

EPISOSE: An Epistemology-Based Social Search Framework for Exploratory Information Seeking

Yuqing Mao, Haifeng Shen, Chengzheng Sun

► **To cite this version:**

Yuqing Mao, Haifeng Shen, Chengzheng Sun. EPISOSE: An Epistemology-Based Social Search Framework for Exploratory Information Seeking. Peter Forbrig; Fabio Paternó; Annelise Mark Pejtersen. Second IFIP TC 13 Symposium on Human-Computer Interaction (HCIS)/ Held as Part of World Computer Congress (WCC), Sep 2010, Brisbane, Australia. Springer, IFIP Advances in Information and Communication Technology, AICT-332, pp.211-222, 2010, Human-Computer Interaction. <10.1007/978-3-642-15231-3_21>. <hal-01055480>

HAL Id: hal-01055480

<https://hal.inria.fr/hal-01055480>

Submitted on 12 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



EPISOSE: An Epistemology-based Social Search Framework for Exploratory Information Seeking

Yuqing Mao¹ Haifeng Shen² Chengzheng Sun¹

¹ School of Computer Engineering
Nanyang Technological University
Block N4, Nanyang Avenue, Singapore, 639798
{maoy0002, czsun}@ntu.edu.sg

² School of Computer Science,
Engineering and Mathematics
Flinders University, Adelaide, Australia
hfshen@csem.flinders.edu.au

Abstract. Search engines are indispensable for locating information in WWW, but encounter great difficulties in handling exploratory information seeking, where precise keywords are hard to be formulated. A viable solution is to improve efficiency and quality of exploratory search by utilizing the wisdom of crowds (i.e., taking advantage of collective knowledge and efforts from a mass of searchers who share common or relevant search interests/goals). In this paper, we present an epistemology-based social search framework for supporting exploratory information seeking, which makes the best of both search engines' immense power of information collection and pre-processing and human users' knowledge of information filtering and post-processing. To validate the feasibility and effectiveness of the framework, we have designed and implemented a prototype system with the guidance of the framework. Our experimental results show that an epistemology-based social search system outperforms a conventional search engine for most exploratory information seeking tasks.

Keywords: Exploratory Search, Information Seeking, Social Search, Search Epistemology, Collaborative Search.

1 Introduction

Web search has become the best source of information for many people. The following factors are relevant to the success of a web search: 1) to provide information that can be easily accessed (by information providers), 2) to formulate precise keywords that can express search goals (by information consumers), and 3) to rank search results according to their relevance to the search goals (by search engines).

To information providers, some research on search engine optimization (SEO) [10] has been done to achieve easy accessibility of their information. However, the abuse of SEO can cause information overload by search engines.

To search engines, the ultimate goal is to understand the goals of information consumers and to return what they want. Unfortunately, the processing of natural lan-

guages and the deployment of semantic web still has a long way to go. Researchers are currently focusing on improving the intelligence of search engines in order to rank the results pertinent to a search query. For example, personalized web search systems collect users' profiles together with users' queries to measure similarities with techniques such as collaborative filtering for web search [2]. On the other hand, studies have shown that nearly 40% of queries were repeated [11]; a lot of work has been done to re-use previous successful queries. For example, one will see the keyword suggestions from a search engine like Google when she/he types an incomplete keyword in the search box. Queries can also be enhanced and modified for query expansion by the analysis of similar query logs in a search engine [1].

To information consumers, field knowledge and search skills are essential to a successful web search task. As individual users' knowledge and skills are limited, they can complement each other by turning to social search, which utilizes "the wisdom of crowds" [4]. As opposed to algorithm-based searches, human-labor-based searches are collectively called "social search". Representative examples include ChaCha¹, which offers live expert assistance to guide ordinary users in performing search tasks and SearchTogether [8], which provides a platform for a group of people to communicate and share search results in real time. As collaborative tagging has grown in popularity on the web, social media and social annotations have also attracted a lot of attention. Successful examples include question and answer systems like Yahoo! Answers², social annotations like Mahalo³, and websites for discovering expert in social networks like Google Adarvark⁴. Social search has changed a search process from an individual activity to a social one. It has been proved effective in information filtering (e.g., understanding users' search goals) and helpful in information post-processing (e.g., annotating search results).

Although these social search systems are good complements to conventional search engines, they are generally inadequate for supporting effective "exploratory search" [7], where a search process starts with a vague goal rather than precise keywords, and the information consumer is not clear about what search results she/he is expecting. She/he can only evaluate every search result by measuring its relevance to her/his search goal. This is mainly due to the fact that information providers are not actively involved in search processes to help information consumers easily access the information they have provided. According to Heymann et al. [3], although some information providers provide information not currently available from other sources (e.g., answers to a question), they need to make extra efforts to make the information easily searchable by information consumers.

Our solution to supporting effective exploratory search is based on the effective sharing of "search epistemology" – the knowledge acquired by users in information collection, pre-processing, filtering and post-processing – among a social search community. Search epistemologies are aggregated and well-structured information packages derived from successful search processes, such as queries, results, rankings, annotations, comments, and evaluations. In this solution, information consumers are

1 <http://www.chacha.com>

2 <http://answers.yahoo.com>

3 <http://www.mahalo.com>

4 <http://www.vark.com>

also information providers, who are actively involved in a social search process to help get their contributed information easily accessed by others.

In this paper, we present an Epistemology-based Social Search (EPISOSE) framework for designing exploratory information seeking systems. The framework leverages existing algorithmic search engines to collect and pre-process information and human users to construct search epistemologies by filtering and post-processing information, and allows a social search community to effectively share their intimate search epistemologies. The EPISOSE framework can be applied to the design and implementation of a range of social search systems with different strategies and algorithms. To validate the feasibility and effectiveness of the framework, we have designed and implemented a prototype system *Baijia* with the guidance of the framework and conducted a set of experiments to measure the system’s performances in supporting exploratory information seeking.

The rest of the paper is organized as follows. Section 2 describes the architecture and components of the EPISOSE framework. Section 3 presents an implementation of the framework – the *Baijia* prototype system, and Section 4 discusses the setting and results of experiments. Finally, Section 5 concludes the paper with a summary of major contributions and future work.

2 The EPISOSE Framework

Contrary to the linear relationship between a user and a search engine in current web search systems, one of EPISOSE’s distinctive characteristics is a positive feedback cycle: information consumers are also information providers so that their search epistemologies can be contributed to subsequent search processes for the benefit of other consumers. As depicted in Fig. 1, the EPISOSE framework consists of the following major components.

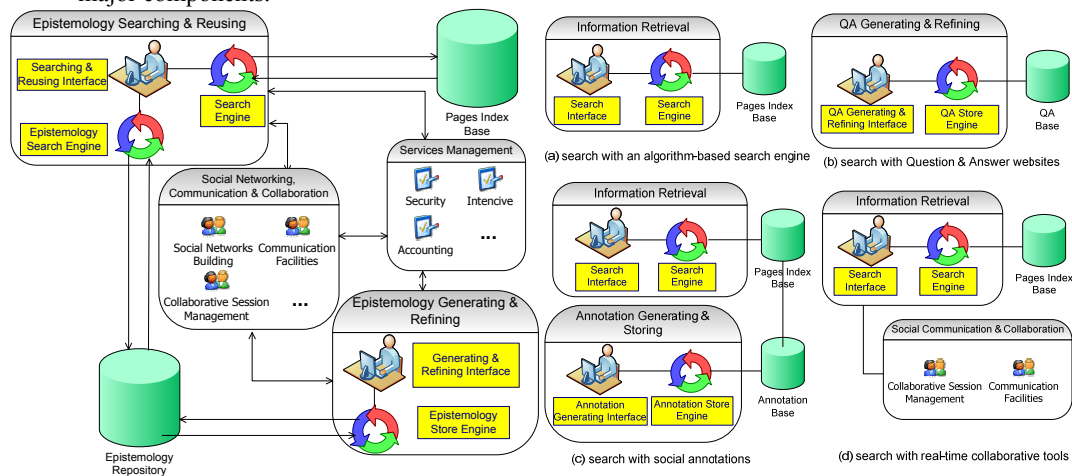


Fig. 1. The EPISOSE Framework

Fig. 2. The A comparison of EPISOSE with other Web search frameworks

Epistemology Searching & Reusing: this component is for a social search community to share and reuse search epistemologies. While a user types a query through the *Searching & Reusing Interface*, the *Epistemology Search Engine* will first search the epistemology repository and return the relevant search epistemologies. These epistemologies were contributed by other users with the same or relevant search interests or goals through the *Generating & Refining Interface* powered by the *Epistemology Store Engine*. If no relevant epistemology is found, the *Search Engine* (e.g., Google) will search the *Pages Index Base* and return relevant pages according to the keywords. Users can generate their own epistemologies from the result pages returned by the search engine or refine others' relevant epistemologies retrieved from the epistemology repository. In this component, knowledge discovery techniques (such as classification or clustering) can be applied to analyze the relevance of the input queries to the stored search epistemologies in the epistemology repository. Because each epistemology is tagged with plenty of additional information by processing raw information returned by a search engine with the user's intimate knowledge and understanding, the *Epistemology Search Engine* has a very high probability to return relevant epistemologies pertinent to a query. Consequently, other users can save their time in repeating the course of collecting and processing raw information through a search engine.

Epistemology Generating & Refining: this component is for users to easily generate new search epistemologies or refine existing ones, and store them into the *Epistemology Repository* through the *Epistemology Generating & Refining Interface* powered by the *Epistemology Store Engine*. A search epistemology includes the subsequence of search keywords, the approbatory results selected by the user, the commentaries on these results added by the user, other useful information about the search topic provided by the user, and evaluation of the search provided by other users.

This component is significantly different from answering questions or adding annotations because stored search epistemologies can be easily retrieved by *Epistemology Searching & Reusing* without relying on search engines, which have difficulties in retrieving most relevant information if precise keywords are hard to be formulated. Since epistemologies are generated during exploratory information seeking processes, the retrieval of epistemologies would be a heuristic search. Users can be quickly led to the final search goal by the epistemologies generated from others' successful searches. Furthermore, as existing epistemologies can be refined by subsequent users, they tend to be more relevant and accurate as they are refined by more users.

Epistemology Repository: this component stores search epistemologies contributed by the users in a social search community. Search epistemologies are information packages regarding specific search goals. They are indexed by the combination of the search goal and relevant keywords for the sake of easy retrieval by the *Epistemology Search Engine*. The *Epistemology Repository* also stores extra information about epistemologies for building social search communities. For example, building a social network needs the information about who contribute relevant epistemologies.

Social Networking, Communication & Collaboration: this component has the following functions: *Social Networks Building* helps users with the same or similar search goals build up social networks to complete search tasks together. While users are doing exploratory information seeking, they would be likely to look for help from others. Users can find people with same hobbies or similar information requirements

from search epistemologies and thence build a social network with them. Furthermore, the EPISOSE framework can adopt effective strategies to search for expertise in social networks, so that exploratory searches can benefit from the building of social networks. When one is not sure about what she/he is looking for, seeking advices from experts is always a good option.

Communication Facilities allow users in a social search community to communicate via tools such as messenger or email. *Collaborative Session Management* allows a group of users to share their search epistemologies in an ongoing search process synchronously or asynchronously. While users are doing exploratory information seeking, they may invite others to work with them together. Users can discuss with the contributors of certain epistemologies in order to better refine them, or invite buddies in their social networks to join their ongoing search processes. *Services Management* provides some common services for making exploratory social search viable, reliable, and sustainable. For example, *Incentive* encourages users to share their epistemologies, *Security* handles issues related to privacy and security in a social search community, and *Accounting* can estimate users' contribution to establish a profit model for social search.

EPISOSE is uniquely distinguished from current web search frameworks listed in Fig. 2(a)-(d). Fig. 2(a) shows the conventional method of web search with a search engine. In these systems, the *Search Engine* is an algorithm-based search engine such as Google. Search knowledge cannot be shared since there is no knowledge generation and storage mechanism. Peer to peer search engines can support each peer to publish its local documents (or local index), but the purpose is to build efficient topic-specific search engines rather than supporting general exploratory search through knowledge sharing. Fig. 2 (b) shows the method of web search with Question & Answer websites, e.g. asking for help in forums or Yahoo! Answers, which is also known as the man-powered search engines. Users can share knowledge through those social web sites, but knowledge sharing is not effective as shared knowledge still has to be retrieved by conventional search engines. Fig. 2(c) shows the method of web search with social annotations, such as Yoogle⁵, which allows people to publish their attitudes toward certain search results through voting or editing. But annotations are not as rich in content as search epistemologies (annotations are only an integral part of epistemologies) and sharing of annotations is not effective as they have also to be retrieved by search engines. Fig. 2(d) shows the method of web search with real-time collaborative tools such as SearchTogether [8], which supports search and sharing of Web pages with others through communication facilities. Searched Web pages cannot be shared to people who are not invited or online when the search process is ongoing. Moreover, it shares only Web pages rather than a package of knowledge about a specific topic.

The primary goal of a social search system for exploratory information seeking is to utilize existing successful searches. Many people have searched for the same topic, for example, wedding planning, but they couldn't re-use previous successful searches because they were not shared or fell into oblivion. Yet it's not true that people are unwilling to share the knowledge gained from their searches. The problem is that they couldn't find convenient and effective ways to share it. As a matter of fact, many

⁵ <http://www.yoogle.net>

people write their knowledge of wedding planning in blogs or discussion forums after effortful web searches. However, only a small group of people in a specific period of time may benefit from such kind of knowledge sharing, while others still need to spend tremendous time repeating the process of searching for the same topic.

The EPISOSE framework addresses this problem as follows. First, it provides a social search platform and interfaces for a wide range of users to share their search epistemologies directly. Second, search epistemology is based not only on a rich set of objective results returned by a search engine, but also on a user’s subjective judgment and intimate knowledge. It is far beyond just a set of search keywords and a list of linear results. Third, searching and sharing are seamlessly integrated in the framework so that users can enrich their knowledge about a search topic and improve their search skills by learning from their peers. It is worth pointing out that an epistemology-based social search system is self-reinforcing in the sense that epistemologies tend to be more relevant and accurate as they are refined by more users.

3 Implementation

To validate the EPISOSE framework, we have designed and implemented a prototype system *Baijia* (see Fig. 3 & 4), which allows a social search community to share their search epistemologies.

3.1 Data Collection

We selected the AOL query logs [9] as the initial dataset for the *Baijia* system. The AOL query logs consist of about 20 million search queries from about 650,000 users. Although the dataset doesn’t contain explicit user’s feedback on the search results, the URL clicking can be regarded as the implicit positive feedback because relative feedback signals generated from the search engine users’ clicking behaviors have been proved to correspond well with explicit judgments [6].

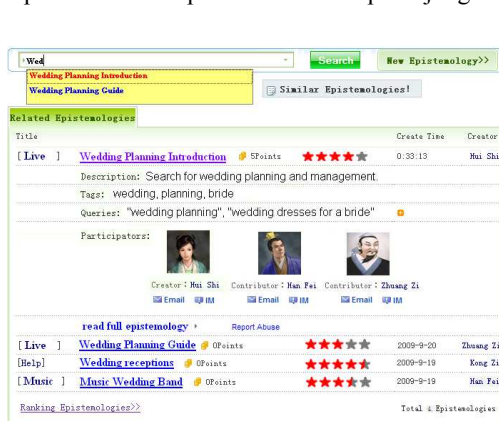


Fig. 3. The *Baijia* system interface

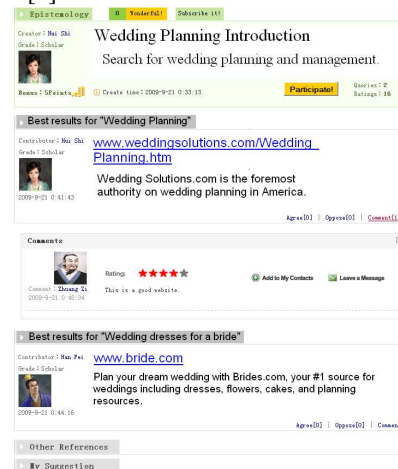


Fig. 4. Epistemology generation and refinement

The dataset includes {AnonID, Query, QueryTime, ItemRank, ClickURL}, where AnonID presents an anonymous user ID number, ClickURL is the URL user clicked and ItemRank is the rank of the clicked item on the listed results. The initial epistemology repository of *Baijia* is built by importing the dataset and the search epistemologies are automatically generated for each user. For some selected topics, search epistemologies are clustered by keywords of the queries, and formed by specified rules, e.g., different weights are given to URLs according to the number of click times for the same queries. With the initial epistemology repository, we can also setup experiments to evaluate the effectiveness of *Baijia* system by contrasting the new search processes with those in the query logs.

3.2 Searching and Sharing

Take a search of wedding planning as an exploratory information seeking example. A user (AnonID=19913) begins with the keyword “wedding planning”. After the user inputs “wedding planning” in the concise search interface (Fig. 3), the system will return a linear list of results through the common API of a search engine (currently the Google AJAX Search API). If previous search epistemologies exist, a re-ranked results list will be returned. The details of these results show who have contributed to the search epistemologies and how do other users evaluate these epistemologies, which will help the user judge their relevance.

After browsing a few pages of the results, the user regards the page from “www.weddingsolutions.com” as the best page about the solutions of wedding planning and drags the item from “search results” to “best results”. The user can also add some personal comments on this result, e.g., “This is a good website about wedding planning”. The interfaces for search and sharing are presented in one page so that users can easily drag and drop items and read and write comments without any popup windows.

Having gained the basic knowledge about wedding planning, the user may continue the search for “wedding receptions” if she/he considers having a wedding reception (AnonID=884092), or “wedding dresses” if she/he wants to get some good suggestions about the dressing in the wedding ceremony (AnonID=5761104). She/he can choose “share epistemology” to publish the epistemology after completing the whole search process.

She/he may entitle the epistemology as “Wedding Planning Introduction” and classify it into a “successful search” or “partially successful search” category. She/he can even supplement the search with her/his own epistemology: “All those preparations are not enough and I think there should be...” (Fig. 4).

When another user also starts exploratory information seeking about wedding planning with the keyword “wedding planning”, she/he will get a list of related epistemologies from others (Fig. 3). The relevant search epistemologies can be collected by filtering according to the similarity to the query. For a user who knows little about wedding planning, what has been searched by others is a good starting point for her/him. Hence, the user can browse the list and dig out the details of those in which she/he is interested, e.g. “Wedding dresses for a bride”, if she/he (e.g. AnonID=16852248) wants to know more about wedding dresses. The user can also get

help from others' epistemologies if she/he is not sure about what she/he is searching for. For example, the user (AnonID=12199341) has only a vague idea about "how can I write the wedding invitations". After browsing the list, she/he finds that many others are looking for "wedding invitation wording", which effectively suggests her/him how to start the search.

Furthermore, a user can learn search skills from others' epistemologies. For example, when a user (AnonID=2200929) wants to search information about a band playing music in the wedding ceremony and she/he uses the keyword "wedding band", she/he would be puzzled since all results returned by a search engine in the first page are all about the wedding rings. If a previous successful search epistemology named "Music Wedding Band" has been generated in the epistemology repository, the user can easily find relevant information from the related epistemologies list. Meanwhile, she/he knows how to formulate keywords and to filter out unwanted ones. Such knowledge will probably help accelerate her/his search process.

The user can continue her/his search on the basis of these searches, add new good results and commentaries or remove expired or faulty information. If a page is dynamic, she/he may add annotations or modify the link with parameters to ensure it will not expire. The user can also evaluate others' epistemologies, e.g., give positive or negative feedback. All search epistemologies will be accumulated in the epistemology repository for further utilization.

3.3 Epistemology Search Engine & Epistemology Repository

The epistemology search engine derives the relevance between search epistemology and a search query. For example, if the "www.weddingsolutions.com" page of wedding has appeared as the best result in both the search epistemologies of "Wedding Planning Guide" and "Traditional Wedding Planning", we are likely to draw the conclusion that it is relevant to the query about "wedding planning". Besides, some methods such as feature extraction and clustering will be employed to accurately discover more relevant search epistemologies to certain search goals.

For conflicting search epistemologies, the reputation and the expertise of users can be used to help make the judgment. One's reputation will increase/decrease when one's published search epistemologies receive positive/negative evaluation by others. This will discourage users from publishing misleading or irresponsible search epistemologies for malicious purposes or undeserved reward points. The epistemologies published by users with better reputation will be assigned heavier weights

Since all search epistemologies are stored in the epistemology repository, its volume grows fast with time. Therefore, it is important to build suitable indices for search epistemologies in order to improve the retrieval efficiency and accuracy. Moreover, with the epistemology repository, we can also support semantic epistemology retrieval by building various ontologies. For example, the "band" would be modeled with the meaning of "ring" for an ontology about the domain of wedding, but with the meaning of "instrumentalists" for another ontology about the domain of music. As such, our system could process users' requests more accurately based on the context of their queries.

3.4 Communication & Personal Information

Baijia allows users to communicate with the system and with each other via email or a web-based instant messenger (Fig. 3). It can support synchronous sharing of search epistemologies among users and allow users to communicate with other instant messaging users through gateways of the jabber server so that users can get instant help from their online buddies or experts during a search process. For example, the bride who is searching for the wedding planning may have difficulties in choosing the dress to wear in the wedding, she can immediately ask her online buddies or parents to generate their search epistemologies for her or modify her rudimentary epistemologies. Users can manage personal information through a control panel. Security settings can also be done here. For privacy-sensitive information, a user can choose to share within a small group, e.g., family members, or not to share the search epistemologies at all. It is helpful in some cases, e.g., colleagues searching for new techniques in order to get a project done. All members can take part in the search task and share epistemologies with each other. Users are motivated to share or evaluate search epistemologies by an incentive mechanism based on accumulated points. Besides, a user can subscribe to an interested search epistemology in order to receive notifications when it is updated (Fig. 4).

4 Experimental Evaluations

We have done some experiments to evaluate to what extent the *Baijia* system can improve exploratory information seeking as compared to a conventional search engine (e.g., AOL). We plan to conduct user evaluations after the system has accumulated adequate epistemologies.

4.1 Experiments Setup

For the sake of simplicity, clicked URLs are regarded as relevant search results pertaining to the corresponding queries in our experiments. It is a reasonable hypothesis since clicked URLs must have attracted users' attentions although they might not necessarily be good matches. More importantly, it is consistently applied to both the *Baijia* system and the AOL search engine; therefore, the comparison results should be fair and conclusive.

In our experiments, search epistemologies are contributed and shared through the following steps:

Step 1: "Users" completed their searches through iterative interaction with the system and contributed their search epistemologies. To simulate the contribution from users, we extracted every user's exploratory searches from their queries. Each exploratory search contains several queries that are contextually related. Cosine distance function is used to measure the contextual similarity between every two queries. We have totally extracted 1,201,497 exploratory searches.

Step 2: The system returned other users' search epistemologies that are relevant to the current user's queries from its epistemology repository. To simulate the sharing of

epistemologies, we retrieved the epistemology repository for relevant search epistemologies. An epistemology is relevant to an exploratory search if its queries are similar to the search queries, and the selected pages of the epistemology completely/partially match the clicked URLs of the search.

Step 3: If no relevant epistemologies are found at step 2, the exploratory search itself will be formulated as a search epistemology; otherwise, it will be integrated into existing relevant epistemologies. “Users” participated in the search activity by re-ranking the re-ranked results from other users or the ranked results from the search engine. To simulate the refinement of epistemologies, a random score is assigned to every clicked URL to represent the judgment from the current user. Actually it is common that users may have different opinions on the same search result. As the motivation of social search is to believe the wisdom of crowds, the result that is ranked highest by the majority of searchers is regarded as the best. For a social search system, a ranking mechanism that is based on the average scores of all participators’ follows this rationale. This ranking mechanism is adopted by the *Bajia* system, which is adaptive to accumulative users’ rankings, no matter whether they are computer-generated random scores or real human evaluation scores. Therefore, the random score assignment serves our purpose for the experiments and we could envisage an even better performance gain if real human scores were used in the future experiments. In addition, URLs that are repeatedly clicked are given higher scores. The selected URLs of every epistemology are re-ranked according to the scores. We have finally built 480,254 records in the epistemology repository.

Following the above steps, we have build up the initial epistemology repository for our system by importing all exploratory searches derived from the AOL query logs. It is worth pointing out that the initial epistemology repository can immediately benefit new exploratory information seekers, but the system can actually work without it. The system relies on the search engine to build up the epistemology repository at its initial stage and gradually relies more on the epistemology repository itself.

Several metrics have been adopted to evaluate *Bajia*’s performance, such as Mean Average Precision (MAP) [13], Precision at K (e.g., Precision@10) [13], and Normalized Discounted Cumulative Gain (NDCG) [5]. In this article, we present our experimental results based on MAP, which shows the overall performance of *Bajia* is superior to that of AOL search engine in exploratory information seeking.

Average precision of a query is defined as follows:

$$AP = \frac{1}{\text{Rel}} \sum_{r=1}^{\text{Rel}} \frac{\text{Pos}_r}{r} \quad (1)$$

where Rel is the total number of documents relevant to the query, and Pos_r is the position of the r^{th} relevant document in the list of all resultant documents. MAP is defined as the mean AP over all queries. It can stably reflect the overall performance of a search system [12].

4.2 Results and Discussion

We traced the generation of the epistemology repository. Our major concern is whether an exploratory search can benefit from the system’s epistemology repository.

Therefore, we introduce the Epistemology-repository Acquisition Rate (EAR) metric – the ratio of exploratory searches that successfully retrieve relevant epistemologies.

We computed the EAR and MAP scores of the *Baijia* system at different stages. In our experiments, the MAP is computed as the mean of every exploratory search’s AP, which is the weighted mean of precisions of all queries it consists of.

$$MAP = \frac{1}{N} \sum_{i=1}^N \left(\sum_{j=1}^{Q_i} w_j AP_j \right) \quad (2)$$

where N is the total number of exploratory searches, Q_i is the total number of queries in the i^{th} exploratory search, AP_j is the AP of the j^{th} query in the i^{th} exploratory search and w_j is the weight of it according to its importance to the exploratory search (e.g. a query with more clicked URLs will be assigned a heavier weight than a query without any clicked URL).

Fig. 5 shows the MAP scores of the *Baijia* system as compared to those of the AOL search engine (the original data). The results show that increase of the number of exploratory searches imported leads to improvement of MAP scores in the *Baijia* system while the MAP scores of AOL search engine remain steady. This can be explained with Table 1: the EAR increases as more exploratory searches are imported, which implies a user who submits an exploratory search will have a higher probability to get relevant search epistemology from the epistemology repository. Since search epistemology is extracted from clicked URLs and unclicked URLs have been filtered out, the MAP scores of the *Baijia* system are clearly higher than those of the AOL search engine, which can never benefit from the search epistemologies at all. Furthermore, when the search epistemologies in the epistemology repository are re-ranked according to users’ feedback rather than random weights, the MAP scores will even be significantly improved.

Unlike previous studies which mainly focus on augmenting search results with relevant data aggregated from the Semantic Web by pre-defined ontologies [29], our work aims at automatically constructing ontologies based on the sequence of queries in every exploratory information seeking process.

5 Conclusions & Future Work

We have proposed a novel epistemology-based social search framework EPISOSE for supporting exploratory information seeking, where search epistemologies – aggregated and well-structured information packages derived from successful search processes contributed by a mass of searchers – are effectively shared, reused, and refined by others with same or relevant search interests or goals. We have also

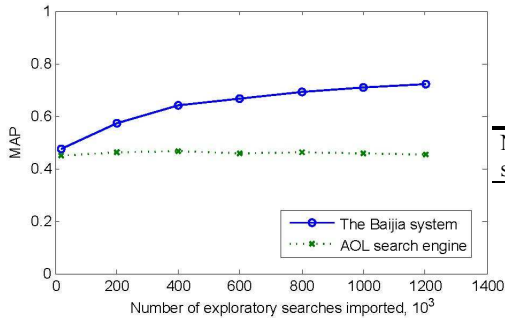


Fig. 5. MAP scores of the *Baijia* system and the AOL search engine

Table 1. Epistemology repository size and EAR at different stages

Number of exploratory searches imported	Epistemology repository size	EAR
20,000	7,612	18.35%
200,000	74,634	29.67%
400,000	151,392	34.20%
600,000	230,273	37.37%
800,000	311,167	39.93%
1,000,000	394,266	41.71%
1,201,497	480,254	42.52%

implemented a prototype system *Baijia* based on the framework and conducted a set of experiments to prove that the proposed solution can outperform a conventional search engine in supporting exploratory information seeking. Preliminary usage study indicates that utilizing search engines' immense power and human users' intelligence is an effective and pragmatic solution to exploratory web search.

We have introduced *Baijia* on our intranet to get some initial usage feedback. Most feedback confirms the improvement of the search efficiency in various situations. A main dissatisfaction is that not enough well-refined up-to-date search epistemologies were available to benefit from at the elementary stage. This situation will be improved as the user number increases and the epistemology repository grows. As EPISOSE is an epistemology-based social search framework, *Baijia* relies more on search engines at its initial stage and is self-adaptive to the growing human factors of the system.

Currently, we are conducting a user study to validate the proposed approach. As a social search framework based on users' contribution and reusing, a user satisfaction study is helpful. For example, we could tune the configuration of the prototype system after testing the protocol and algorithm in the real world. In the future, data mining and other artificial intelligence technologies will be adopted to discover the definite objective of an exploratory web search task from the accumulated search epistemologies. We also plan to research and develop components for users to conveniently build social networks and share their intimate search epistemologies for exploratory information seeking.

References

1. Baeza-Yates, R., Tiberi, A. (2007) Extracting semantic relations from query logs. In Proceedings of KDD'07, 76-85.
2. Herlocker J, Konstan J, Borchers A, Riedl J. An algorithmic framework for performing collaborative filtering. In: Proceedings of SIGIR '99, 230-237.
3. Heymann, P., Koutrika, G. and Garcia-Molina, H. (2008) Can social bookmarking improve web search? In Proceedings of WSDM'08, 195-205.
4. James Surowiecki. (2004) *The Wisdom of Crowds: Why the many are smarter than the few and How collective wisdom shapes Business, Economies, Societies and Nations*,
5. Jarvelin, K. and Kekalainen, J. (2000) IR evaluation methods for retrieving highly relevant documents. In Proceedings of SIGIR '00, 41-48.
6. Joachims, T. (2002) Optimizing search engines using clickthrough data. In Proceedings of KDD'02, 133-142.
7. Marchionini, G. (2006) Exploratory search: from finding to understanding. *Communications of the ACM*, 49, 4 (2006), 41-46.
8. Morris, M. R. and Horvitz, E. (2007) SearchTogether: an interface for collaborative web search. In Proceedings of UIST '07, 3-12.
9. Pass, G., Chowdhury, A. and Torgeson, C. (2006) A Picture of Search. In Proceedings of the 1st International Conference on Scalable Information Systems. Infoscale '06.
10. Ravi, S. (2005) Optimal search engine marketing strategy, *International Journal of Electronic Commerce* 10, 1 (2005), 9-25.
11. Teevan, J., Adar, E., Jones, R., and Potts, M. (2007) Information Re-retrieval: Repeat queries in Yahoo's Logs. In Proceedings of SIGIR'07, 151-158.
12. Voorhees, E. M. and Harman, D. (1999) Overview of the 7th Text REtrieVal Conference (TREC-7). In Proceedings of the 7th Text REtrieVal Conference (TREC-7), 1-23.