FOCUSED SEARCH USING COMMUNITY SEARCH LOGS

Brendan Flanagan[a], Takahiko Suzuki[a*], Jun Zeng[a], Chengjiu Yin[b], Toshihiko Sakai[a], Kiyota Hashimoto[c] & Sachio Hirokawa[b]
[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
suzuki@cc.kyushu-u.ac.jp

ABSTRACT

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

KEYWORDS

Community Search Logs, Search Engine, Knowledge and Experience Sharing, Focused Search.

1 INTRODUCTION

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 Internet novice users encounter when they try to search 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 like 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. They can 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", they will return more specific information.

Users may also have difficulty in selecting the relevant information from millions of search results. Although most

search engines employ methods to rank results in order to provide the "best" results first, not all search results which are ranked high are relevant to the users' needs.

We propose a search engine system which shares search queries and browsing history among students. We call the search queries and browsing history "community search logs". This system analyzes the community search logs, and shares users' knowledge and experience with each other. Our purpose is to help students, especially those who are not good at searching, to improve their searching efficiency.

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

2 RELATED WORKS

Previous research has been conducted on using access log of Web pages. Toda et al. [6] paid attention to the Web pages visited often by many users in order to improve the efficiency of Web browsing routines as well as discovering new information. They analyzed the time and frequency of users who visit to particular pages in their access log to evaluate the importance of the pages for recommendation. They used collaborative filtering for Web page recommendation.

Nakao et al. [7] uses the web page access log of each user to calculate the similarity of Web pages. They propose a Web page recommendation method based on similarity. They considered not only the similarity of Web pages but also the similarity between sub-trees. They then combined these two similarities in a link structure.

So far great importance has been placed on qualitative analysis. Recently, learning logs have become popular with many researchers to help 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. So, participating members of the community can share and learn from others' activity logs even if they experienced.

Watanabe et al. [9] used collaborative learning logs to help participants form 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 issue for learners.

Anderson et al. [10] designed and implemented a montage system to improve the experience for 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 a user's browsing log and preferences.

Prior study by X Wu et al. [11] proposed a community type search platform which combines a search engine and SNS search. By seamless use of various activities in search and the mutual comments, users can share problems or new knowledge.

3 DATA PREPARATION AND PRELIMINARY EXPERIMENT

There are two kinds of data that are used in the proposed system. One kind of data

is the database of the search target. We call this data "search data". The other kind of data is the users' search query and browsing history. We call this data "community search logs". In this section, we will introduce the procedure of collecting these two kinds of data.

3.1 Collecting Search Data

We collected information of papers about "mobile learning" as the search data. We used "mobile learning" as keywords and conduct a search using Scopus [1] which is the world's largest abstract and citation database of peer-reviewed literature. As a result, we collected information about 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

Then, we created a frequency file for the words appeared in those components. Finally, we used "GETA"[2], which is a generic engine for transposable association, to build an index based on the frequency file.

3.2 Preliminary Experiment

At initial stage of research no data for community search logs were available. A preliminary experiment was conducted in order to create bootstrapping community search logs. In this experiment, we used another system called "Milky Way" search engine, which has been introduced in paper [12]. The experiment data are the same as the search data mentioned above.

The participants were two classes of sophomore students of a university in north Taiwan. A total of sixty-nine students participated in the study, including twenty-five females and forty-four 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 thirty-six students (twelve females and twenty-four males), while the control group had thirty-three students (thirteen females and twenty males). In order to avoid the influence of different instructors on the experimental results, 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 with a conventional interface (i.e., displaying the searched results by showing 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 evaluation of "Milky Way" system are reported in [13].

The present paper does not concern the system or the evaluation of the system, but concerns how to utilize the search logs of the system.

[1] http:// www-scopus-com
[2] http://geta.ex.nii.ac.jp/

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, we collected 84 search queries and 192 browsed web page history. Figure 1 shows the segments of search query. Each search queries contains the following:

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

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

- IP address of the PC used by participants
- 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. Segment 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. Segment of log data

4 THE PROPOSED SYSTEM

4.1 Design Goal

As shown in Figure 3, the design goal of the proposed system is to share search queries and browsing history in order to help the users to improve their searching efficiency. When users use the search engine to conduct a search their search queries and browsing history will be recorded reflecting the users search behaviors. User's search behaviors can be influenced knowledge, experience and other factors [15]. In other words, sharing search queries and browsing history is sharing the user's knowledge and experience.

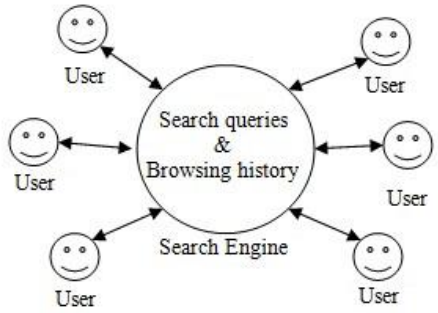

Figure 3. Design goal of our system

The effect of just one user's search queries and browsing history is limited. However a large number of users' searching history data can not only provide other users with a recommendation, but also provide basic data to perform statistical analysis of users' search behavior. In this way, search engines will become a large data warehouse of users' behavior.

However, at the initial stage of research, we will only analyze the log data mentioned in section III. In other words, no new log data will be created when users use the proposed system.

4.2 Functions And The Interface Of The Proposed System

This section will introduce the functions and interface of the proposed system. Figure 4 shows the main interface which contains three frames.

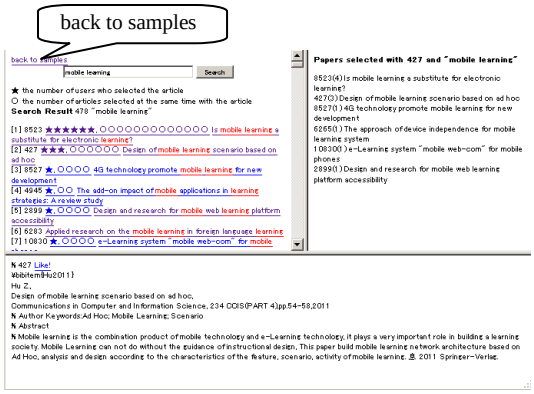

Figure 4. Interface of the proposed system

1) Frame A displays the search results: in this frame, users can conduct keyword-based searches. The search results will be displayed below. Each search result contains the following parts:

- Paper Number (1~13353)
- Title of the Paper
  ★ (stars) indicate how many users had selected the paper
- ○ (circles) indicate how many other papers were browsed in the past search of the paper.

There is a link "back to samples" at the top of frame A. When the link is clicked, the system will redirect to sample list page as shown in Figure 5.

Figure 5. Search query list page of the system

This page not only displays the search queries but also displays how many times the queries have been used. When the links of queries are clicked, the system will automatically conduct a search without inputting a search query manually.

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 times that the user selected the papers.

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 the papers have 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 detail of the paper. In frame C, the following items are displayed:

- Title of the paper
- Authors

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

4.3 Search Instance Of The Proposed System

This section will introduce a search instance of the proposed system.

1) Conducting a search: Firstly, we used "mobile learning" as a search query to conduct a search. Figure 6 shows the search results in frame A.

There are 478 search results for the query "mobile learning". In the search results for example, paper "8523" has 6 ★ and 13 ○. It means that 6 users browsed the paper and 13 other papers were browsed at the same time as paper "8523".

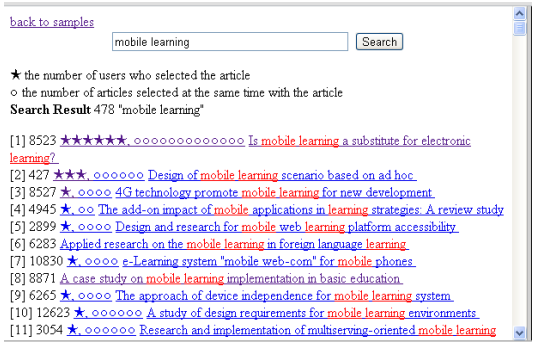

Figure 6. Search results in frame A.

Analyzing the related users: When the ★ of paper "8523" is clicked, frame B displays all the users' ID who browsed this paper. Figure 7 shows the list of the users in frame B. The number of users in frame B and the number of ★ for paper "8523" in frame A are equal. The number in the parentheses following the user ID indicates how many times that the user browsed this paper.

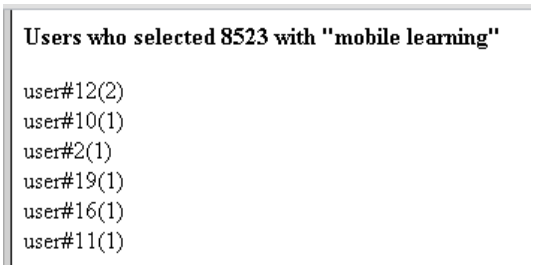

Figure 7. User list in frame B

Analyzing the related papers: When the ○ of paper "8523" is clicked, frame B displays all the papers that were browsed at the same time as this paper. Figure 8 shows the list of papers in frame B. The number of the papers in frame B and the number of ○ for paper "8523" in frame A are equal. The number in the parentheses following the paper number indicates how many times this paper has been browsed. For example, the paper "8523" has been browsed 7 times. This is because user#12 browsed this paper twice, and other users browsed this paper only once.

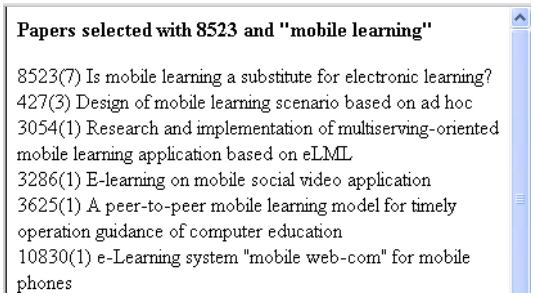

Figure 8. A segment of paper list in frame B

5 CONCLUSIONS AND FUTURE WORK

When a student conducts a search, his/her search log can be valuable information 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 which can share search queries and browsing history among students. We call the search queries and browsing history "community search logs". This system analyzed the community search logs, and shares 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 Yahooligans! 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. C. Yin, Y. Tabata, & S. Hirokawa, "A "Milky Way Research Trend" system for Survey of Scientific Literature", Proc. of the 10th International Conference on Web-Based Learning (ICWL 2011) workshop on "Enhancing Learning with Social Media(ELSM)", Springer LNCS, Hongkong, China, Dec.8-Dec.10 , 2011 (in press).
13. 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.
14. 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.