (1996). Email Overload: Exploring Personal Information Management of Email. *Proc. CHI '96*, 276-283.
2. Duggan, G.B. and Payne, S.J. (2008). Knowledge in the Head and on the Web: Using Topic Expertise to Aid Search. *Proc. CHI '08* (check).
3. White, R., Dumais, S., & Teevan, J. (2009). Characterizing the Influence of Domain Expertise on Web Search Behavior. *Proc. WSDM '09* (check)
4. Morris, M.R. (2008). A survey of collaborative web search practices. *Proc. CHI '08*, 1657-1660.
5. Evans, B.M. and Chi, E. H. (2008). Towards a model of understanding social search. *Proc. CHI '08*, 485-494.
6. Furnas, G.W., Landauer, T.K., Gomez, L.S., & Dumais, S.T. (1987) . The vocabulary problem in human-system communication. *Comm. ACM*, 30(11): 964-971.
7. Evans, B.M., Kairam, S., & Pirolli, P. (2009). Do Your Friends Make You Smarter?: An Analysis of Social Strategies in Online Information Seeking. To Appear in *Information Processing and Management*.