A Web Retrieval Support System with a Comment Sharing Environment: Toward an Adaptive Web-based IR System

Hiroyuki Mitsuhara[1], Youji Ochi[1], Kazuhide Kanenishi[2] and Yoneo Yano[1]
[1] Faculty of Engineering, Tokushima University
[2] Center for Advanced Information Technology, Tokushima University
2-1 Minamijosanjima, Tokushima, Japan
E-mail: {mituhara, ochi, marukin, yano}@is.tokushima-u.ac.jp

Abstract

A search engine is frequently used to look up interesting/unknown topics in web-based exploratory learning, but occasionally it returns meaningless pages (hyperlinks), which can interrupt finding suitable pages. This paper describes a web retrieval support system with a comment sharing environment, which aims at reducing loads in search activity. The effectiveness of this system has been shown through an experiment. This system is not currently adaptive but can be extended to a web-based adaptive IR (Information Retrieval) system.

Keywords: Comment sharing environment, search engine, adaptive web-based IR system, loads in search activity, exploratory learning.

1. Introduction

The web can be regarded as vast teaching material since it has numerous informative pages. Nowadays, learners actively learn through exploring the web. Such a learning method is called exploratory learning and practiced in not only school education but also life long learning. At the beginning of exploration or during exploration, learners frequently try to find suitable pages for fulfilling their needs, goals, or interests by using a search engine. A search engine is very powerful tool but occasionally it returns meaningless pages (hyperlinks), which can interrupt finding suitable pages. This is caused by an insufficient search query. In particular, novice learners who are unfamiliar with a learning topic can have difficulty in designating a sufficient search query to narrow down a search focus. Even if meaningless pages are eliminated with a sufficient search query, it is hard for learners to select a suitable page quickly from similar pages. Although search results generally consist of page title, URL, summary text, and statistic, these are not useful enough to select a suitable page. In this situation, learners have to visit a page and estimate its value repeatedly. Thus, learners are excessively given loads in search activity and consequently lose learning volition.

Recent search engines are equipped with functionalities of supporting search activity. For example, Alexa Web Search presents users with information for clarifying retrieved pages such as ratings, comments, related pages, etc, which are offered by users [1]. Lycos analyzes search query logs with the technique of agglomerative clustering and suggests alternative search queries for rephrasing users' needs [2]. SearchPad recalls users to successful search by presenting individual users' histories of search queries and visited pages [3]. In addition, many adaptive web-based IR (Information Retrieval) systems have been developed that function as a meta-search engine [4]. Syskill&Webert suggests suitable search queries and annotates pages (hyperlinks) of a search result on the basis of individual users' interests [5]. A system developed by Marinilli et al. sorts pages on the basis of individual users' preferences [6]. A mechanism proposed by Huang et al. suggests additional search queries on the basis of individual users' needs [7]. GOOSE elicits suitable search queries from natural language that articulates individual users' goals [8].

A web search retrieval support system WebCOBALT (Web Contents Observable System) has been developed in order to reduce loads in search activity. This system enables learners to write comments on pages retrieved by an existing search engine and share the comments. Learners can find a suitable page without excessive loads, referring to the comments. This system, which adopts Alexa Web Search's approach, is not currently adaptive but can be extended to a web-based adaptive IR system.

The remainder of this paper is organized as follows. Section 2 describes the outline of WebCOBALT, focusing on its fundamental idea and functions. Section 3 illustrates its user interface. Section 4 reports the results of a small-scale experiment. Section 5 considers prospective adaptivity of WebCOBALT.

Proceedings of the International Conference on Computers in Education (ICCE'02)
0-7695-1509-6/02 $17.00 © 2002 IEEE

2. WebCOBALT

WebCOBALT reduces loads in search activity by supplementing drawback in search engines with a comment sharing environment, which enables users to write comments (e.g. complementary explanations, questions, and personal remarks) on information (e.g. web pages) and share the comments. This environment can heighten learning effectiveness. For example, SharlokII facilitates problem solving by enabling learners to share various kinds of information such as questions, answers, comments, and hyperlinks [9]. ReCoNote facilitates knowledge construction by enabling learners to share web notes structured with mutual links with comments [10]. Annotea, which enables users to extend shared comments (annotations) based on XML technologies, may facilitate long-term discussion resulting from the comments [11].

In WebCOBALT, learners can write comments on pages retrieved by an existing search engine and share the comments attached to the search result. Each comment has one of three attributes, "informative", "neutral", "useless", which is designated by a learner. The comments may indicate informative pages, which are not shown in summary text generated by a search engine. For example, "This page is easy to understand thanks to useful diagrams" and "This page is proper for review because there are many quizzes", which have the "informative" attribute, can be clues used to find a suitable page. Conversely, "This page does not describe truth" and "This page is filled with harmful information", which have the "useless" attribute, can be clues not to visit unsuitable pages. Learners refer to the comments and efficiently find a suitable page without actually visiting the pages. Thus, enabling learners to share the comments leads to reduction of loads in search activity. Figure 1 shows the framework of WebCOBALT. As shown in Fig.1, this system is similar to a meta-search engine[1]; there are advantages from the viewpoint of system developers: no intricate retrieval algorithms and no huge databases. WebCOBALT has the following functions in addition to comment sharing.

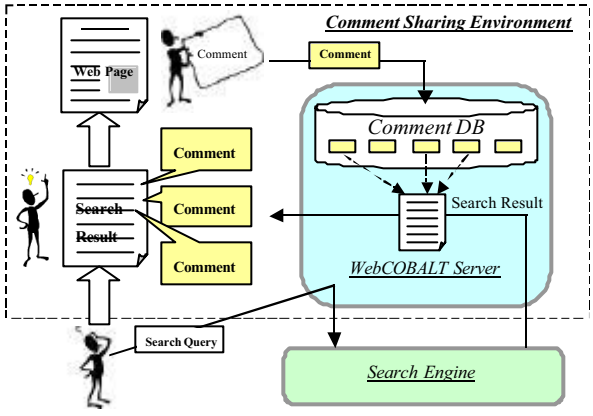

Figure 1. Framework of WebCOBALT

Bookmark: When finding an informative page during exploration, learners usually record the page as bookmarks, which are subsequently used to reflect on learning. WebCOBALT automatically builds a personal bookmark including pages with informative comments (i.e. comments with the "informative" attribute) written by a learner. This function facilitates not only reflection but also the increase of comments.

Sorting: The informative contained in a page will be clarified as the comments on the page increase. WebCOBALT simplifies finding a suitable page by sorting pages retrieved in descending order of the number of comments on each page.

3. User Interface

Learners can use this system on a web browser without plug-in software.

Main Frame: In this frame (Fig. 2(A)), the search result retrieved by goo[2] is displayed using stretch-text, which reduces information overload by hiding comments in the initial state; page titles and the number of comments are displayed. In order for a learner to notice pages with comments written by him/herself, a red circular icon is embedded in the titles of such pages. As soon as he/she clicks on a page title, its comment area is expanded/closed under the title. The comment area describes URL, summary text generated by goo, comment, commentator's e-mail address, etc. If a page has some comments, the learner can choose one from a commentator list (pull-down menu). By clicking on the "Read this page" button, the page is presented in this frame.

Retrieval Frame: In this frame (Fig. 2(B)), components for designating a search query are displayed. In addition, the learner can invoke his/her bookmark by clicking on the anchor text of "My Bookmark".

Bookmark Window: In this window, a personal bookmark is displayed in the same design of the search result. The learner can effectively reflect on learning, referring to comments attached to his/her bookmark. He/she can delete a page from his/her bookmark.

Comment Window: In this window (Fig. 2(C)), components for writing comments are displayed. This window disappears immediately after the learner clicks on the "Submit" button.

4. Experiment

[1] Strictly, a meta-search engine may utilize several existing search engines.

[2] A commercial Japanese search engine (robot search service). http://www.goo.ne.jp/

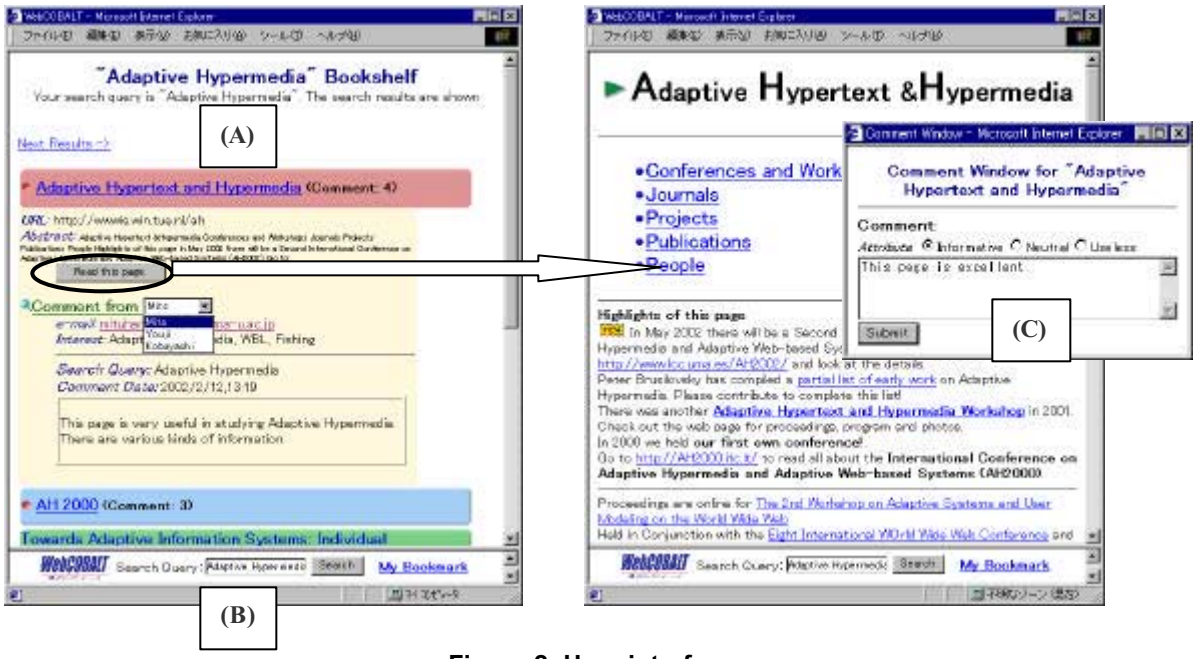

Figure 2. User interface

A comparison experiment between WebCOBALT and an existing search engine was done on a small scale in order to evaluate the effectiveness of comment sharing. Subjects of this experiment were graduate and undergraduate students who belonged to the department of computer science at Tokushima University. Twenty subjects were randomly divided into the following groups and were given the same task.

Group A (10 persons): They used goo.

Group B (10 persons): They used WebCOBALT without a bookmark function or a comment window.

Task: They had to look up nine items of three topics shown in Table 1 by using an allocated system (goo or WebCOBALT) and record answers to each item in a DOC file by copy & paste.

Specific evaluating points were the number of pages that were visited in each topic and the time needed to complete each topic. In this experiment, search queries and search options were fixed to ensure that all the subjects always received the same search result. Experiment time was not restricted and the subjects were allowed to abandon the items when they were not able to find suitable pages. WebCOBALT had attached the following types of comments to the search result in advance.

Comment type 1: It is clear that the page satisfies an item (e.g. "This page has a photomicrograph of bacteria that are the origin of puffer's poison").

Comment type 2: It is clear that the page does not satisfy an item (e.g. "This page describes how to cook Afghan foods").

Comment type 3: It is unknown whether or not the page satisfies an item (e.g. "This page is good to study copyright law superficially").

Topic A: Puffer's poison (Tetrodotoxin)	
Traits of this topic in search activity: The subjects have to find suitable pages from many pages that include similar information.	
A-1	Find a photomicrograph of bacteria that make puffer's poison.
A-2	Does a horseshoe crab have puffer's poison?
A-3	What are symptoms of puffer poisoning and measures against it?
Topic B: Afghanistan	
Traits of this topic in search activity: The subjects have to find suitable pages from many pages unrelated with the task (e.g. pages concerning fund raising).	
B-1	Enumerate races, languages, religions, and the government of Afghanistan.
B-2	How many Afghan refugees exist?
B-3	Look up recent medical treatment and humanitarian supports in Afghanistan.
Topic C: Copyright law	
Traits of this topic in search activity: The subjects will have difficulty in finding suitable pages directly from the search result. They have to find suitable pages by traversing a few layers.	
C-1	What kinds of computer programs are protected by copyright law?
C-2	Under what conditions can we quote information with copyright?
C-3	Look up copyright law in education.

Table 1. Tasks in the experiment

4.1 Results

DOC files collected indicated that one subject in group A abandoned two items and one subject in group B abandoned one item. The others completed all the items.

Time: Figure 3 shows the mean time (min.) needed for the subjects to complete each topic. Figure 4 shows the subjective time needed to complete the task (all the items), which was asked with a questionnaire (five degrees) "Did you complete this task earlier than you predicted?"

For the two topics, actual time consumption by group B exceeds that of group A. On the other hand, subjective time consumption by group B is fairly superior to that of group A.

Visit: Figure 5 shows the mean number of pages that the subjects visited in each topic. In topic A and B, the mean number of group B is notably lower than that of group A.

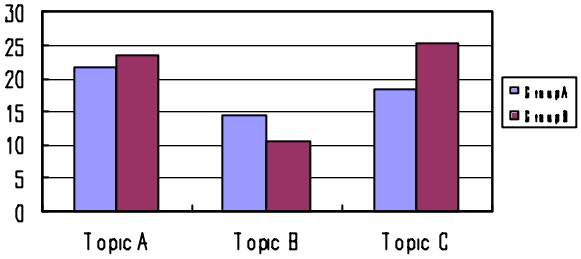

Figure 3. The mean time (min.) needed by the

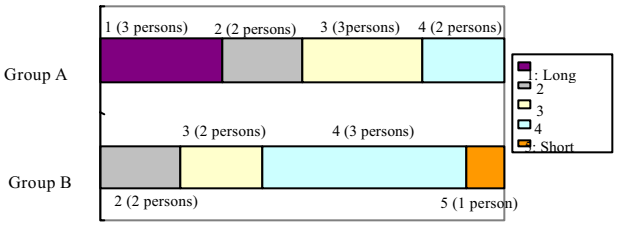

Figure 4. Subjective time needed to complete the task

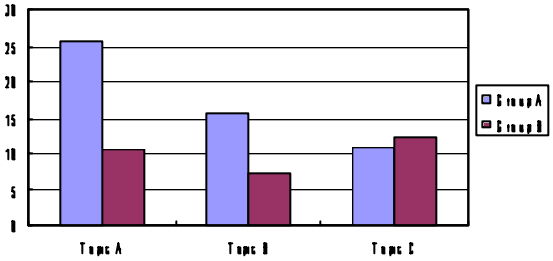

Figure 5. The mean number of pages that the subjects visited in each topic

Question	Answer
Q1: How many comments did you refer to?	All comments: 5 persons Almost all comments: 3persons Half comments: 2 persons
Q2: Did you take interests in the topics by referring to comments?	Definitely yes: 1 person Yes: 3 persons No: 6 persons
Q3: Are comments helpful for your search activity?	Definitely yes: 2 persons Yes: 8 persons

Table 2. Results of questionnaires

4.2 Considerations

Although the effectiveness of comment sharing is doubtful according to Fig. 3, Fig. 4 indicates that group B completed the task without excessive loads. Actual time consumption of group B was increased by thorough reference to comments. This ground appears as the results of questionnaires shown in Table 2. Group B noticed the importance of comments and made active use of the comments to select suitable pages from the search results. Furthermore, several of them took interest in the topics by referring to comments.

Figure 5 indicates that group B found suitable pages efficiently by referring to comments. It is clearly conceivable that comments attached to the search result reduced a load in search activity. Group A had to infer to the content of a page from summary texts generated by goo and actually visit the page to confirm whether it satisfied an item. Repetition of this activity increased the number of pages visited and subjective time consumption of group A. Overall from the above considerations, it may be concluded that comment sharing reduced loads in search activity.

The fact that there is no notable difference between two groups in Topic C may be caused by its trait in search activity that both groups had to traverse a few layers to find suitable pages. In other words, almost all comments in Topic C were superficial (i.e. Comment type 3) and did not describe paths to those suitable pages. Therefore the same traverse was imposed on both groups however the number of pages visited by group B was unexpectedly increased. From this situation, WebCOBALT was recognized to be effective under the condition that the trait of comments corresponds to that of a task.

5. Prospective Adaptivity

WebCOBALT is not currently an adaptive system. Does it need to have adaptivity? What does it adapt? How does it build a learner model? To expand this system to a web-based adaptive IR system, these questions need to be considered.

The increase of comments causes information overload. As described above, the current system displays meaningless comments unrelated with a task. To make matters worse, it may rank pages with many meaningless comments higher. This situation disrupts the search and

consequently lowers learning effectiveness. Adaptivity can be applied to remedy this problem. A significant issue is how the system identifies learners' tasks. Our intuition is that a task may be revealed as a search query. For example, a learner who is interested in Adaptive Hypermedia (AH for short) will designate "Adaptive Hypermedia" as a search query in order to find pages with comprehensive contents. A learner who understands AH roughly but is unfamiliar with AI technology used in AH will designate "Adaptive Hypermedia AI technology" in order to find pages with detailed contents. The following assumptions are therefore derived.

- *Similarity between tasks can be replaced with similarity between learners.*
- *Similarity between learners can be replaced with similarity between search queries.*
- *A learner prefers pages that similar learners estimated to be informative.*

The current system already stores search queries, URLs, and attributes of page estimation. These data can be used as a learner model. Similarity between search queries can be calculated by means of simple keyword matching, the Pearson's correlation coefficient or other methods. A principal element to be adapted in the current system is comments. Meaningless comments unrelated to the learner's current task should be eliminated. If that happens, pages related with the task will be ranked higher and the learner may find a suitable page smoothly. The prospective adaptivity will be actualized using the following adaptation technologies.

Adaptive Link Sorting: To sort pages (hyperlinks) in order of high similarity between tasks (specifically, the current task of a learner and the past tasks of other learners) is very profitable for learners. WebCOBALT first calculates the similarity between search queries from search queries of comments on pages retrieved. Comments with low similarity are eliminated. Secondly, it picks up pages with informative comments from pages with high similarity. The pages picked up will be suitable for the current task. Thirdly, it sorts these pages in order of useful information.

Adaptive Link Generation: Numerous web pages are created daily and search results are frequently renewed. In this situation, pages that similar learners estimated to be informative are not always retrieved. WebCOBALT copes with this problem by generating hyperlinks that connect to suitable pages independently of the search result. After performing the above adaptive link sorting, it first picks up pages that do not exist in the current search result but were previously evaluated to be informative by similar learners. Secondly, it sorts these pages in order of useful information. Thirdly, it generates hyperlinks that connect to these pages.

6. Conclusions

This paper has described the web retrieval support system with a comment sharing environment and indicated that this system can be extended to a web-based adaptive IR system. To combine collaborativity with adaptivity will enhance the quality of search engines. However it is true that this research is currently insufficient. We recognize that further considerations, surveys, and experiments are required in order to prove soundness of our idea for the adaptivity.

Acknowledgements

This research was supported in part by the grant-in-aid for Scientific Research (B)(2) No.13480047 from Japan Society for the Promotion of Science.

This research was supported by a grant to RCAST at Doshisha University from the Ministry of Education, Science, Sports and Culture, Japan.

References

1. Alexa Web Search. http://www.alexa.com/

2. Beeferman, D., and Berger, A.: Agglomerative clustering of a search engine query log, Proc. of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.407-415 (2000).

3. Bharat, K.: SearchPad: Explicit Capture of Search Context to Support Web Search, Proc. of WWW9 International Conference (2000). http://www9.org/w9cdrom/173/173.html

4. Brusilovsky, P.: Adaptive Hypermedia, User Models and User-Adapted Interaction 11, pp87-110 (2001).

5. Pazzani, M., Muramatsu, J., Billsus, D.: Syskill & Webert: Identifying interesting web sites, Proc. of the Thirteenth National Conference on Artificial Intelligence (AAAI'96), pp.54-61 (1996). http://www.ics.uci.edu/~pazzani/Sy skill.html

6. Marinilli, M., Micarelli, A., Sciarrone, F.: A Case-Based Approach to Adaptive Information Filtering for the WWW, Proc. of the Second Workshop on Adaptive Systems and User Modeling on the World Wide Web, pp.81-87 (1999).

7. Huang, C. K., Oyang, Y. J., and Chien L. F.: A Contextual Term Suggestion Mechanism for Interactive Web Search, Proc. of the First Asia-Pacific Conference on Web Intelligence: Research and Development (WI2001), pp.272-281 (2001).

8. Liu, H., Lieberman, H., and Selker T.: GOOSE: A Goal-Oriented Search Engine with Commonsense, Proc. of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2002), pp.253-263 (2002).

9. Ogata, H., Imai, K., Matsuura, K., and Yano, Y.: Knowledge Awareness Map for Open-ended and Collaborative Learning on the World Wide Web, Proc. of the 9th International Conference on Computers in Education (ICCE'99), pp.319-326 (1999).

10. Miyake, N. and Masukawa, H.: Relation-Making to Sense-Making: Supporting College Students' Constructive Understanding with an Enriched Collaborative Note-Sharing System, Proc. of the 4th International Conference of the Learning Science, pp.41-47 (2000).

11. Kahan, J., Koivunen, M.R., Prud'Hommeaux, E., and Swick, R.R.: Annotea: An Open RDF Infrastructure for Shared Web Annotations, Proc. of the WWW10 International Conference (2001). Available online at http://www10.org/cdrom/papers/488/