

Peer Knowledge Assisted Search Using Community Search Logs

Brendan Flanagan^b, Takahiko Suzuki^{a*}, Jun Zeng^b, Chengjiu Yin^a, Toshihiko Sakai^b,
Kiyota Hashimoto^c and Sachio Hirokawa^a

^a Research Institute for Information Technology, Kyushu University, Japan

^b Graduate School of Information Science, Kyushu University, Japan

^c School of Humanities and Social Sciences, Osaka Prefecture University, Japan

*6-10-1, Hakozaki, Higashi, Fukuoka, Japan

*suzuki@cc.kyushu-u.ac.jp

ABSTRACT

Search engines have become an increasingly important educational tool. By just typing a few keywords into a search engine text-box, finding information has become an easy exercise. However, some people cannot use search engines effectively for various reasons. Students may not know how to choose proper keywords. They may have difficulty in selecting relevant information from millions of search results. In this paper, we propose a search engine system that shares search queries and browsing history among students. We refer to the search queries and browsing history as “community search logs”. This system analyzes these community search logs, and shares the students’ knowledge and experience with each other. Our purpose is to help students to improve their searching efficiency, especially those who are not good at searching.

KEYWORDS

Community Search Logs, Search Engine, Knowledge and Experience Sharing, Peer Knowledge Assisted Search.

1 INTRODUCTION

Along with the development of the Internet and search engine technology, search engines have become an increasingly important educational tool. Researchers have indicated that allowing students to learn as active and self-

directed participants is one of the greatest benefits of web-based learning activities, which often involve information searching tasks [1, 2]. Marchionini [3] and Dias et al. [4] have reported the difficulty that novice Internet users often encounter when they try to search for information on the web effectively and efficiently. Training novice users in information searching skills, including abstracting and summarizing the collected information from the Internet, has become an important and challenging issue [5].

Search engines are programs that search documents for specified keywords and return a list of the documents where the keywords were found. However, the term is often used to specifically describe systems such as Google, Bing and Yahoo! Search that enable users to search for documents on the World Wide Web. Nowadays search engines have become an increasingly important tool. Finding information has become an easy exercise by just typing a few keywords into a search engine text-box. Some people can effectively use search engines and always find the most suitable keywords to use that return the most relevant search results. This can be of great advantage to their work and life. However, not everyone can use search engines effectively. Users may not know how to choose proper keywords. For example, when looking for science

experiments, if “School Science Fair Projects” is entered instead of “science projects”, this will return more specific information.

Users may also have difficulty in selecting the relevant information from a large number of search results. Although most search engines employ methods to rank results in order to provide the “best” results first, not all search results that are ranked high are relevant to the users’ needs.

We propose a search engine system that shares search queries and browsing history among a group of users. We refer to the search queries and browsing history as “community search logs”. This system analyzes these community search logs, and shares the users’ knowledge and experience with each other. Our purpose is to help students to improve their searching efficiency, especially those who are not good at searching.

This paper is structured as follows: Section 2 is a brief introduction of the related works. In section 3, we explain the preparation of data and a preliminary experiment. Section 4 describes the proposed system in detail. Section 5 describes our conclusions and future work.

2 RELATED WORKS

Previous research has been conducted on using the access logs of Web pages. Toda et al. [6] focused on Web pages that were often visited by many users in order to improve the efficiency of Web browsing routines as well as the discovery of new information. In order to evaluate the importance of the pages for recommendation, they analyzed the time and frequency of users who visited particular pages in access logs. This was

then used to provide collaborative filtering for Web page recommendation. Nakao et al. [7] used the web page access logs of each user to calculate the similarity of Web pages. They then proposed a Web page recommendation method based on similarity. They considered not only the similarity of Web pages, but also the similarity of the sub-trees in which the pages are located. These two methods were then combined in a link structure.

So far great importance has been placed on qualitative analysis. Recently, learning logs have become popular with many researchers using them to enable communication between users. Gotoda et al. [8] used the activity logs of runners in creating regional and focused communities of runners. Any member can read and comment on the training plan of other members. This enables participating members of the community to share and learn from each other’s activity logs even if they are experienced.

Watanabe et al. [9] used collaborative learning logs to help participants in a team to solve the same problem. They evaluated the number of users who made the same annotation of their utterances, and constructed a knowledge graph. They claimed that the graph is useful to point out key issues for learners.

Anderson et al. [10] designed and implemented a montage system to improve the experience of routine web browsing that users tend to repeat over and over in similar situations by providing a start page. The start page shows an ensemble of links and content based on the user’s browsing log and preferences.

Prior study by X Wu et al. [11] proposed a community based search platform which combines a search engine and

SNS search. By seamless use of various activities performed while searching and mutual comments, users can share information on problems or new knowledge.

3 DATA PREPARATION AND PRELIMINARY EXPERIMENT

Two kinds of data are utilized in the proposed system: a database containing search targets that we refer to as “search data”, and the users’ search query and browsing history that we refer to as “community search logs”. In this section, we will introduce the procedure used to collect both of these kinds of data.

3.1 Collecting Search Data

Information on papers about “mobile learning” was collected as the search data. The keywords “mobile learning” was used to conduct a search using Scopus¹, which is the world’s largest abstract and citation database of peer-reviewed literature. As a result, we collected the information of 13353 papers. Each paper contains the following components:

- Title of the paper
- Authors of the paper
- Departments, organizations and countries of the authors
- Publisher
- Publication date
- Keywords
- Abstract

A frequency file of the words appearing in those components was generated. Finally, this frequency file was then used to build an index using “GETA”², which

is a generic engine for transposable association.

3.2 Preliminary Experiment

At the initial stage of research no community search log data was available. A preliminary experiment was conducted to create community search logs by bootstrapping. In this experiment, we used another system called the “Milky Way” search engine, which was introduced in paper [12]. This system uses the same experimental data as the aforementioned search data.

The participants were two sophomore student classes of a north Taiwan university. A total of 69 students participated in the study, including 25 females and 44 males with an average age of 20 years old. One class was assigned to be the experimental group, and the other was the control group. The experimental group included 36 students (12 females and 24 males), while the control group had 32 students (13 females and 20 males). In order to avoid any effect on the experiment results that may be caused by different instructors, the two classes were taught by the same instructor. The experimental group learned with the “Milky Way” search engine, while those in the control group learned with a search engine that has a conventional interface (i.e., displaying the searched results as a list of document titles and abstracts). Both groups received the information searching and summarizing instructions for analyzing the research trends of mobile learning articles before the learning activity. The details of the “Milky Way” system are reported in [12].

In general, the “Milky Way” could help users in their learning activities by providing various search hints. However,

¹ [http:// www-scopus-com](http://www-scopus-com)

² <http://geta.ex.nii.ac.jp/>

some novice users might complain that they found it difficult to utilize those various search hints.

3.3 Collecting Community Search Logs

The search queries and browsing history of the experimental group were collected as the log data. As a result, 84 search queries and 192 browsed web page histories were collected. Figure 1 shows the segments of the search queries. Each search query contains the following:

- The number of users who used the query
- The keywords of the query

Figure 2 shows the segments of log data. Each entry of the log contains the following:

- The IP address of the PC used by the participants
- The date and time when the participants browsed the search results
- Paper number (1~13353)
- Keywords used by participants

```

...
22 mobile learning
17 mobile learning y:2010
9 mobile learning device y:2010
7 mobile learning y:2001
7 mobile learning c:Taiwan
...
    
```

Figure 1. Segments of search data

```

...
111.249.47.11,201201041255,4230,y:2011 mobile learn
133.5.7.108,201201041256, 5109,mobile learn ubiquitous
163.14.7.115,201201101200,8413,mobile learning c:Japan
163.14.7.82,201201101209,8523,mobile learning
...
    
```

Figure 2. Segments of log data

4 THE PROPOSED SYSTEM

4.1 Design Goal

Figure 3 shows the design goal of the proposed system. The system shares the search query history and browsing activities each user has conducted so far in order to help other users improve their searching efficiency. When a user conducts a search using the search engine, his/her query and browsing history are recorded as data reflecting their search behaviors. The user's search behaviors can be influenced by his/her knowledge, experience and other factors [13]. In other words, sharing search queries and browsing history is also sharing the user's knowledge and experience.

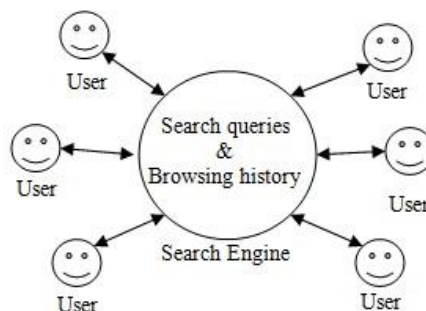


Figure 3. Design goal of our system

The effect of just one user's search queries and browsing history is limited. However by using the searching history from a large number of users' can not only provide other users with recommendations, but can also provide basic data to perform statistical analysis of users' search behaviors. When used in this way, search engines will become large data warehouse of users' behaviors.

However, at the initial stage of the research only the log data mentioned in section 3 was utilized, and no new log data was created while users were using the proposed system.

4.2 The Functions and User Interface Of The Proposed System

This section introduces the functions and user interface of the proposed system. Figure 4 shows the main user interface containing three frames.

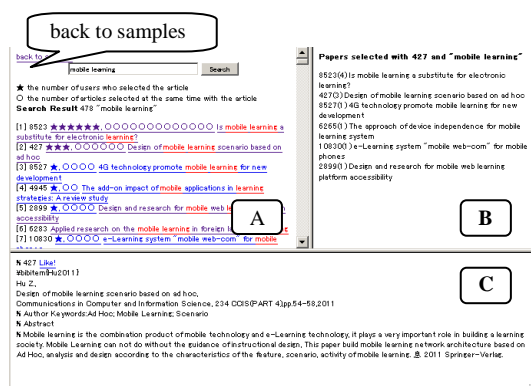


Figure 4. User interface of the proposed system

1) Frame A displays the search results. This can also be used to conduct keyword-based searches. The search results will be displayed below the query text-box. Each search result contains the following parts:

- Paper Number (1~13353)
- Title of the paper
- ★ (stars) indicate how many users have selected the paper
- ○ (circles) indicate how many other papers were browsed in the past search of the paper.
- This page not only displays the search queries, but also how many times the queries have been used. When the links of queries are clicked, the system will automatically conduct a

search without having to inputting a search query manually.

There is a “back to samples” link at the top of frame A that redirects the user back to sample list page as shown in Figure 5.



Figure 5. Search query list in frame A

2) Frame B displays the related papers and users: When the ★ of a paper in frame A is clicked, frame B will display the list of the users who have browsed the paper. In frame B, the following items are displayed:

- User ID
- The number of times that the user selected the paper.

When the ○ of a paper in frame A is clicked, frame B will display the papers that were browsed at the same time as the paper. In frame B, the following items are displayed:

- Paper Number (1~13353)
- The times a paper has been browsed
- Title of the paper

3) Frame C displays the detail of a paper: When the paper title in frame A is clicked, frame C will display the details of the paper. The following items are displayed in frame C:

- Title of the paper
- Authors
- Departments, organizations and countries of the authors

- Publisher
- Publication date
- Keywords
- Abstract

There is a “like” button at the top of frame C as shown in Figure 6. Users can express their interest in the search result document by clicking the button.

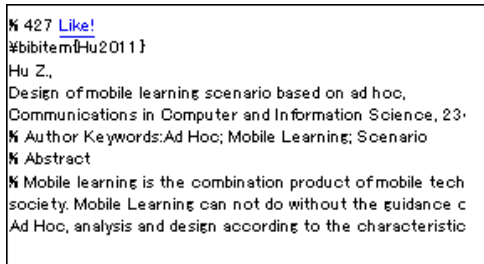


Figure 6. Details of frame C

4.3 An Example Of Searching Using The Proposed System

This section will describe an example of searching using the proposed system.

Conducting a search: Firstly, we use “mobile learning” as a search query to conduct a search. The search results in frame A are shown in Figure 7.

There are 478 search results for the query “mobile learning”. An example of how to read the results can be seen in paper “8523” which has 6 ★ and 13 ○. This means that 6 users have browsed the paper and 13 other papers were browsed at the same time as paper “8523”.

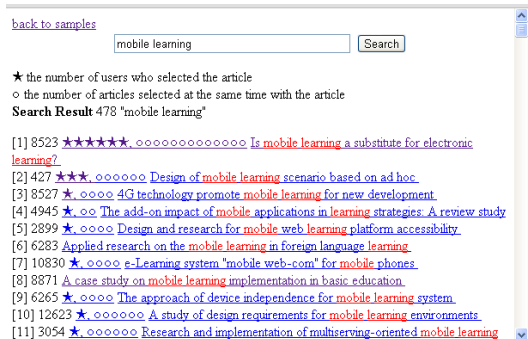


Figure 7. Search results in frame A.

Analyzing the related users: When one of the ★ associated with paper “8523” is clicked, frame B displays all the ID’s of the users who browsed the paper. Figure 8 shows a list of the users in frame B. The number of users in frame B is equal to the number of ★ display next to paper “8523” in frame A. The number in the parentheses following the user ID indicates how many times the user browsed the paper.

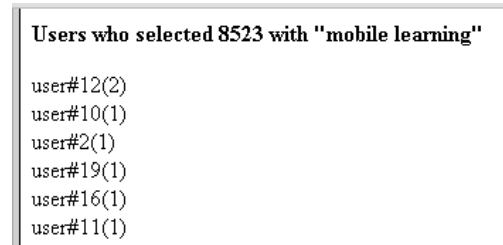


Figure 8. User list in frame B

Analyzing the related papers: When one of the ○ associated with paper “8523” is clicked, frame B displays all the papers that were browsed at the same time as this paper. Figure 9 shows the list of papers in frame B. The number of the papers in frame B is equal to the number of ○ for paper “8523” in frame A. The number in the parentheses following the paper number indicates how many times the paper has been browsed. For example, paper “8523” has been browsed 7 times. This is because user#12 browsed this paper twice, and the other users only browsed this paper once.

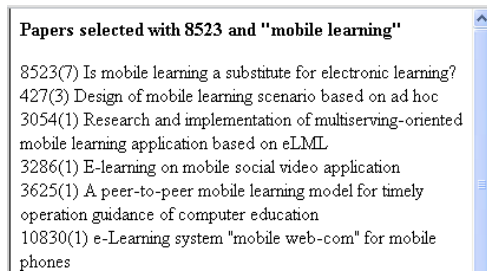


Figure 9. A segment of a paper list in frame B

5 CONCLUSIONS AND FUTURE WORK

When a student conducts a search, his/her search log, which contains valuable information, can be a precious resource to other users. The information can help students, especially those who are not good at searching, to improve their searching efficiency. In this paper, we proposed a search engine system that can share search queries and browsing history among students. We refer to the search queries and browsing history as “community search logs”. This system analyzes the community search logs, and shares the students’ knowledge and experience with each other.

In the future, we are planning to analyze the relation between the students who browsed the same papers and the common features of the papers that have been browsed at the same time.

REFERENCES

1. D. Bilal. “Children’s use of the Yahoo!igans! web search engine: I. cognitive, physical, and affective behaviors on fact-based search tasks”, *Journal of the American Society for Information Science*, 51, 2000, pp.646–665.
2. G.J. Hwang, P. S. Tsai, C. C. Tsai, & J. C.R. Tseng, “A novel approach for assisting teachers in analyzing student web-searching behaviors”, *Computers & Education*, 51(2), 2008, pp.926-938.
3. G. Marchionini, *Information Seeking in Electronic Environments*. Cambridge University Press, New York, 1995.
4. P. Dias, M. J. Gomes, & A. P. Correia. “Disorientation in hypermedia environments: mechanisms to support navigation”, *Journal of Educational Computing Research*, 20(2), 1995, pp.93-117.
5. G.J. Hwang & F. R. Kuo, “An information-summarizing instruction strategy for improving web-based problem solving abilities of students”, *Australasian Journal of Educational Technology*, 27(2), 2011, pp.290-306.
6. T. Toda, T. Mine, & M. Amamiya, “Active Browsing: A Novel Personalized Web Browsing Support System”, *Tech. Rep. IEICE SIG-KBSE 106(473)*, 2007, pp.1-6 (in Japanese).
7. K. Nakao & T. Mine, “A Method for Relevance Calculation between Browsed Web Pages toward Personalized Recommendation System”, *Tech. Rep. IEICE SIG-AI 109(51)*, 2009, pp.31-36 (in Japanese).
8. N. Gotoda, K. Matsuura, K. Kanenishi, & Y. Yano, “Support of the Running Motivation based on activity Log”, *Tech. Rep. IEICE SIG-ET, 107(536)*, 2008, pp.65-68 (in Japanese).
9. Y. Watanabe, T. Kojiri, & T. Watanabe, “Knowledge Organization Framework from Discussion Records of Collaborative Learning”, *Tech. Rep. IEICE SIG-ET, 109(82)*, 2009, pp.21-26 (in Japanese).
10. C. R. Anderson & E. Horvitz, “Web Montage: A Dynamic Personalized Start Page”, *WWW’02: Proc. of the 11th international conference on World Wide Web*, 2002, pp.704-712.
11. X. Wu, J. Zeng, C. Yin, & S. Hirokawa, “Sharing Knowledge and Experience of Search with SNS”, *The 17th International Symposium on Artificial Life and Robotics 2012 (AROB 17th 2012)*, B-Con Plaza, Beppu, Oita, Japan, Jan 19-21, 2012, pp.101-104.
12. B. Flanagan, C. Yin, S. Hirokawa, H. Y. Sung, & G.J. Hwang, “Analyzing Research Trends of Mobile Learning with the Milky Way”, *Proc. 7th International Conference on Wireless, Mobile and Ubiquitous Technology in Education (WMUTE2012)*, 2012, pp.249-253.

13. Y. Chuang & L. Wu, "User-Based Evaluations of Search Engines: Hygiene Factors and Motivation Factors", Proc. of 40th Annual Hawaii International Conference on System Sciences, 2007, pp.1-10.