2009

Searching for Authoritative Documents in Knowledge-Base Communities

G. Alan Wang
Virginia Tech University, alanwang@vt.edu

Jian Jiao
Virginia Tech University, jjiao@vt.edu

Weiguo Fan
Virginia Tech University, wfan@vt.edu

SEARCHING FOR AUTHORITATIVE DOCUMENTS IN KNOWLEDGE-BASED COMMUNITIES

Research-in-Progress

G. Alan Wang
Virginia Tech
Blacksburg, VA
alanwang@vt.edu

Jian Jiao
Virginia Tech
Blacksburg, VA
jjiao@vt.edu

Weiguo Fan
Virginia Tech
Blacksburg, VA
wfan@vt.edu

Abstract

Knowledge-based communities are popular Web-based tools that allow members to share and seek knowledge globally. However, research on how to search effectively within such knowledge repositories is scant. In this paper we study the problem of finding authoritative documents for user queries within a knowledge-based community. Unlike prior research on the ranking function design which considers only content or hyperlink information, we leverage the social network information embedded in the rich social media, in addition to content, to design novel ranking strategies. Using the Knowledge Adoption Model as the guiding theoretical framework, we design features that gauge the two major factors affecting users' knowledge adoption decisions: argument quality (AQ) and source credibility (SC). We design two ranking strategies that blend these two sources of evidence with the content-based relevance judgment. A preliminary study using a real world knowledge-based community showed that both AQ and SC features improved search effectiveness.

Keywords: Knowledge-based communities, knowledge adoption, information retrieval, social network analysis

Introduction

Online communities is an emerging discipline because more and more people are getting online with Internet usage doubling every fifty-two days (Preece 2000). Hinds and Lee (2008) classified online communities into different types based on the needs that members expect to be satisfied. In this research we particularly investigate the knowledge-based communities where activities of learning, information dissemination, knowledge seeking, and opinion expression are supported. As a knowledge-based community evolves over time, a knowledge repository grows through its members' participations and has become a valuable asset for the community and its users. Due to its voluntary nature, the knowledge base is not as structured as those professionally maintained such as digital libraries. It is comprised of a large collection of documents authored by many different people. The knowledge encoded in each document is not verified and up to the knowledge seeker to judge its relevance and validity. It is an open research question on how to utilize the knowledge repositories in knowledge-based communities and better serve the needs of people seeking knowledge from the communities.

In an organizational context, Ackerman (1998) envisioned two ways of utilizing a knowledge base: making recorded knowledge retrievable; and making individuals with knowledge accessible. There are emerging studies in the context of online communities that mainly focus on the latter issue of utilizing a knowledge base, i.e., locating the experts who have the expertise to answer a question (Campbell et al. 2003; Dom et al. 2003; Kautz et al. 1997; Krulwich and Burkey 1996; Yang and Dia 2008; Zhang et al. 2007). However, there is little research effort to better retrieve knowledge with regard to a knowledge-seeking task. In practice, most online communities still rely on keyword-based search techniques. Generally speaking, keyword-based search determines the relevance of a document with regard to a user's query based on the number of occurrences that the query keywords appear in the document as well as other heuristics, such as query word rarity in the entire collection. However, for people who seek useful knowledge for a particular task, simple keyword matching may not offer the best solution. The knowledge adoption model shows that one's willingness to adopt knowledge is determined by argument quality that includes information relevance, and source credibility such as the perceived expertise level and trustworthiness of the knowledge source (Sussman and Siegal 2003).

With the limitations of current keyword-based search in mind, this research takes the knowledge adoption model into the community search engine design process. We propose a new search method that helps users of knowledge-based communities effectively locate knowledge they are seeking. The retrieved knowledge is not only highly relevant based on keyword matching, but has a high likelihood to be adopted by the knowledge seeker. We follow the guidelines of the design science research (Hevner et al. 2004) in the design and evaluation of the new search method for knowledge-based communities. We propose two alternative ranking strategies based on different assumptions when combining argument quality and source credibility factors during the knowledge retrieval/seeking process. The goal of the evaluation is to search for a context where the proposed method achieves an optimal performance.

Related Work

In this section we review various issues related to knowledge seeking in knowledge-based communities. These include the unique characteristics of knowledge-based communities, current search techniques in online communities, and the knowledge adoption model.

Knowledge-Based Communities

A knowledge-based community is defined as an online community that primarily supports knowledge seeking and learning activities based on common interests (Hinds and Lee 2008; Lin et al. 2007). The knowledge base of a knowledge-based community is represented by a collection of posts made by its members over a long period of time. The posts are organized into threads, each of which focuses on a particular topic defined by the initial post. Each thread belongs to a higher-level topic category called a forum. Examples of knowledge-based communities include VBCity.com and GameDev.net. VBCity is an 8-year-old online community where over 230,000 VB and .NET developers share programming knowledge and news. GameDev is a 10-year-old community for game developers with over 350,000 participants from around the world.

Knowledge Seeking in Knowledge-Based Communities

Most knowledge-based communities provide a search function that users can use to seek knowledge pertinent to their interests. Ackerman (1998) presented a tool called Answer Garden for searching an organizational knowledge base. Similar to an expert system, the tool "travels down a set of diagnostic questions that lead the user to the information sought" (Ackerman 1998). However, most knowledge-based communities do not have the resources to construct and maintain the structured question set. Instead, they often implement their search functions using traditional information retrieval (IR) techniques. For example, VBCity employs a keyword-based search function. Keyword-based search considers a document relevant to a query if the document contains the search keywords. GameDev's search function employs the Google Site Search® technology that is developed based on Google's proprietary PageRank algorithm (Brin and Page 1998). PageRank considers both keyword-based relevancy and document authority indicated by the number of other relevant pages that refers to it (i.e., hyperlinks). However, the use of hyperlinks in online communities is not as popular as those Web sites professionally developed and maintained. Many documents in knowledge-based communities do not contain hyperlinks. The retrieval of those documents is again determined solely by keyword-based relevance matching.

Although keyword-based relevance has been used by various search engines for many years, it may not exactly represent the true relevance that the searchers would expect. Bade (2007) found that there was a discrepancy between the artificial relevance ranking by statistical methods and the true relevance judgments as determined by the searcher. He further suggested that the design of search techniques would benefit from incorporating relevance indications from the user's perspective.

The Knowledge Adoption Model

Users of knowledge-based communities often come to seek knowledge that can be applied to their own cases. Therefore, we consider the search behavior in a knowledge-based community as a knowledge adoption process. A knowledge seeker is interested in not only information relevant to a search query, but information that the seeker is able to comprehend and has confidence in the credibility and authority of the source. Sussman *et al*. (2003) proposed the Knowledge Adoption Model (KAM) based on the Elaboration Likelihood Model (ELM) (Petty and Cacioppo 1986). They found that a person's intention to adopt knowledge was determined by two factors: the argument quality of the received message and source credibility of the information source. Argument quality refers to the quality of the information content such as relevance, accuracy, and timeliness, while source credibility indicates the credibility and authority of the information source reflecting nothing about the information content. According to the ELM, argument quality is a critical determinant of information influence when the message recipient is able to comprehend the message well. When the recipient is either unable or unwilling to process the message, indications of source credibility become a more critical role in the information influence process. However, source credibility alone cannot be used to judge the message relevancy. Message argument quality and source credibility have to work collectively to influence the knowledge seeker's adoption decision. The mechanisms underlying the interactions of the two factors, however, appear to be complex (Sussman and Siegal 2003).

Searching for Authoritative Documents in Knowledge-Based Communities

Kleinberg (1999) proposes to search for authoritative documents in a hyperlinked environment. The intention is to effectively locate information that is not only relevant to a query but corresponds to human notions of quality. The notion of authority means standing, impact, importance, and influence (Kleinberg 1999). Documents in knowledge-based communities are mostly organized in a hierarchical structure. There are very few hyperlinks that link between documents (i.e., threads). Therefore, the link-based method for searching authoritative documents that Kleinberg proposes does not apply directly in this context. However, we can leverage on the social network built upon the post-reply activities. We consider an authoritative document in a knowledge-based community as a thread that is relevant to a particular topic and participated by influential members in its social network.

A number of studies have leveraged the social network within an organization or community to locate experts in the community with certain expertise (Foner 1997; Kautz et al. 1997; Zhang et al. 2007). However, there are very few research efforts on locating authoritative knowledge that can be directly applied. Jeon *et al*. (2006) proposed a new technique that is potentially useful evaluating the argument quality of a document. However, it does not provide a means to assess the source credibility of the document. In the context of question-answering, Bian *et al*. (2008)

combined both content-based features and user interactions when determining the rank of retrieved answers. However, the design of their features is ad-hoc. Many of the features are specifically designed for Yahoo! Answer, which is their test bed.

Research Questions

Based on the previous discussion, we pose the following research questions in the current study: (1) Given the large amount of social interaction data among members and the large content archive of user postings, how can we design a better search method that serves user needs in seeking useful knowledge? (2) What is the mechanism underlying the interaction of user perceived argument quality and source credibility in the knowledge adoption process?

Research Design

In this section we propose a new search method for knowledge seeking in knowledge-based communities based on the KAM. The proposed method aims to evaluate the usefulness of retrieved documents in terms of two important determinants that affect a user's knowledge adoption intention: argument quality and source credibility. Figure 1 presents an overview of our proposed search method. The keyword-based search function can be accomplished by any existing keyword-based search engine. In the rest of the section we focus our discussion on the computational features of the two determinants. Two ranking strategies are also proposed when combining the features to determine the relevance of a document.

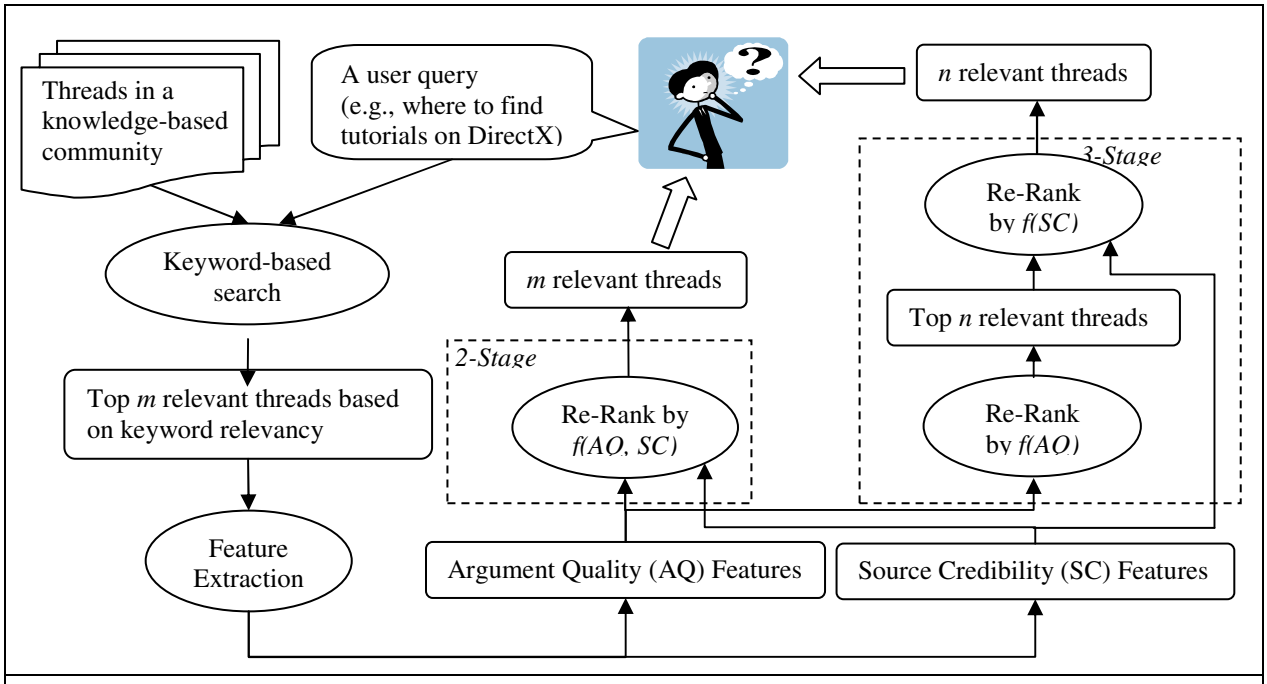

Figure 1. An Overview of the Proposed Search Method

Argument Quality Feature Extraction

Existing IS literature evaluates argument quality in terms of information content, accuracy, and format (Al-Hakim 2007; Bailey and Pearson 1983; Rai et al. 2002; Strong et al. 1997; Sussman and Siegal 2003). However, those measures were mainly developed for conducting surveys. Many of them require subjective judgments from human users and cannot be automatically computed. After examining those measures, we compiled the measures into four dimensions: relevance, amount of information, timeliness, and completeness. We then identified features that to some extent reflect those dimensions and can be computed objectively and automatically. Previous information quality literatures consider timeliness as information recency. However, the interpretation cannot be directly applied

to knowledge-based communities. Information in a replying post needs to be understood in the context of the original post. The further away a reply is made to the original post, the more likely the topic has drifted away from the original context (Herring 1999). Therefore, we consider the timeliness of an online discussion thread as having prompt discussions around the time when a thread is started. Those threads that elicit adequate discussions within a short period of time immediately after the initial post are considered to have higher timeliness than others. Table 1 summarizes the features that we propose to measure the four dimensions of argument quality in knowledge-based communities.

Source Credibility Feature Extraction

Previous research studied source credibility both at individual and organizational levels (Mowen et al. 1987; Newell and Goldsmith 2001; Weiner and Mowen 1986). Most of the measures aim to evaluate two dimensions: the perceived expertise and trustworthiness of an information source. In knowledge-based communities, perceived expertise refers to the knowledge seeker's perception of a post author's reputation as an expert and ability to effectively help. Perceived trustworthiness can be assessed by the post author's role in the social network of the community (Uternark 2002). Table 2 summarizes our proposed computational features that measure the source credibility of a post author in a knowledge-based community. The source credibility of a thread is calculated as the average over all the authors who posted in the thread. The measures of perceived trustworthiness are calculated based on the centrality measures in social network analysis. The social network is constructed based on the post-rely activities in a knowledge-based community. Community members are represented as the network nodes. A link between two members indicates that one member has replied to a thread initiated by the other member. Each link is directional in the sense that a member who receives a reply gets an incoming link.

Table 1. Argument Quality (AQ) Features

Dimensions	Features	Description
Relevance	TF-IDF	A measure that indicates a document's relevance to a search query
Amount of Information	Thread Length	Number of non-stop words
	Number of posts	Number of posts in a thread
Timeliness	Posting frequency	Number of posts per hour in a thread
	Thread duration	Number of hours elapsed between the first post and the last
Diversity	Distinct opinions	Number of unique members in a thread

Table 2. Source Credibility (SC) Features

Dimensions	Features	Description
Perceived Expertise	Membership Length	Number of days elapsed between first and last posting dates
	Productivity	Number of posts per year
	Knowledge Broadness	Number of forums participated by the member
Perceived Trustworthiness	Closeness	Communication distance of the member from all other members
	Betweenness	Importance in forming a coherent communication network
	Out-degree	Number of people being replied by the member (i.e., generosity)
	In-degree	Number of people who helped the member (i.e., popularity)
	Clustering coefficient	How well connected the neighbors of the member are (i.e., brokerage)

Multistage Ranking Methods

Once the features are available, the next question is how to use the features to improve the search results of a user's query. As we mentioned earlier, traditional IR techniques rank searching results primarily based on the query-

document relevance. Our goal is to improve the ranking of retrieved documents so that those ranked on top are not only relevant but will most likely be considered useful by the knowledge seeker. Although all the features proposed in Tables 1 and 2 may affect the ranking of a document, only the relevance measure has a direct effect while others surely provide noisier sources of information. Therefore, they cannot be treated equally when contributing to the document ranking. To compensate the problem, we adopt a multistage ranking strategy similar to the one used in (Kurland and Lee 2005; Wu et al. 2006). The query-document relevance is only used to obtain an initial result set of top *m* documents ranked by the relevance score (i.e., TF-IDF), which is referred to as Stage one results. Other non-relevance-related AQ and SC features will be used to re-rank the documents in the initial set. As it is unknown how AQ and SC interact during a knowledge adoption process, we propose two alternative re-ranking methods based on different assumptions.

Two-Stage Ranking

The two-stage ranking assumes that non-relevance AQ and SC features contribute collectively to the final document ranking. Based on the assumption, after the Stage one ranking using TF-IDF, the score function that we use to re-rank the *m* documents in the initial set in Stage two is:

$$f(AQ, SC) = \sum_{i=1}^{x} w1_i \, AQ_i + \sum_{j=1}^{y} w2_j \, SC_j \,, \tag{1}$$

where AQ_i and SC_i are normalized between 0 and 1, $w1_i$ and $w2_j$ are importance weights. When $w1_i$=0 ($1 \le i \le x$), the SC features alone are used to re-rank the *m* documents obtained in Stage one. The AQ features re-rank the *m* documents alone when $w2_j$=0 ($1 \le j \le y$).

Three-Stage Ranking

The three-stage ranking assumes that non-relevance AQ and SC features may have different importance when contributing to the final document ranking. Research suggests that source credibility has little impact on attitude change towards a message when the message recipient is familiar with the topic (Heesacker et al. 1983). That way the recipient will focus on analyzing the message argument rather than evaluating the expertness and trustworthiness of the information source. Therefore, we argue that SC features may be even noisier than non-relevance-related AQ features. Thus we propose to first re-rank the *m* documents in the initial set from Stage one using non-relevance-related AQ features only (Stage two). SC features will then be used to re-rank the top *n* documents with high AQ ratings (Stage three). The ranking functions for the three-stage ranking are listed as the following:

$$f(AQ) = \sum_{i=1}^{x} w1_i \, AQ_i \,, \;\; f(SC) = \sum_{j=1}^{y} w2_j \, SC_j \,. \tag{2}$$

Preliminary Study

We have conducted a preliminary study that evaluated the effectiveness of the AQ and SC features using data gathered from VBCity.com. VBCity has a large knowledge base of more than 0.67 million posts contributed by more than 235,900 registered members over the past 8 years. There are 36 forums at VBCity specifically focusing on the technical aspects of VB programming.

We manually constructed 11 queries by analyzing thread titles in the 36 forums. After indexing the thread titles, stop words were first removed and stemming was used to convert related words into the same stem. We selected a set of frequently used keywords whose document frequency ratings were greater than a threshold. We then calculated the co-occurrence of any two frequent keywords that appeared together in the thread titles. The query candidates were chosen from the list of keyword pairs ranked high by the co-occurrence. In this pilot study we mainly focused on search tasks intended to solve specific problems. The 11 VB questions listed in Table 3 were used as our search queries when evaluating our proposed search methods.

Similar to the experimental design in (Chakrabarti et al. 1998), we used human judgment to assess the quality of each retrieved authoritative document. We asked 3 experts, who were Ph.D. students in computer science, to evaluate the top 50 search results of the 11 queries using a keyword-based search algorithm, the Okapi BM25 (Robertson et al. 1995). Okapi BM25 is designed based on the 2-Poisson model and has consistently performed very well in TREC competitions (Hawking and Craswell 2001). It ranks documents based on query-document relevance similar to a TF-IDF score. The experts labeled each search result as "helpful" or "not helpful" in terms of their utility in learning about the search query. The labeled document set became the gold standard for the evaluation of our pilot study.

Table 3. Search Queries for Preliminary Study	
Q1: how to delete a file that is already open	Q7: how to get date from DateTime in SQL Server
Q2: how to add WinXP themes to VB application	Q8: how to get pixel color
Q3: how to change font size while resizing the window	Q9: how to search the text in a .pdf file
Q4: how to combine data from multiple .csv files to one excel file	Q10: how to validate a phone number in VB
Q5: how to convert XML to database	Q11: how to view a single crystal report with multiple subreports
Q6: how to get CPU ID	

Performance Measures

We used three IR measures to evaluate the document ranking with regard to a user query: Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), and Normalized Discounted Cumulative Gain (NDCG). Those measures are standard performance measures widely in information retrieval and question answering evaluations (Chen et al. 1998; Fan et al. 2005a; Fan et al. 2005b; Harman 1996; Radev et al. 2005) and provide a balanced view about the search performance.

Given a query q, the corresponding ranked document set $\{d_i\}$, and a relevance function $rel(d_i)$ that takes on 1 and 0 as values representing d_i being relevant or irrelevant, we can present the three measures as the following. The mean values were computed as the average over all queries.

$$AP@10 = \frac{\sum_{i=1}^{10} P@i * rel(d_i)}{\sum_{i=1}^{n} rel(d_i)}, \quad RR@10 = \begin{cases} \frac{1}{i}, \text{ if } \exists i < 10 : rel(d_i) = 1 \wedge \forall j < i : rel(d_j) = 0 \\ 0 \end{cases},$$

$$NDCG@10 = \frac{\sum_{i=1}^{2} rel(d_i) + \sum_{i=3}^{10} \frac{rel(d_i)}{\log_2(i)}}{\sum_{j=1}^{2} rel^{ideal}(d_j) + \sum_{j=3}^{10} \frac{rel^{ideal}(d_j)}{\log_2(j)}}$$

Evaluation Procedure

To evaluate the effectiveness of the AQ and SC features, we applied the two-stage ranking method to get the ranked results for each search query. At Stage one, we used Okapi BM25 to select the top m search results for a search query. At Stage two, we used either AQ or SC features to re-rank the search results obtained in Stage one. We varied the parameter m in the range of 10-50 incrementing 10 each time. We compared the search performance with the Okapi BM25 baseline method. We will evaluate the re-ranking results using both AQ and SC features at future studies.

In this preliminary study we assigned equal weights to those features. However, it does not imply that all features should be considered equally important for the ranking. We plan to examine the feature importance in the near future. We believe that a better weighing scheme would further improve our results if this preliminary study is promising.

Preliminary Results

Figure 2 shows the effectiveness of using the AQ and SC features to improve search performance in comparison to the baseline method, namely the Okapi BM25. The charts show the following findings. First, the AQ features improved the search results returned by the Okapi BM25 across the board. The improvements in the three measures were more prominent when *m* (i.e., the number of top Okapi search results) was small. The AQ features are reliable in boosting the search performance, although the inclusion of less relevant documents may reduce the magnitude of the improvement. Secondly, AQ features outperformed SC features in most cases. SC features improved the MAP and MRR measures when *m* was equal to 10. However, both measures dropped significantly as *m* increased. The observation shows that SC features may be noisy and not useful when re-ranking less relevant documents. The result prompts us that SC features should be used with care when determining the ranking of retrieved documents. We expect the three-stage ranking strategy to outperform the two-stage ranking when SC features are used to re-rank those documents with high AQ.

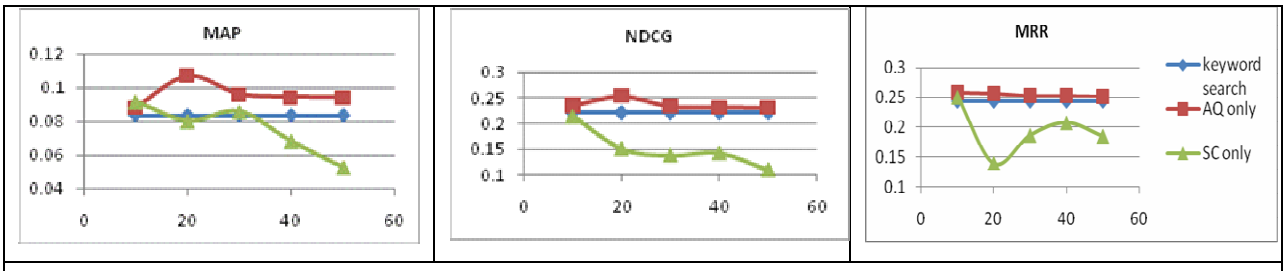

Figure 2. The Effectiveness of AQ and SC Features in Searching Authoritative Documents

Discussions and Conclusions

In this paper, we study the problem of finding authoritative documents for user queries within a knowledge-based community. Unlike many of the prior research on ranking function design which considers only content or hyperlink information in relevance matching between user queries and documents, we leverage the social network information embedded in the rich social interaction data, in addition to the content information, to design novel ranking strategies that find the most useful knowledge for end users. In particular, using the Knowledge Adoption Model as the guiding theoretical framework, we design features that can be automatically extracted and used to gauge the two major influencing factors affecting user knowledge adoption decision: Argument Quality and Source Credibility. Our preliminary experiment using a real world knowledge-based community showed promising results. Both AQ and SC features that we identified helped improve search performance. We propose two new ranking strategies that seek to blend these two sources of evidence with the content-based relevance judgment. We plan to conduct more comprehensive studies to evaluate the two ranking strategies and their variations.

There are several directions for future research. One direction is to design more features along the two knowledge adoption dimensions to capture various heuristics that users may use for knowledge adoption. Another direction is to design algorithms and strategies that can find the optimal ways of combining these two sources of evidence. We are currently working on a feature weighting scheme using the Genetic Algorithm so that an optimal ranking solution can be achieved. In addition, we can use the weights to assess the influence of individual and combinations of features on ranking quality. The third direction is to test our proposed ranking strategies on different knowledge-based communities to validate the generalizability of our approach on different collections. The fourth direction is to design ranking strategies that can search across different communities for any given user queries. Since different knowledge-based communities may exhibit different network of relationships and contents, finding a common ranking strategy that can work well across the heterogeneous collections will be a big challenge.

References

Ackerman, M.S. 1998. "Augmenting Organizational Memory: A Field Study of Answer Garden," *ACM Transactions on Information Systems (TOIS)* (16:3), pp 203-224.

Al-Hakim, A. 2007. "E-Innovation: Information Quality Perspective," *Fifth International Symposium on Management of Technology*, Hangzhou, China.

Bade, D. 2007. "Relevance Ranking Is Not Relevance Ranking or, When the User Is Not the User, the Search Results Are Not Search Results," *Online Information Review* (31:6), pp 831-844.

Bailey, J.E., and Pearson, S.W. 1983. "Development of a Tool for Measuring and Analyzing Computer User Satisfaction," *Management Science* (29:5), pp 530-545.

Bian, J., Agichtein, E., Liu, Y., and Zha, H. 2008. "Finding the Right Facts in the Crowd: Factoid Question Answering over Social Media," *Proceedings of the International World Wide Web Conference (WWW)*, Beijing, China.

Brin, S., and Page, L. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks* (30), pp 107-117.

Campbell, C.S., Maglio, P.P., Cozzi, A., and Dom, B. 2003. "Expertise Identification Using Email Communications," *Proceedings of the 12th International Conference on Information and Knowledge Management*, New Orleans, LA.

Chakrabarti, S., Dom, B., Gibson, D., Kumar, S., Raghavan, P., Rajagopalan, S., and Tomkins, A. 1998. "Experiments in Topic Distillation," Melbourne, Australia, pp. 13-21.

Chen, H., Shankaranarayanan, G., She, L., and Lyer, A. 1998. "A Machine Learning Approach to Inductive Query by Examples: An Experiment Using Relevance Feedback, Id3, Genetic Algorithms, and Simulated Annealing," *Journal of the American Society for Information Science* (49), pp 693-705.

Dom, B., Eiron, I., Cozzi, A., and Zhang, Y. 2003. "Graph-Based Ranking Algorithms for E-Mail Expertise Analysis," *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, San Diego, CA.

Fan, W., Gordon, M., and Pathak, P. 2005a. "On Linear Mixture of Experts Approach to Information Retrieval," *Decision Support Systems* (42:2), pp 975-987.

Fan, W., Gordon, M.D., and Pathak, P. 2005b. "Genetic Programming-Based Discovery of Ranking Functions for Effective Web Search," *Journal of Management Information Systems* (21:4), Spr, pp 37-56.

Foner, L.N. 1997. "Yenta: A Multi-Agent, Referral-Based Matchmaking System," *Proceedings of the First International Conference on Autonomous Agents*, Marina Del Rey, CA.

Harman, D.K. 1996. "Overview of the Fourth Text Retrieval Conference (Trec-4)," in: *The Proceedings of the Fourth Text Retrieval Conference,* D.K. Harman (ed.). pp. 1-24.

Hawking, D., and Craswell, N. 2001. "Overview of the Trec-2001 Web Track," in: *NIST special publication*. pp. 61-67.

Heesacker, M., Petty, R.E., and Cacioppo, J.T. 1983. "Field Dependence and Attitude Change: Source Credibility Can Alter Persuasion by Affecting Message Relevant Thinking," *Journal of Personality* (51:4), pp 653-666.

Herring, S. 1999. "Interactional Coherence in Cmc," *Journal of computer-mediated communication* (4:4), pp 1-24.

Hevner, A.R., March, S.T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp 75-105.

Hinds, D., and Lee, R.M. 2008. "Social Network Structure as a Critical Success Condition for Virtual Communities," *Proceedings of the 41st Hawaii International Conference on System Sciences (HICSS)*, Big Island, Hawaii.

Jeon, J., Croft, W., Lee, J., and Park, S. 2006. "A Framework to Predict the Quality of Answers with Non-Textual Features," *Proceedings of SIGIR*.

Kautz, H., Selman, B., and Shah, M. 1997. "Referral Web: Combining Social Networks and Collaborative Filtering," *Communications of the ACM* (40:3), pp 63-65.

Kleinberg, J. 1999. "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM* (46:5), pp 604-632.

Krulwich, B., and Burkey, C. 1996. "Contactfinder Agent: Answering Bulletin Board Questions with Referrals," *Proceedings of the 13th National Conference on Artificial Intelligence*, Portland, OR.

Kurland, O., and Lee, L. 2005. "Pagerank without Hyperlinks: Structural Re-Ranking Using Links Induced by Language Models," *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brazil, pp. 306-313.

Lin, H., Fan, W., Wallace, L., and Zhang, Z. 2007. "An Empirical Study of Knowledge Community Success," *Proceedings of the 40th Hawaii International Conference on System Sciences (HICSS)*, Big Islands, Hawaii.

Mowen, J.C., Wiener, J.L., and Joag, S. 1987. "An Information Integration Analysis of How Trust and Expertise Combine to Influence Source Credibility and Persuation," *Advances in Consumer Research* (14:564-567).

Newell, S.J., and Goldsmith, R.E. 2001. "The Development of a Scale to Measure Perceived Corporate Credibility," *Journal of Business Research* (52), pp 235-247.

Petty, R.E., and Cacioppo, J.T. 1986. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag.

Preece, J. 2000. *Thriving Online Communities: Usability and Sociability*. Chichester, England: John Wiley & Sons.

Radev, D., Fan, W., Qi, H., Wu, H., and Grewal, A. 2005. "Probabilistic Question Answering on the Web," *Journal of the American Society for Information Science and Technology* (56:6), Apr, pp 571-583.

Rai, A., Lang, S., and Welker, R. 2002. "Assessing the Validity of Is Success Models: An Empirical Test and Theoretical Analysis," *Information Systems Research* (13:1), pp 50-69.

Robertson, S., Walker, S., Jones, S., Hancock-beaulieu, M., and Gatford, M. 1995. "Okapi at Trec--3," *Proceedings of the Third Text TEtrieval Conference (TREC-4)*.

Strong, D.M., Yang, W.L., and Wang, R.Y. 1997. "Data Quality in Context," *Communications of the ACM* (40:5).

Sussman, S.W., and Siegal, W.S. 2003. "Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption," *Information Systems Research* (14:1), pp 47-65.

Uternark, T.L. 2002. "Cohort Differences in Perceptions of Helpful Counselor Characteristics," in: *Department of Psychology*. University of North Texas.

Weiner, J.L., and Mowen, J.C. 1986. "Source Credibility: On the Independent Effects of Trust and Expertise," *Advances in Consumer Research* (13), pp 306-310.

Wu, C.-H., Yeh, J.-F., and Lai, Y.-S. 2006. "Semantic Segment Extraction and Matching for Internet Faq Retrieval," *IEEE Transactions on Knowledge and Data Engineering* (18:7), pp 930-940.

Yang, W.-S., and Dia, J.-B. 2008. "Discovering Cohesive Subgroups from Social Networks for Targeted Advertising," *Expert Systems with Applications: An International Journal* (34:3), pp 2029-2038.

Zhang, J., Ackerman, M.S., and Adamic, L. 2007. "Expertise Networks in Online Communities: Structure and Algorithms," *Proceedings of the 16th International World Wide Web Conference (WWW)*, Banff, Canada.