Discovery of Memorized Features from Web Browsing History for Re-finding Vaguely-Memorized Pages

Shinichi Ikura
Graduate School of Information Science and Engineering, Ritsumeikan University
Kusatsu-shi, Shiga, Japan
Email: ikura@de.is.ritsumei.ac.jp

Tsukasa Kondo
Graduate School of Science and Engineering, Ritsumeikan University
Kusatsu-shi, Shiga, Japan
Email: tsukasa@de.is.ritsumei.ac.jp

Fumiko Harada and Hiromitsu Shimakawa
College of Information Science and Engineering, Ritsumeikan University
Kusatsu-shi, Shiga, Japan
Email: {harada, simakawa}@cs.ritsumei.ac.jp

Abstract—**A user browses huge number of web pages. There is the case that he is tempted to revisit a web page he has perused in the past. However, when the user remembers the contents of the web page vaguely, he cannot find the suitable keywords for rediscovering the web page by a search engine. When a user peruses a web page, he memorizes the fact of having perused the web page or the vague contents of the web page by various information. This paper proposes a method to support rediscovery of a web page which the user wants to peruse once again when the keywords which express the contents of the web page are not remembered by him. In order to use the proposed method, proper setup of questions is indispensable. Therefore, the experiment which investigates whether the question about what kind of contents should be set up or not was conducted. Then, the result that pictures in web pages are useful was obtained.**

Index Terms—**web browsing history, vaguely-memorized, information searching**

I. INTRODUCTION

In recent years, by the spread of the WWW, users are able to easily get information they want from web pages. Users visit a lot of web pages, and want to visit again the web pages they read in the past. In this case, a user will need to rediscover the web page by using the search function of the browsing history. However, it is sometimes difficult for them to rediscover the web page with a search engine. If a user memorizes the contents of the web page vaguely, the user cannot think of appropriate keywords in order to rediscover the web page. In this case, any search engine in useless for the users. The paper refers to the state users do not remember

keywords to rediscover the web page, as a vaguely-memorized state. Because of a vaguely-memorized state, users cannot rediscover the web page that they want to view again. When users are in vaguely-memorized states, let web pages they want to view again as purpose web pages. In this paper, we propose a method to narrow down candidates of purpose web pages, using the information other than keywords.

II. SEARCH IN BROWSING HISTORY

When a user views the web page, a user vaguely remembers information except keywords. It involves the background color of the page, the location of image in the page, the assignment of frames in the page, and so on. That information could be clues of finding purpose web pages from browsing history. And, when users search web page, they don't browse all the web page carefully obtained by the search[1]. Users watch each content in a short period of time, and also they skip many areas of the user's screen when they view the WWW [2]. As a result, that information is remembered in a vague way. That information except keywords could be too weak to narrow down web pages. Therefore, when that information is used to search web pages, it could be necessary to use that information together with many others. Since the user has visited the purpose web page, it is included in the browsing history of the user. When the user is not sure for keywords to identify the purpose web page, he would use many kinds of information such as the time he visits and the impression he has received from the page. If the browsing history is examined using those information, we can narrow down candidates of the purpose web page. In fact, there is the method that users

search the purpose web page to use memorized layout of web page[3]. When the user vaguely remembers the purpose web pages, history is used together with that information. Then, when users do not know keyword that represent context of purpose web page, they could be find purpose web page from browsing history.

III. SUPPORT THE REDISCOVERY OF THE VAGUELY-MEMORIZED WEB PAGE

A. The Rediscovery of Purpose Web Page by Question and Answer

In this paper, we propose a method to support the rediscovery of the purpose web page in a vaguely-memorized state. This method questions users interactively. It repeats to get answers for questions. This method scores web pages matching answers for questions. It presents candidate web pages in the form of ranking in order to rediscover the purpose web page. Figure 1 shows the entire flow of this method.

- *Questions for users:* at first, proposed method asks users yes-no questions. It makes users reply some features except keyword of the purpose web page. In this method, some questions have been set in advance.
- *Acquisition of the user answer:* since users are vaguely-memorized state, they could not type in detailed information for the purpose web page. The method permits users to reply to the question in either of "yes", "no", and "I don't know". The reply in the selection of the three, options makes users easy to answer questions. Users simply judge whether the purpose web page includes features of questions. Even though there are simple, they contribute to narrow down candidates of the purpose web page. When users cannot reply in question, they can select "I don't know".
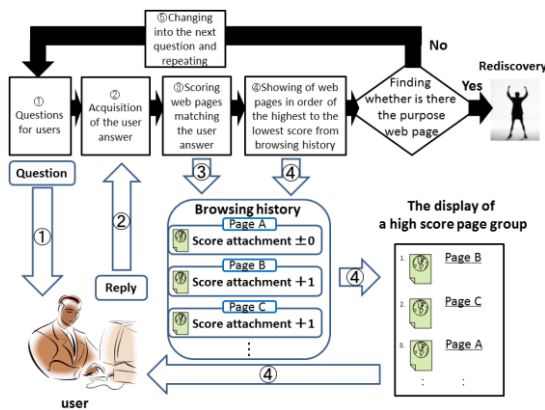

Figure 1. The entire flow of this method.

- *Scoring web pages matching the user answer:* the proposed method scores web pages matching the answer to question in browsing history. It means the proposed method narrow down candidates of the purpose web page from browsing history.
- *Showing of web pages in order of the highest to the lowest score from browsing history:* The

proposed method shows web pages in order of highest to lowest score from browsing history. Every time web pages of browsing history are scored, web pages to be presented are updated. Because users can confirm results of the scoring every question, they confirm whether the purpose web page is included in candidates each time. Users judge whether they repeat the questions and answers once again. When users answer to repeat the above four steps, they repeat steps with the next question prepared in advance. When users can find the purpose web page from candidate web pages which is currently presented, users respond that they stop searching.

B. Indices for Questions

The proposed method asks a user a question. The method narrows down candidates for the purpose web page from browsing history by understanding features of the purpose web page from users answers. As it is mentioned in Section 2, information that represents vaguely content of the purpose web page could be clue to find the purpose web page from browsing history. Here, we define an index as contents that are targeted by questions. When users are vaguely-memorized state, satisfied condition for index is content to be easy to remember by users.

IV. DISCOVERY AND VARIDATION OF FEATURES PAGE UNDER STORAGE

A. Preliminary Experiment

In the proposed method, in order to obtain the characteristics of the purpose web page, it is necessary to set of questions in advance. As mentioned in section 3.2, the question will be set based on the index. Even in a vaguely-memorized state, users often remember some characteristics of the purpose web page, such as the color, the layout, and images involved in it. The questions use indexes related to web page contents that are likely to be remembered by users. Experiment 1 is conducted in order to investigate the high possibility contents that users are likely to remember in a vaguely-memorized state. 10 subjects join to the experiment. Each of them is pointed to a person for him to ask questions. The experiment is carried out in accordance with the following steps. IN pairs of a person to ask and a subject. First, giving a random theme to each subject, he browses several web pages a search engine recommends for the theme. When themes are changed, users search web pages again and browse them. This process is repeated 3 times. The browsing history of web pages is collected. The total time is 15 minutes for the first step. Next, a person selects one of the web pages at random from web pages viewed by the subject. The page is not indicated to the subject. The subject guesses the page, asking questions to the person, based on the memory of the subject. Here, we refer to the web page selected by the person as the selected web page.

The subject gives the person a question relating to features of web pages in order to guess the selected web

page. At this time, the subject does not know which in the selected web page in the browsing history. The subject gives yes-no question. Subjects repeat questions, until he cannot imagine any other question or he specifies the selected web page.

TABLE I. THE SUMMARY OF QUESTIONS OBTAINED FROM SUBJECTS DURING THE EXPERIMENT

Questionnaire	Number of persons to ask
Image	9 persons
Background color	8 persons
Text	6 persons
Link	5 persons
Movie	4 persons
Menu bar	4 persons
Advertising	4 persons
Blog	4 persons
Flash	3 persons

Table I is the summary of questions obtained from subjects during the experiment. That were classified subjectively represents what the subjects ask on the contents. This defines questionnaire. The number of question persons represents how many persons ask the questionnaire. From Table I, there are many questions about the contents of images and background colors.

That is, the image and the background color could be high useful potential as the index by searching the web browsing history as the background color conditions, even if you do not know the keywords that represent the web page, the purpose web page is rediscovered can be identified. In fact, the research to find a Web page using the color is performed [4]. Next, let us consider the image. The experiment reveals that may images are memorized by many users. However, the questions can be obtained from the subjects, was just what some people have different criteria for definition of "Is there an image of a person?" And "Is there an image of the brand logo?". It is necessary for users to clarify the question criteria in order to answer questions with "Yes" or "No". From existing researches, users read less carefully from the beginning from the contents of the web page, and it is known that users are said to be only 79% of look over them also [5]. It is important to the size of the image in order to recognize that there is the image for users in the short viewing time.

B. Experiment for Images

In this experiment, it verifies the size of the image and memorability for 10 subjects in order to set the criteria for determining questions about the image. First, we provide web page *'a', 'b', 'c', 'd'* and *'e'*. Here, let us specify web page *'c'* whose image in expanded as web page *' c '*, while web page *'c'* whose image is shrunk as web page *' c'' '*. We divide 10 subjects into 2 groups *'A'* and *'B'* of 5 persons. Subjects in group 'A' look at the web pages in order *'a', 'b', ' c ', 'd',* and *'e'*. Subjects in group *'B'* look at the web pages in order *'a', 'b', ' c'' ', 'd',* and *'e'*.

And then, we have subjects answer the order of impressed web pages from viewed web pages. We verify whether there is a difference in the order of web pages that left an impression for subjects of group A and group B. Table II shows the results of this experiment. Table II shows that web pages with expanded images gives a better easily impression than web pages of with shrunk images, can be confirmed. We can conclude that the sign of the image is effective as the index to specify the characteristics of the purpose web page. The image size is a memorable factor to be used to search the purpose web page from the web access log.

TABLE II. THE ORDER OF THE WEB PAGE THAT REMAIN IN THE IMPRESSION

Subjects	Group A
1	$b \rightarrow e \rightarrow c' \rightarrow d \rightarrow a$
2	$b \rightarrow c'' \rightarrow e \rightarrow a \rightarrow d$
3	$c' \rightarrow a \rightarrow e \rightarrow d \rightarrow b$
4	$c' \rightarrow e \rightarrow b \rightarrow a \rightarrow d$
5	$b \rightarrow c' \rightarrow e \rightarrow a \rightarrow d$
Subjects	**Group B**
6	$b \rightarrow c'' \rightarrow d \rightarrow e \rightarrow a$
7	$e \rightarrow b \rightarrow c'' \rightarrow d \rightarrow a$
8	$b \rightarrow e \rightarrow c'' \rightarrow a \rightarrow d$
9	$b \rightarrow e \rightarrow a \rightarrow c'' \rightarrow d$
10	$e \rightarrow b \rightarrow c'' \rightarrow a \rightarrow d$

V. CONCLUSION

I have conducted experiments to obtain characteristics to set a question to the user and the usefulness of the image. As a result, image and background in the web page was found to be effective as a feature to set the question. When I examine the result of the size of the image, the bigger the size of the image, the more it tends to remain in the user's memory. Therefore, by setting the questions related to the size of the image, it is considered likely that the question can be set useful for the user.

REFERENCES

[1] Takehiko Ohno: Where You Look While You Navigate the Web? –Eye mark analysis of WWW pages, 2000.???
[2] Jakob Nielsen: Report From a 1994 Web Usability Study, 1994.
[3] Yasunari Hashimoto: Web Page Search by Sketching Layout, [Online]. Available: http://www-ui.is.s.u-tokyo.ac.jp/~gyasu/projects/layoutsearch/overview.html
[4] Motonao Ishikawa: An Implementation Method of the Web Page Retrieval based on the Visual Impression, DEWS 2005.
[5] Jakob Nielsen, "How People Read" on the Web, Jakob Nielsen's AlertBox, October 1, 1997.

Shinichi Ikura received B.E from Ritsumeikan University in 2012. He advanced Graduate School of Ritsumeikan University. He engages in the research on data engineering. He is member of IPSJ.

Tsukasa Kondo received B.E from Ritsumeikan University in 2011. He advanced Graduate School of Ritsumeikan University. He engages in the research on data engineering. He is member of IPSJ.

Hiromitsu Shimakawa received Ph.D degree from Kyoto Univ. in 1999. He joined Ritsumeikan Univ. in 2002. Currently, he is a professor in Ritsumeikan Univ. His research interests include data engineering, usability, and integration of psychology with IT. He is a member of IEEE and ACM.

Fumiko Harada received B.E. and M.E, and Ph.D degrees from Osaka University in 2003, 2004, and 2007, respectively. She joined Ritsumeikan University as an assistant professor in Ritsumeikan University in 2007, and is currently a lecturer. She engages in the research on real-time systems and data engineering. She is a member of IEEE.